

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/331506634>

Co-occurring genomic capacity for anaerobic methane and dissimilatory sulfur metabolisms discovered in the Korarchaeota

Article in *Nature Microbiology* · April 2019

DOI: 10.1038/s41564-019-0362-4

CITATIONS

4

READS

228

9 authors, including:



Luke Justin McKay

Montana State University

26 PUBLICATIONS 324 CITATIONS

[SEE PROFILE](#)



Matthew W Fields

Montana State University

248 PUBLICATIONS 4,294 CITATIONS

[SEE PROFILE](#)



Tom Delmont

Genoscope - Centre National de Séquençage

75 PUBLICATIONS 1,341 CITATIONS

[SEE PROFILE](#)



A. Murat Eren

University of Chicago

191 PUBLICATIONS 2,507 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Terragenome and French ANR research project METASOIL [View project](#)



Microbially Enhanced Coalbed Methane [View project](#)

Co-occurring genomic capacity for anaerobic methane and dissimilatory sulfur metabolisms discovered in the Korarchaeota

Luke J. McKay^{1,2*}, Mensur Dlakić³, Matthew W. Fields^{2,3}, Tom O. Delmont^{4,8}, A. Murat Eren^{4,5}, Zackary J. Jay², Korinne B. Klingensmith², Douglas B. Rusch⁶ and William P. Inskeep^{1,7*}

Phylogenetic and geological evidence supports the hypothesis that life on Earth originated in thermal environments and conserved energy through methanogenesis or sulfur reduction. Here we describe two populations of the deeply rooted archaeal phylum Korarchaeota, which were retrieved from the metagenome of a circumneutral, suboxic hot spring that contains high levels of sulfate, sulfide, methane, hydrogen and carbon dioxide. One population is closely related to '*Candidatus Korarchaeum cryptofilum* OPF8', while the more abundant korarchaeote, '*Candidatus Methanodesulfokores washburnensis*', contains genes that are necessary for anaerobic methane and dissimilatory sulfur metabolisms. Phylogenetic and ancestral reconstruction analyses suggest that methane metabolism originated in the Korarchaeota, whereas genes for dissimilatory sulfite reduction were horizontally transferred to the Korarchaeota from the Firmicutes. Interactions among enzymes involved in both metabolisms could facilitate exergonic, sulfite-dependent, anaerobic oxidation of methane to methanol; alternatively, '*Ca. M. washburnensis*' could conduct methanogenesis and sulfur reduction independently. Metabolic reconstruction suggests that '*Ca. M. washburnensis*' is a mixotroph, capable of amino acid uptake, assimilation of methane-derived carbon and/or CO₂ fixation by archaeal type III-b RuBisCO for scavenging ribose carbon. Our findings link anaerobic methane metabolism and dissimilatory sulfur reduction within a single deeply rooted archaeal population and have implications for the evolution of these traits throughout the Archaea.

Methanogenesis and sulfate reduction are considered two of the earliest evolved mechanisms for microbial energy conservation^{1–4}. Genes associated with both of these energy metabolisms have recently been discovered in diverse taxonomic lineages. Genes that encode methyl:coenzyme M reductase (Mcr), which catalyses the final step in methanogenesis, were recently discovered in the Bathyarchaeota⁵ and Verstraetearchaeota⁶, and overturned the long-held paradigm that this functional capacity was restricted to the Euryarchaeota. Distantly related *mcr* homologues were shown to be involved in butane oxidation in members of '*Candidatus Syntrophoarchaeum*'⁷, calling into question the presumed methanogenic function of distant homologues from the Bathyarchaeota⁸. Similarly, dissimilatory sulfite reductases (Dsr), which catalyse the conversion of sulfite to hydrogen sulfide during the process of sulfate reduction, were recently expanded to thirteen additional bacterial and archaeal lineages⁹. Discoveries of Mcr- and Dsr-encoding genes among diverse microbial phyla continue to expand our understanding of the evolution of methane and sulfur metabolisms.

Both Mcr and Dsr can function in reverse to mediate the oxidation of methane¹⁰ and reduced sulfur¹¹, respectively, by phylogenetically distinct groups of archaea and bacteria. The predominant mode of anaerobic oxidation of methane is thought to be completed by microbial consortia of anaerobic methanotrophic (ANME) archaea and sulfate-reducing bacteria living in

marine sediments^{12,13} and involves reverse Mcr and forward Dsr functions (reaction 1).



Several mechanisms have been proposed for this interspecies redox couple, including the use of cytochromes¹⁴, sulfur disproportionation¹⁵ and pili- or flagellum-like proteins¹⁶. Whether a single organism can conduct both methane and sulfur metabolism through Mcr and Dsr has not yet been established. Here we performed metagenome sequencing of highly sulfidic and pyritic sediments from a suboxic geothermal spring in Yellowstone National Park that contained high levels of methane, carbon dioxide, hydrogen and sulfate. Two populations of the poorly understood archaeal phylum Korarchaeota were discovered, including one with co-occurring *mcr* and *dsr* complexes. Phylogenetic analyses, energetic calculations and detailed metabolic considerations highlight three possible energy-conserving strategies, including the highly exergonic, incomplete oxidation of methane to methanol with sulfite. These observations link anaerobic methane oxidation and sulfur reduction within a single deeply rooted population and provide far-reaching implications for the evolution of methane and sulfur metabolism in the Archaea.

Results and discussion

Comparison of korarchaeotal genomes. We recovered two near-complete korarchaeotal genomes from a metagenome of sulfidic,

¹Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT, USA. ²Center for Biofilm Engineering, Montana State University, Bozeman, MT, USA. ³Department of Microbiology and Immunology, Montana State University, Bozeman, MT, USA. ⁴Department of Medicine, University of Chicago, Chicago, IL, USA. ⁵Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole, MA, USA. ⁶Informatics Group, Indiana University, Bloomington, IN, USA. ⁷Thermal Biology Institute, Montana State University, Bozeman, MT, USA. ⁸Present address: Genoscope, Evry, France. *e-mail: luke.mckay@montana.edu; binskeep@montana.edu

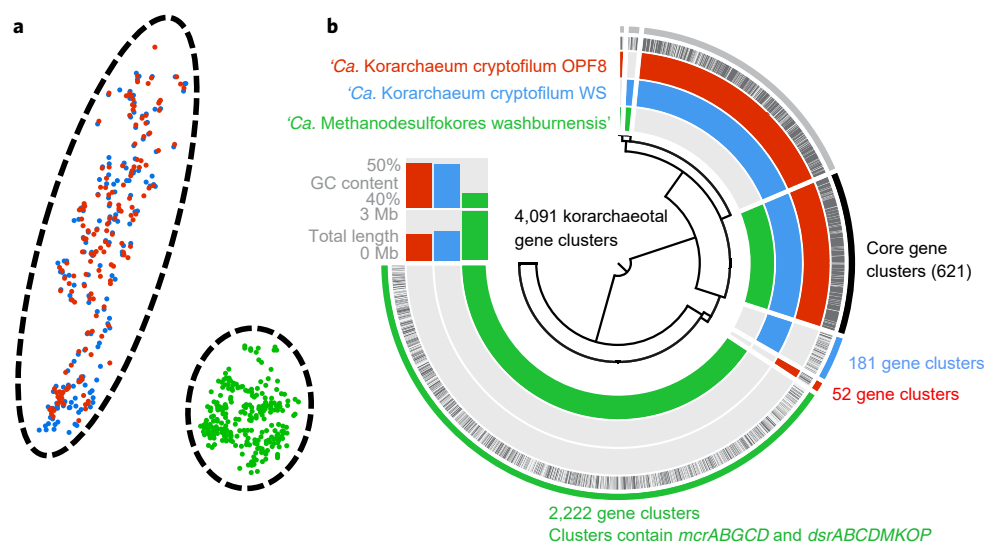


Fig. 1 | Tetranucleotide frequency display and comparison of genome sequence for three populations of Korarchaeota. a, A Two-dimensional snapshot of a three-dimensional analysis generated by tSNE of tetranucleotide frequencies. Dots represent assembled scaffolds chopped at a maximum length of 20 kb. $n = 203$ for 'Ca. M. washburnensis' and $n = 92$ for 'Ca. K. cryptofilum WS'. **b,** Differential distribution of homologous gene clusters⁵⁸ (based on translated open reading frames) across the three genomes. Red, 'Ca. K. cryptofilum OPF8'; blue, 'Ca. K. cryptofilum WS'; green, 'Ca. M. washburnensis'.

thermal sediments from Washburn Hot Springs (Fig. 1 and Table 1). One population was highly similar to 'Ca. K. cryptofilum OPF8'¹⁷ with an average nucleotide identity (ANI) of 98.7% (Supplementary Table 1), hereafter referred to as 'Candidatus Korarchaeum cryptofilum WS' (for Washburn Hot Springs). 'Ca. K. cryptofilum OPF8' was enriched from Obsidian Pool, another hot spring in Yellowstone National Park that exhibits a similar pH, temperature and low-oxygen status to that of Washburn Hot Springs^{18,19}. The second korarchaeotal genome was significantly different from 'Ca. K. cryptofilum OPF8' with an ANI of 70%, average amino acid identity of 45% (Fig. 1, Table 1 and Supplementary Table 1) and markedly lower GC content (42.7%) relative to the two 'Ca. K. cryptofilum' genomes (48.4 and 49.0%; Fig. 1 and Table 1).

Several gene clusters (15.2%) were common to the two Korarchaeota populations described here as well as 'Ca. K. cryptofilum OPF8'¹⁷ (Fig. 1b and Supplementary Table 3). Marker genes for anaerobic methane metabolism (for example, *mcrA*) and dissimilatory sulfite reduction (for example, *dsrAB*) were found within a large set of gene clusters belonging exclusively to the divergent korarchaeotal population (comprising 54.3% of the total number of detected clusters). Previously, genes involved in methane metabolism and dissimilatory sulfite reduction have not been found within a single population. The proposed name, 'Ca. M. washburnensis', combines these putative functions with the Greek root used for Korarchaeota (kore, young woman²⁰) and the location where this genome was detected (Washburn Hot Springs). Genome completeness estimates were 95.1% for 'Ca. M. washburnensis', 95.7% for 'Ca. K. cryptofilum WS' and 95.7% for 'Ca. K. cryptofilum OPF8' (Table 1). Redundancy values estimated from multiple copies of single-copy genes ranged from 1.5 to 4.3% for all three Korarchaeota. Together, these results suggest that the curated genome assemblies are highly complete, contain very few redundancies and are of high quality²¹.

An analysis of single-nucleotide variants showed that both genomes represented near-clonal environmental populations with low levels of strain heterogeneity (Table 1). However, we observed variable coverage patterns across both korarchaeotal genomes and in that of 'Ca. M. washburnensis', contigs containing genes for methane and sulfur metabolisms fell within different coverage groups at 109× and 818×, respectively (Supplementary Fig. 2). We

performed multiple tests to ensure that each sequence cluster identified through *t*-stochastic neighbour embedding (tSNE) analysis²² of tetranucleotide frequencies accurately corresponded to unique korarchaeotal populations. These tests included additional analyses with five separate assemblies (Supplementary Table 2), multiple variations of three clustering algorithms (Fig. 1 and Supplementary Figs. 1, 2, 4), careful examination of specific sequence content (that is, single-copy genes, necessary cellular processes, and methane and sulfur metabolisms), sequence character and contig overlaps, and taxonomic assignments (Supplementary Figs. 1–3 and Supplementary Discussion). We also included positive controls with two clustering algorithms—tSNE and emergent self-organizing maps²³. These analyses demonstrated a perfect overlap of the complete genome of 'Ca. K. cryptofilum OPF8'¹⁷ on the sequence for the closely related 'Ca. K. cryptofilum WS' (Fig. 1a) and complete recovery of *Methanopyrus kandleri* AV19²⁴ and *Metallosphaera sedula* DSM5348²⁵ reference genomes added to the metagenome from Washburn Hot Springs (Supplementary Fig. 4). In summary, our analyses of different assemblies and sequence clustering techniques indicated that the genome of 'Ca. M. washburnensis' is representative of either a single microbial population with significant genomic sequence variability or a highly related group of korarchaeotal subpopulations that have a nearly identical core genome in addition to strain-specific genes (that is, a 'pangenome'²⁶). Notably, genes for sulfur and methane metabolisms co-occur in 'Ca. M. washburnensis' in either scenario (Supplementary Discussion).

The estimated genome size for 'Ca. M. washburnensis' was 2.9 Mb (3,578 coding genes), which was considerably larger than the genome sizes of 1.8 and 1.6 Mb (1,989 and 1,617 coding genes) for 'Ca. K. cryptofilum WS' and 'Ca. K. cryptofilum OPF8', respectively. Coverage values and resulting abundance estimates were much higher for 'Ca. M. washburnensis' compared to 'Ca. K. cryptofilum WS', which indicates that 'Ca. M. washburnensis' represented a greater fraction (2.5% of metagenomic reads) of the microbial community at Washburn Hot Springs than 'Ca. K. cryptofilum WS' (0.8% of metagenomic reads; Table 1). 'Ca. M. washburnensis' was the fourth most abundant of 135 putative genomes recovered from this metagenome, which suggests that this organism has a key role in the Washburn ecosystem.

Table 1 | Genomic characteristics of the Korarchaeota

	' <i>Ca. Methanodesulfokores washburnensis</i> '	' <i>Ca. Korarchaeum cryptofilum</i> WS'	' <i>Ca. Korarchaeum cryptofilum</i> OPF8'
Length (bp)	2,942,065	1,742,982	1,590,757
Completeness (%)	95.1	95.7	95.7 ^a
Redundancy (%)	4.3	3.1	1.5 ^a
N50 (bp)	24,753	78,545	NA
Number of scaffolds	179	51	1
Number of genes	3,578	1,989	1,617
GC content (%)	42.7	48.4	49.0
Number of tRNAs	48	47	45
Number of rRNAs	3	3	3
Mean coverage	377	114	NA
Relative abundance (%)	2.5	0.8	NA
SNV density	1.22	0.94	NA

Estimated completeness, redundancy, median scaffold size (N50), GC content, total number of genes encoding tRNAs and rRNAs (that is, 5S, 16S and, 23S) and mean coverage of the Korarchaeota populations are compared with '*Ca. K. cryptofilum* OPF8'¹⁷. The genome for '*Ca. K. cryptofilum* OPF8' has been published previously¹⁷. Relative abundance represents the percentage of metagenomic short reads recruited by a particular population. Single-nucleotide variant (SNV) density is the percentage of genomic positions that show nucleotide-level variation based on metagenomic short read recruitment. NA, not applicable. ^aCompleteness and redundancy estimates of a complete Korarchaeota genome based on archaeal single-copy core genes.

Phylogenomic analysis of the Korarchaeota. Phylogenomic analyses using an alignment of 56 conserved proteins (Supplementary Table 4) showed that '*Ca. K. cryptofilum* OPF8' and '*Ca. K. cryptofilum* WS', as well as the more distantly related '*Ca. M. washburnensis*', formed a deeply rooted monophyletic clade among the Archaea (Fig. 2a). Strong posterior probabilities at all nodes reinforced the basal position of this phylum. Additional phylogenomic analyses with representatives from all three domains of life confirmed the deeply rooted placement of the Korarchaeota but yielded different relationships between the Archaea and Eukarya (Supplementary Fig. 5 and Supplementary Discussion). Most trees indicated monophyly of the Korarchaeota with the Eukarya, whereas a minority of trees supported an Asgard–Eukarya clade^{27,28}; both phylogenomic results were accepted with significance by some posterior predictive tests (Supplementary Table 5). Collectively, these results are qualitatively similar to previous observations^{29,30} and an important reminder that the selection of proteins, microbial species and the parameters for phylogenomic analyses strongly affect interpretations of the perceived evolutionary history of the Eukarya. More genomic entries from the Korarchaeota and Asgard archaea will help to clarify these discrepancies. A phylogenetic comparison of full-length 16S rRNA gene sequences demonstrated that '*Ca. K. cryptofilum* WS' was 99% similar to that of '*Ca. K. cryptofilum* OPF8'¹⁷ and 92% similar to '*Ca. M. washburnensis*', which was more closely related to korarchaeotal sequences from thermal environments in Iceland, Russia and other sites in North America (Supplementary Fig. 6 and Supplementary Discussion).

Phylogenetic analysis of korarchaeotal McrA and DsrAB. The deduced McrA sequence from '*Ca. M. washburnensis*' represents a basal entry relative to other methane-metabolizing organisms (Fig. 2b), which is consistent with the deeply rooted position of the Korarchaeota within the Archaea (Fig. 2a) and the universal tree of life (Supplementary Fig. 5). The McrA sequence from '*Ca. M. washburnensis*' was placed with strong posterior probability near the branches of two recently proposed lineages of methylotrophic methanogens, the Verstraetearchaeota⁶ and the Methanomassiliicoccales³¹. Sequences from the uncultivated ANME-1 group of methane-oxidizing archaea formed an adjacent clade to the Korarchaeota, but the branch length between these groups was relatively large. These findings support the involvement of methylated compounds and the possibility of anaerobic oxidation of methane in the Korarchaeota. McrA sequences from

Korarchaeota and Verstraetearchaeota⁶ resolved the previously uncharacterized clades, 'deeply branching *mcrA* groups' 1 and 3³² (Supplementary Fig. 7 and Supplementary Discussion).

The wide taxonomic distribution of McrA among diverse archaea suggests that lateral gene transfer (LGT) may have occurred throughout the evolution of methane metabolism. For example, McrA sequences from the Methanomassiliicoccales are more closely related to those from the Korarchaeota and Verstraetearchaeota than to other Euryarchaeota such as the Methanocellales and Methanomicrobiales. Ancestral reconstructions of McrA from two deep node selections (Supplementary Fig. 8a) were more closely related to McrA from '*Ca. M. washburnensis*' (84.2% and 80.2%) than any other McrA sequence available in the National Center for Biotechnology Information (NCBI) database. Additionally, we extracted all other deduced McrA sequences from the metagenome from Washburn Hot Springs (Integrated Microbial Genomes and Microbiomes (IMG) genome identifier: 3300005860), but none were as similar to the ancestral sequence as that of '*Ca. M. washburnensis*'. These observations provide further support that korarchaeotal *mcrA* may have developed relatively early in the evolution of archaeal methane metabolism. However, given the frequency of phylum-level additions to the McrA tree in the past three years, it is probable that our knowledge of Mcr-containing lineages remains incomplete.

Phylogenetic analysis of deduced DsrAB proteins from '*Ca. M. washburnensis*' showed that these sequences were most closely related to Aigarchaeota from Great Boiling Springs, Nevada, USA³³ and Jinze Pool, Yunnan, China³⁴. The Korarchaeota and Aigarchaeota DsrAB sequences fell within a larger clade that also contained three cultivated groups of sulfate-reducing Clostridia (*Carboxydotherrmus*, *Desulfosporosinus* and *Desulfotomaculum*) and several clades of uncharacterized environmental sequences. This large cluster was adjacent to other reductive-type DsrAB clades that belong to the Desulfobacca and Deltaproteobacteria. By contrast, oxidative-type DsrAB sequences (for example, Chlorobi, Alpha-, Beta-, Gamma-proteobacteria) formed a separate, distant clade consistent with previous observations^{9,35}. These findings together with the presence of *dsrD*, which is absent from oxidative-type Dsr-expressing organisms³⁶, provide strong support that the deduced DsrAB proteins found in '*Ca. M. washburnensis*' are involved in the dissimilatory reduction of sulfur compounds (that is, sulfite, thio-sulfate and/or elemental sulfur).

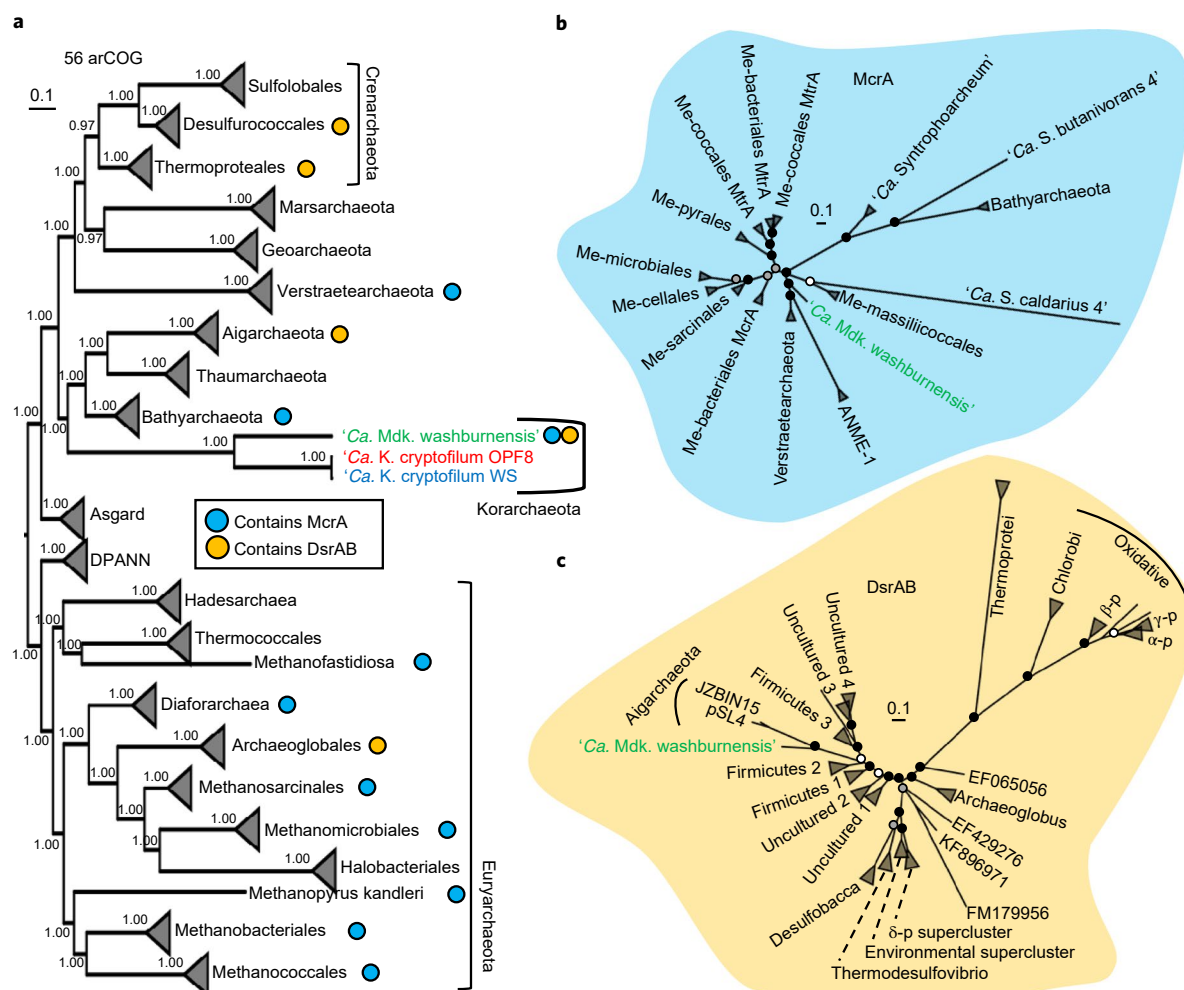


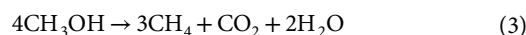
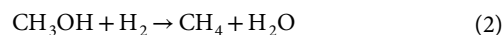
Fig. 2 | Phylogeny of the Korarchaeota. a, Phylogenomic analysis of representative archaea ($n = 109$ genomes; 56 concatenated proteins; Supplementary Table 4) using Bayesian methods. Blue and orange circles indicate lineages in which *mcrA* and *dsrAB* have been detected, respectively. **b, c**, Bayesian phylogenetic analyses of *McrA* (**b**; $n = 126$ sequences) and *DsrAB* (**c**; $n = 138$ sequences) were also performed. Posterior probabilities for the phylogenomic tree (**a**) are reported on each node. Posterior probabilities for *McrA* and *DsrAB* trees are indicated by black, grey and white circles at each node corresponding to >0.9 , >0.7 and >0.5 , respectively. Me, Methano; Mdk, Methanodesulfokores.

The association between the Firmicutes and the two Aigarchaeota sequences has previously been attributed to LGT³⁴, and other researchers have noted that LGT was probably frequent in *dsrAB* evolution⁹. Interspersed lineages of Archaea and Bacteria in the *DsrAB* tree support LGT, but the directionality of lateral events is unclear. To examine this issue, ancestral reconstructions of *DsrAB* were built from three basal nodes (Supplementary Fig. 8b). The reconstructed sequences from the deeper two nodes were more closely related to Firmicutes *DsrAB* sequences than ‘*Ca. M. washburnensis*’ or Aigarchaeota. Out of a total of 112 *Dsr* homologues found in the metagenome from Washburn Hot Springs, six were more similar to the ancestral sequence than the *DsrAB* from Korarchaeota (Supplementary Table 6). These six sequences were most closely related to members of the Firmicutes (for example, *Thermoanaeromonas*, *Pelotomaculum* and *Calderihabitans*). Consequently, the ancestral reconstruction of *DsrAB* successfully identified members of the Firmicutes living in the same habitat that are candidate donors of *dsr* transfer to ‘*Ca. M. washburnensis*’.

Energy metabolism in ‘*Ca. M. washburnensis*’. Metabolic reconstruction of ‘*Ca. M. washburnensis*’ revealed enzyme-encoding genes for two major energy conservation pathways—methanogenesis

and dissimilatory sulfite reduction. Here we discuss three metabolic possibilities (Fig. 3) and outline observations that support or challenge each.

Methanogenesis from methanol and hydrogen. The presence of a full *mcr* complex (ABG subunits and CD-related proteins) and complete methanol:coenzyme M methyltransferase (*mtaABC*) suggests that ‘*Ca. M. washburnensis*’ is capable of using methanol as a substrate for methanogenesis (Fig. 3a). The absence of the methyltetra hydromethanopterin:coenzyme M methyltransferase (*mtr*) sodium ion-translocating system, the absence of the methyl branch of the Wood–Ljungdahl pathway³⁷ and the presence of a complete F_{420} -nonreducing hydrogenase (*mvhADG*) each indicates that ‘*Ca. M. washburnensis*’ is capable of using hydrogen to reduce methanol (reaction 2) rather than disproportionating methanol to methane and CO_2 (reaction 3)³⁸.



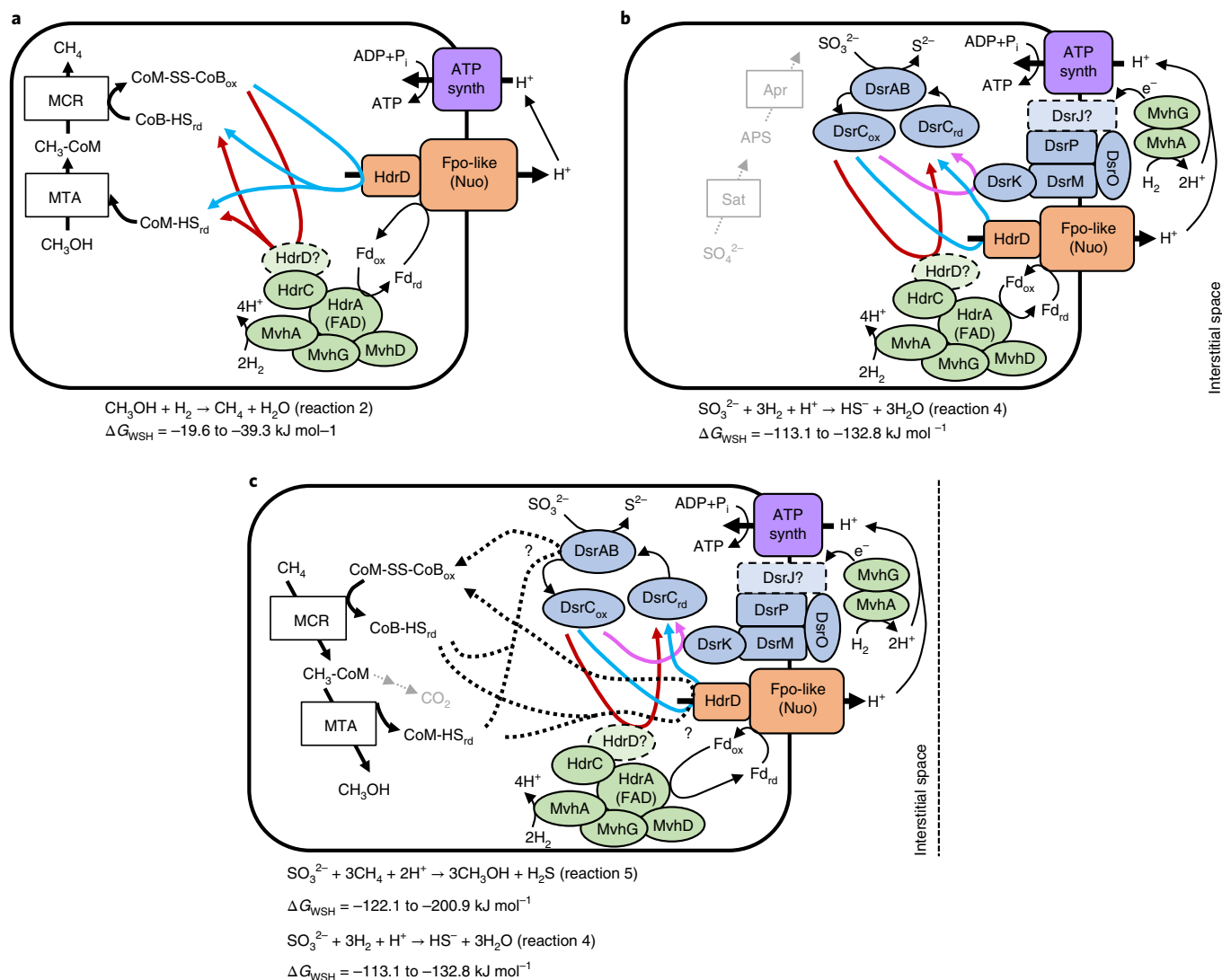


Fig. 3 | Possible energy metabolisms for ‘Ca. M. washburnensis’. **a–c.** Three energy-conserving strategies are presented based on genetic content, including methanogenesis from methanol and hydrogen (**a**), sulfite reduction with hydrogen (**b**) and sulfite reduction with methane and hydrogen (**c**). Black dotted lines represent subunits with proposed homologous replacements (for example, HdrD replacing HdrB) or hypothetical interactions. Grey dotted lines represent absence of pathways. No classical DsrJ was found, but a cytochrome *b* subunit of the *bc* complex was encoded in the same operon as DsrABCKM (see dashed lines around DsrJ; Supplementary Fig. 11). Fpo, F₄₂₀:methanophenazine oxidoreductase; Hdr, heterodisulfide reductase; MCR, methyl coenzyme M reductase; MTA, methanol:coenzyme M methyltransferase; Mvh, NiFe F₄₂₀-nonreducing hydrogenase; rd, reduced; ox, oxidized; synth, synthase; CoM, coenzyme M; CoB, coenzyme B; SS, disulfide bond; HS, monosulfide.

The A and C subunits of a heterodisulfide reductase (*hdrAC*) were also encoded within the same operon as *mvh*. The Mvh NiFe-hydrogenase forms a complex with HdrABC and is thought to regulate the oxidation of hydrogen while reducing ferredoxin and CoM-SS-CoB in an electron-bifurcating reaction during hydrogen-dependent methanogenesis^{39,40}. However, the HdrB subunit, which catalyses the direct reduction of CoM-SS-CoB, was not found in the ‘Ca. M. washburnensis’ assembly. It was recently proposed that a homologous heterodisulfide reductase, HdrD, coupled to an Fpo-like proton-translocating pump is responsible for establishing a membrane potential in the Methanomassiliicoccales⁴¹. Indeed, ‘Ca. M. washburnensis’ has genes that encode this Fpo-like complex and the HdrD protein that binds to it; this suggests that the energy metabolism of ‘Ca. M. washburnensis’ may be similar to that proposed for ‘*Candidatus* Methanoplasma termitum’. The shared homology and similar function of HdrD and HdrB suggests that HdrD may also replace HdrB in the Mvh–Hdr complex.

On the basis of the physicochemical properties of Washburn Hot Springs (Supplementary Table 7) and formation energies of all relevant chemical species⁴², the free energy yield (ΔG) of hydrogen-dependent methylotrophic methanogenesis (reaction 2; Fig. 3a) was estimated to range from -19.6 to $-39.3 \text{ kJ mol}^{-1}$. The volatilization of methanol probably limits its availability as a substrate for methanogenesis. At the elevation of Washburn Hot Springs (1,883 m), methanol has a boiling point of 54°C , while the temperature of the spring is $65\text{--}70^\circ\text{C}$.

Sulfite reduction with hydrogen. ‘Ca. M. washburnensis’ may perform sulfite or thiosulfate reduction as an alternative energy-conservation strategy (Fig. 3b). In addition to DsrAB, genes were detected for the related proteins DsrC and DsrD, and 4 out of the 5 subunits of the membrane-bound DsrMKJOP complex. Subunit J was missing from *dsrMKJOP*, but the function of this trihaeme cytochrome might be replaced by a dihaeme cytochrome *b* of the *bc*

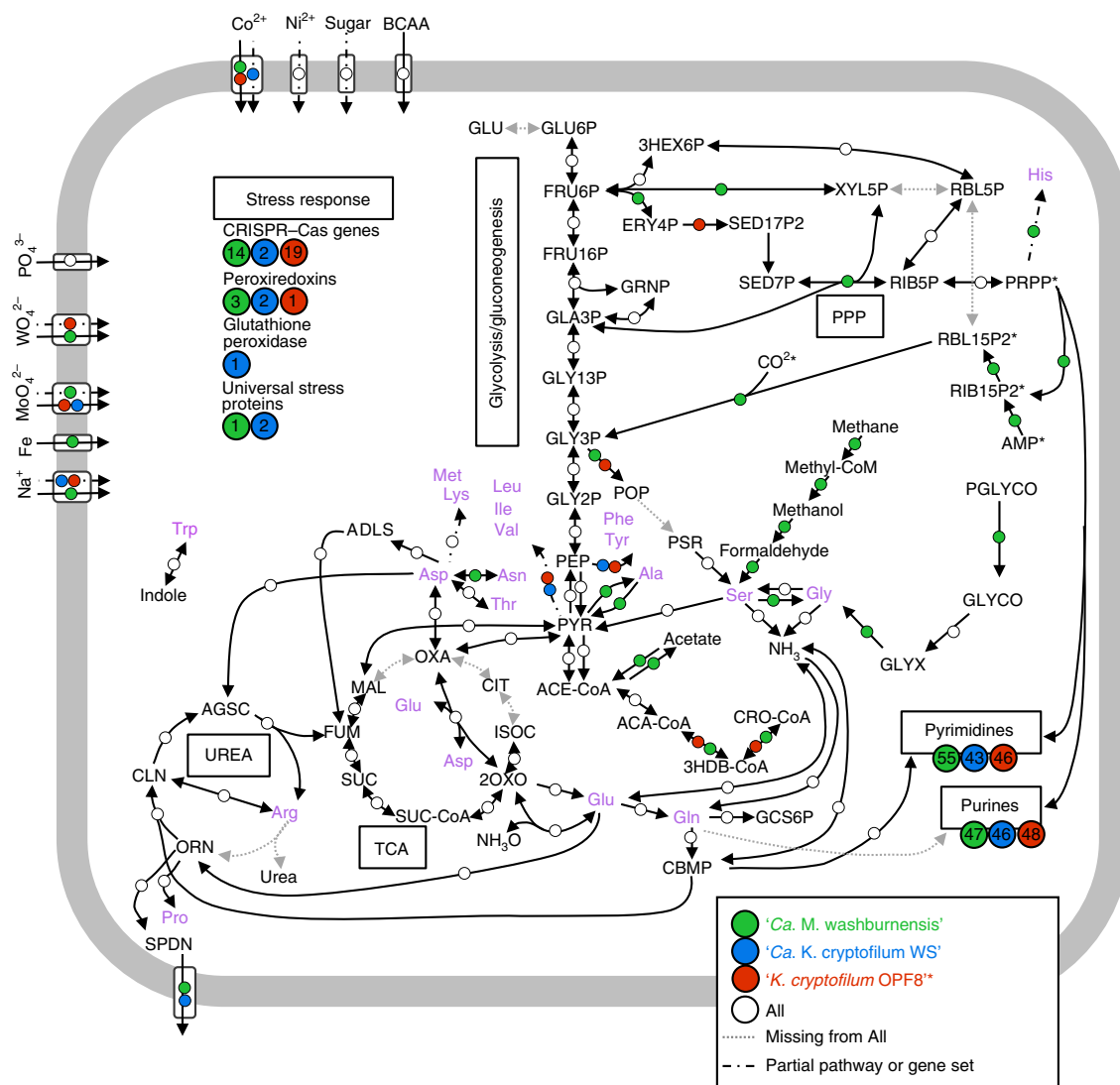
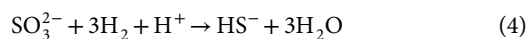


Fig. 4 | Carbon and intermediate metabolism in the Korarchaeota. Metabolic reconstruction of three Korarchaeota (coloured circles) based on the presence of protein-coding genes. Complete gene or gene complexes are indicated by black arrows; partially complete complexes are indicated by dashdot black arrows; incomplete complexes are indicated by dotted grey arrows. Green, 'Ca. M. washburnensis'; blue, 'Ca. K. cryptofilum WS'; red, 'Ca. K. cryptofilum OPF8'; white, present in all. Cases in which one population has a full complex and another has a partial complex are indicated by two arrows with the appropriate colour designations (for example, the cobalt transporter). Amino acids are in magenta with three-letter codes according to the International Union of Pure and Applied Chemistry. Abbreviations are described in Supplementary Table 8. Asterisks indicate steps that are involved in the carbon-assimilation pathway, which is outlined in more detail in Supplementary Fig. 10.

complex encoded within the same operon. Like hydrogenotrophic methanogens, many sulfate-reducing organisms also have Mvh-Hdr complexes^{43,44}, which are thought to be used for the reduction of sulfite with hydrogen (reaction 4; Fig. 3b). Together with the Mvh hydrogenase, the Dsr components suggest that hydrogen is used as an electron donor for sulfite reduction (reaction 4).



The reduction of sulfite with hydrogen was estimated to be exergonic under the conditions at Washburn Hot Springs ($\Delta G_{\text{WSH}} = -113.1$ to $-132.8 \text{ kJ mol}^{-1}$). This suggests that 'Ca. M. washburnensis' could either be a facultative methanogen/sulfite reducer that alternately employs distinct energy conservation strategies based on environmental conditions (that is, Fig. 3a,b) or a sulfite reducer that uses both hydrogen and methane as electron donors (Fig. 3c).

Anaerobic oxidation of methane to methanol via sulfite reduction. The co-occurrence of *mcr*, *mta* and *dsr* complexes in 'Ca. M. washburnensis' suggests the possibility of a previously unidentified metabolism in which methane oxidation to methanol is coupled with sulfite reduction (reaction 5; Fig. 3c). Calculations of ΔG demonstrate that the sulfite-dependent anaerobic oxidation of methane to methanol is highly exergonic (between -122.1 and $-200.9 \text{ kJ mol}^{-1}$) in Washburn Hot Springs:



In this scenario, Mcr and Mta operate in reverse^{10,45} to produce methanol from methane, which results in the release of reduced CoB-SH and CoM-SH that provide substrates for reverse heterodisulfide reductase (that is, 'dual monosulfide oxidase' activity⁴⁶ at HdrD, DsrAB or DsrK (Fig. 3c and Supplementary Discussion). It is uncertain whether energy can be conserved from this exergonic

process; however, methane oxidation may be an important avenue for carbon assimilation (described below).

Carbon and intermediate metabolism in the Korarchaeota. We hypothesize that methanol produced by Mcr and Mta can be converted to formaldehyde by an alcohol dehydrogenase (Fig. 4 and Supplementary Fig. 9). Eight genes for alcohol dehydrogenases were present in ‘*Ca. M. washburnensis*’. Six of these encoded putative short-chain alcohol dehydrogenases, which were proposed to catalyse formaldehyde production from methanol in *Methanosarcina barkeri*⁴⁷ and ANME-1 spp.⁴⁸. Formaldehyde can then be converted to serine by glycine hydroxymethyltransferase; subsequent pyruvate production is mediated by serine-alanine lyase. The genes involved in the latter half of this pathway ($\text{CH}_2\text{O} \rightarrow \text{pyruvate}$) are present in the ‘*Ca. M. washburnensis*’ genome and well-established in organisms that complete the isocitrate-lyase pathway. Additionally, no genes for the catalytic subunits of aldehyde or formate dehydrogenases were identified in ‘*Ca. M. washburnensis*’, which suggests that formaldehyde would be available for assimilation by the proposed pathway.

The ‘*Ca. M. washburnensis*’ genome also contained an archaeal form III-b⁴⁹ ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), as well as genes encoding the full suite of enzymes that are necessary for ribose-scavenging pathways that involve phosphoribosyl pyrophosphate (PRPP) and/or adenosine monophosphate (AMP) precursors (that is, purine nucleoside phosphorylase, ribose-1,5-phosphate isomerase and thiazole-adenylate synthase). Type III RuBisCOs have been inferred to scavenge ribose carbon (for example, nucleic acids) by incorporating CO_2 for re-entry into glycolysis and/or gluconeogenesis⁵⁰. Consequently, the incorporation of highly abundant CO_2 from Washburn Hot Springs (92% of subsurface gas¹⁹) during recycling of ribose may represent another carbon source for ‘*Ca. M. washburnensis*’ (Supplementary Fig. 10).

General pathways for central carbon metabolism were similar for ‘*Ca. M. washburnensis*’, ‘*Ca. K. cryptofilum WS*’ and ‘*Ca. K. cryptofilum OPF8*’ (Fig. 4). Each of these korarchaeotes contain a glycolysis pathway that lacks genes for the reversible transformation between glucose and glucose-6-phosphate. All three populations had an incomplete tricarboxylic acid (TCA) cycle, which lacked malate dehydrogenase, citrate synthase and aconitate hydratase. However, each organism exhibited a potential ‘shortcut’ for conversion between oxaloacetate and α -ketoglutarate through glutamate and aspartate cycling, which utilizes glutamate synthases and aspartate transaminases. This partial TCA cycle has also been observed in recently described Bathyarchaeota⁵ and Verstraetearchaeota⁶, and suggests that these organisms may use the TCA cycle primarily to generate precursors for biosynthesis. As is common in archaea, these Korarchaeota exhibited several avenues for the incorporation of amino acid carbon into major metabolic pathways; entry points include the production of oxaloacetate or α -ketoglutarate from aspartate, glutamate and glutamine, or possible production of phosphoenolpyruvate from leucine, isoleucine, valine, phenylalanine or tyrosine. In conjunction with the presence of complete membrane transport systems for branched-chain amino acids, this suggests that the Korarchaeota take up and utilize protein-degradation products from the environment.

In summary, genes associated with methane metabolism and sulfur reduction were detected in one of the most deeply rooted archaeal phyla, the Korarchaeota. Our findings support the hypotheses that methane metabolism was an early energy conservation strategy in archaea and, more specifically, that the Korarchaeota served an important role in the distribution of methane- and sulfur-metabolizing proteins in this domain. These observations of extant relatives of ancient archaea influence considerations of early evolution, particularly with respect to sources of energy and carbon in geothermal environments.

Based on data presented here, we propose the name *Methanodesulfokores washburnensis* (L. n. *methanum*, methane; L. pref. *de*, from; L. n. *sulfo*, sulfur; N.L. pref. *Methanodesulfo*-, methane metabolizing and dissimilatory sulfur-reducing, used to characterize a microorganism that participates in the production and/or oxidation of methane and the dissimilatory reduction of sulfite; Gr. n. *kore*, young woman, the previously selected root for phylum Korarchaeota;²⁰ N.L. masc. adj. *washburnensis*, pertaining to Washburn Hot Springs in Yellowstone National Park, USA). This organism contains genes that are necessary for methanogenesis from methanol and hydrogen, anaerobic oxidation of methane with sulfite and/or sulfite reduction with hydrogen. Genome sequences were obtained from Washburn Hot Springs in Yellowstone National Park, USA, with a temperature of 65–70 °C, a pH of 6.4 and elevated concentrations of carbon dioxide, methane, hydrogen, hydrogen sulfide and sulfate.

Methods

Site selection and sample collection. Washburn Hot Springs (Yellowstone National Park) is a circumneutral (pH 6.4), highly sulfidic and anoxic geothermal pool¹⁸ that contains high concentrations of methane, hydrogen and carbon dioxide. Recent work recovered divergent methane metabolism genes from these 65–70 °C sediments¹⁹. Sediments for DNA extraction were collected (4 October 2012) in 50-ml Falcon tubes, immediately placed on dry ice and then transferred to a –80 °C freezer within 12 h.

DNA extraction, sequencing and metagenome analysis. DNA was extracted from sediments from Washburn Hot Springs with the FastDNA Spin Kit for Soil (MP Biomedicals). DNA was eluted in sterile water and frozen at –80 °C until shipment to the Joint Genome Institute (JGI), US Department of Energy, for sequencing on the Illumina HiSeq platform. Initial analyses of the metagenome from Washburn Hot Springs were performed in accordance with standard JGI protocols (<https://img.jgi.doe.gov/>).

Quality-filtered short reads were assembled with SPAdes⁵¹ (v.3.10) according to the JGI analysis pipeline. Four additional assemblies were also analysed for comparison. These included the original JGI ‘manual’ assembly, our own local assembly using SPAdes v.3.11 and Megahit⁵² with default and sensitive parameters. Tetranucleotide frequencies were calculated for DNA scaffolds with a minimum sequence length of 5 kb, chopped to 20 kb and with an overlap size of 10 kb. Tetranucleotide frequency results were plotted two-dimensionally with tSNE²² using a cluster radius that discriminated discrete sequence clusters (Supplementary Fig. 4). Clustered scaffolds were imported into anvio⁵³ (v.3) and further characterized. The curated genome assemblies have been described as ‘metagenome-assembled genomes’ and represent the average genomic content of microbial populations resolved at close to the species level. Coverage values were not used for sequence clustering purposes because of variation in coverage across highly similar sequences that were shown to belong to the same population (Supplementary Discussion).

Genome analysis, pangenomics and metabolic reconstruction. Anvio⁵³ used Prodigal⁵⁴ to identify open reading frames in contigs, HMMER⁵⁵ to search for archaeal single-copy genes⁵³ for estimating genome completeness and redundancy and the BLAST software suite⁵⁶ to assign functions through NCBI’s Clusters of Orthologous Genes (COGs)⁵⁷. Anvio also calculated coverage estimates for individual scaffolds and population genomes along with other statistics (that is, N50 and GC content). Anvio was also used to analyse microdiversity patterns through single-nucleotide variants and for pangenomic analyses⁵⁸. In brief, anvio identifies gene clusters by computing amino acid sequence identity scores between every open reading frame in every genome using the BLASTp program⁵⁶, then uses the MCL algorithm⁵⁹ to demarcate individual gene clusters in search results through graph partitioning and finally visualizes the distribution of gene clusters across genomes. In addition, functional properties of each population genome were examined with the KEGG database⁶⁰ using the IMG system from the JGI. KEGG pathways were used to create metabolic reconstructions. After initial KEGG analysis, certain functions were further analysed using gene neighbourhood and blast comparisons to gene and protein databases (for example, COG⁵⁷ and PFAM⁶¹ databases). Average genome nucleotide identities (ANIb and ANIm) and tetranucleotide frequency regression analysis were performed with JSpecies v.1.2.1⁶². Average genome amino acid identities were calculated by averaging the identities of pairwise protein homologues identified using BLASTp homology cut-offs ($\geq 29\%$ amino acid identity and $\geq 60\%$ protein length).

Phylogenetic analyses. Amino acid sequences from 56 archaeal clusters of orthologous genes (arCOGs; Supplementary Table 4) were aligned, concatenated and used in the phylogenetic analysis of korarchaeotal genomes compared to other major archaeal phyla, as well as to representatives from the Bacteria and

Eukarya²⁹. A similar set of proteins was used by other groups²⁷. To ensure a robust alignment, we only included taxa that contained at least 50 out of 56 arCOGs (approximately 90%). Sequences of individual proteins were aligned with MAFFT-L-INS-i⁶³, visually inspected and adjusted, and trimmed by trimAL⁶⁴ using a 50% gap threshold. Trimmed alignments (15,526 residues) were subjected to Bayesian inference analysis (MrBayes v.3.2.5⁶⁵) in increments of 1 million generations until the standard deviation of split frequencies was <0.01. ProtTest 3⁶⁶ analysis revealed that an LG model with empirical amino acid frequencies and invgamma rates was the most optimal. Nevertheless, we tested multiple combinations of parameters (2–5 million generations, 0.1–0.25 burn-in fraction, four and eight parallel chains, gamma and invgamma models for rate variation, four and eight rate categories for the gamma distribution and temperature factors ranging from 0.075 to 0.15). Figure 2a shows a tree after 2 million generations (0.25 burn-in, two runs with eight parallel chains and eight rate categories for the gamma distribution). Maximum likelihood trees were computed using RAXML⁶⁷ (v.8.2.0) in regular bootstrapping mode with 1,000 replicates (parameters determined by ProtTest 3) and using IQ-TREE (v.1.6.7.2)⁶⁸ with the best parameters determined by internal model selection (LG + F + R10). IQ-TREE reconstructions were done both in non-parametric bootstrap mode (100 replicates) and with ultrafast bootstrap approximation (1,000 replicates). IQ-TREE was also used for posterior tree topology tests (Supplementary Table 5). A comparison of three-domain trees resulting from different programs and their parameter selections is shown in Supplementary Fig. 5. Newick versions of all phylogenomic trees as well as the master concatenated protein alignment used for all phylogenomic trees are available as Supplementary Data 1–13. Tree topologies were locally constrained to infer ancestral reconstructions (MrBayes) for select nodes and ancestral sequences were derived from states with highest probabilities.

16S rRNA sequences were retrieved from IMG for each korarchaeotal genome, aligned to the Silva database (v.132) with SINA⁶⁹ and compared to other (NCBI, Genbank) near-full-length (>1,200 nucleotides) 16S rRNA sequences from environmental samples. Alignments were manually inspected in Arb (v.6.0.6)⁷⁰ and a maximum likelihood phylogenetic tree was constructed using the inverse gamma rate substitution model only on sequences with >1,000 nucleotide positions. Subsequently, 1,000 iterations were performed for calculations of bootstrap support.

Translated amino acid sequences for korarchaeotal McrA and DsrAB were retrieved from the IMG/MG database. McrA and DsrAB sequences were aligned with MAFFT-L-INS-i using previously published databases that were updated to include recently described sequence information^{19,35}. Sequence alignments were visually examined and manually edited in Arb⁷⁰. MrBayes was used for Bayesian analysis of long fragment sequences until the standard deviation in split frequencies was below 0.01. ProtTest 3⁶⁶ analysis revealed that an LG model with empirical amino acid frequencies and gamma (DsrAB) or invgamma (McrA) rates were the most optimal. An additional tree comparing *mcrA* nucleotide sequences (Supplementary Fig. 7) was calculated with MrBayes until the standard deviation in split frequencies was below 0.01 (rates = gamma, eight rate categories for gamma distribution).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Metagenome sequences used in this study are available on IMG/M (DOE-Joint Genome Institute) under genome identifier [3300005860](https://doi.org/10.25554/IMG.M.2018050860). Metagenome-assembled genomes are available under NCBI BioProject accession number [PRJNA492148](https://doi.org/10.6026/97024492148). Access to the tSNE-based nucleotide frequency analysis algorithm can be obtained from the Center for Genomics and Bioinformatics at Indiana University. Newick files for three-domain and archaea-only phylogenomic trees are available as Supplementary Data 1–13.

Received: 1 June 2018; Accepted: 7 January 2019;

Published online: 04 March 2019

References

- Woese, C. R. Bacterial evolution. *Microbiol. Rev.* **51**, 221–271 (1987).
- Ueno, Y., Yamada, K., Yoshida, N., Maruyama, S. & Isozaki, Y. Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. *Nature* **440**, 516–519 (2006).
- Shen, Y., Buick, R. & Canfield, D. E. Isotopic evidence for microbial sulphate reduction in the early Archaean era. *Nature* **410**, 77–81 (2001).
- Dhillon, A., Goswami, S., Riley, M., Teske, A. & Sogin, M. Domain evolution and functional diversification of sulfite reductases. *Astrobiology* **5**, 18–29 (2005).
- Evans, P. N. et al. Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* **350**, 434–438 (2015).
- Vanwonterghem, I. et al. Methylophilic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat. Microbiol.* **1**, 16170 (2016).
- Laso-Perez, R. et al. Thermophilic archaea activate butane via alkyl-coenzyme M formation. *Nature* **539**, 396–401 (2016).
- Spang, A., Caceres, E. F. & Ettema, T. J. G. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* **357**, eaaf3883 (2017).
- Anantharaman, K. et al. Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. *ISME J.* **12**, 1715–1728 (2018).
- Scheller, S., Goenrich, M., Boecher, R., Thauer, R. K. & Jaun, B. The key nickel enzyme of methanogenesis catalyses the anaerobic oxidation of methane. *Nature* **465**, 606–608 (2010).
- Loy, A. et al. Reverse dissimilatory sulfite reductase as phylogenetic marker for a subgroup of sulfur-oxidizing prokaryotes. *Environ. Microbiol.* **11**, 289–299 (2009).
- Hinrichs, K. U., Hayes, J. M., Sylva, S. P., Brewer, P. G. & DeLong, E. F. Methane-consuming archaeobacteria in marine sediments. *Nature* **398**, 802–805 (1999).
- Boetius, A. et al. A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* **407**, 623–626 (2000).
- McGlynn, S. E., Chadwick, G. L., Kempes, C. P. & Orphan, V. J. Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature* **526**, 531–535 (2015).
- Milucka, J. et al. Zero-valent sulphur is a key intermediate in marine methane oxidation. *Nature* **491**, 541–546 (2012).
- Wegener, G., Krukenberg, V., Riedel, D., Tegetmeyer, H. E. & Boetius, A. Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria. *Nature* **526**, 587–590 (2015).
- Elkins, J. G. et al. A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl Acad. Sci. USA* **105**, 8102–8107 (2008).
- Inskeep, W. P. et al. Phylogenetic and functional analysis of metagenome sequence from high-temperature archaeal habitats demonstrate linkages between metabolic potential and geochemistry. *Front. Microbiol.* **4**, 95 (2013).
- McKay, L. J., Hatzepichler, R., Inskeep, W. P. & Fields, M. W. Occurrence and expression of novel methyl-coenzyme M reductase gene (*mcrA*) variants in hot spring sediments. *Sci. Rep.* **7**, 7252 (2017).
- Barns, S. M., Delwiche, C. F., Palmer, J. D. & Pace, N. R. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc. Natl Acad. Sci. USA* **93**, 9188–9193 (1996).
- Bowers, R. M. et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
- van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
- Dick, G. J. et al. Community-wide analysis of microbial genome sequence signatures. *Genome. Biol.* **10**, R85 (2009).
- Slesarev, A. I. et al. The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. *Proc. Natl Acad. Sci. USA* **99**, 4644–4649 (2002).
- Huber, G., Spinnler, C., Gambacorta, A. & Stetter, K. O. *Metallosphaera sedula* gen. and sp. nov. represents a new genus of aerobic, metal-mobilizing, thermoacidophilic Archaeobacteria. *Syst. Appl. Microbiol.* **12**, 38–47 (1989).
- Medini, D., Donati, C., Tettelin, H., Massignani, V. & Rappuoli, R. The microbial pan-genome. *Curr. Opin. Genet. Dev.* **15**, 589–594 (2005).
- Spang, A. et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179 (2015).
- Zaremba-Niedzwiedzka, K. et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).
- Jay, Z. J. et al. Marsarchaeota are an aerobic archaeal lineage abundant in geothermal iron oxide microbial mats. *Nat. Microbiol.* **3**, 732–740 (2018).
- Da Cunha, V., Gaia, M., Gabelle, D., Nasir, A. & Forterre, P. Lokiarchaeae are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLoS Genet.* **13**, e1006810 (2017).
- Dridi, B., Fardeau, M. L., Ollivier, B., Raoult, D. & Drancourt, M. *Methanomassiliicoccus luminyensis* gen. nov., sp. nov., a methanogenic archaeon isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* **62**, 1902–1907 (2012).
- Lever, M. A. & Teske, A. P. Diversity of methane-cycling archaea in hydrothermal sediment investigated by general and group-specific PCR primers. *Appl. Environ. Microbiol.* **81**, 1426–1441 (2015).
- Rinke, C. et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).
- Hua, Z. S. et al. Genomic inference of the metabolism and evolution of the archaeal phylum Aigarchaeota. *Nat. Commun.* **9**, 2832 (2018).
- Müller, A. L., Kjeldsen, K. U., Rattai, T., Pester, M. & Loy, A. Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi)sulfite reductases. *ISME J.* **9**, 1152–1165 (2015).
- Rabus, R. et al. A post-genomic view of the ecophysiology, catabolism and biotechnological relevance of sulphate-reducing prokaryotes. *Adv. Microb. Physiol.* **66**, 55–321 (2015).
- Borrel, G. et al. Wide diversity of methane and short-chain alkane metabolisms in uncultured archaea. *Nat. Microbiol.* <https://doi.org/10.1038/s41564-019-0363-3> (2019).

38. Liu, Y. & Whitman, W. B. Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Ann. NY Acad. Sci.* **1125**, 171–189 (2008).
39. Thauer, R. K., Kaster, A. K., Seedorf, H., Buckel, W. & Hedderich, R. Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat. Rev. Microbiol.* **6**, 579–591 (2008).
40. Wagner, T., Koch, J., Ermler, U. & Shima, S. Methanogenic heterodisulfide reductase (HdrABC-MvhAGD) uses two noncubane [4Fe-4S] clusters for reduction. *Science* **357**, 699–703 (2017).
41. Lang, K. et al. New mode of energy metabolism in the seventh order of methanogens as revealed by comparative genome analysis of “*Candidatus* methanoplasma termitum”. *Appl. Environ. Microbiol.* **81**, 1338–1352 (2015).
42. Amend, J. P. & Shock, E. L. Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria. *FEMS Microbiol. Rev.* **25**, 175–243 (2001).
43. Hedderich, R. et al. Anaerobic respiration with elemental sulfur and with disulfides. *FEMS Microbiol. Rev.* **22**, 353–381 (1998).
44. Pereira, I. A. et al. A comparative genomic analysis of energy metabolism in sulfate reducing bacteria and archaea. *Front. Microbiol.* **2**, 69 (2011).
45. Dong, M. et al. In vitro methanol production from methyl coenzyme M using the *Methanosarcina barkeri* MtaABC protein complex. *Biotechnol. Prog.* **33**, 1243–1249 (2017).
46. Haroon, M. F. et al. Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* **500**, 567–570 (2013).
47. Welander, P. V. & Metcalf, W. W. Mutagenesis of the C1 oxidation pathway in *Methanosarcina barkeri*: new insights into the Mtr/Mer bypass pathway. *J. Bacteriol.* **190**, 1928–1936 (2008).
48. Meyerdierks, A. et al. Metagenome and mRNA expression analyses of anaerobic methanotrophic archaea of the ANME-1 group. *Environ. Microbiol.* **12**, 422–439 (2010).
49. Kono, T. et al. A RuBisCO-mediated carbon metabolic pathway in methanogenic archaea. *Nat. Commun.* **8**, 14007 (2017).
50. Sato, T., Atomi, H. & Imanaka, T. Archaeal type III RuBisCOs function in a pathway for AMP metabolism. *Science* **315**, 1003–1006 (2007).
51. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
52. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
53. Eren, A. M. et al. AnVio: an advanced analysis and visualization platform for ‘omics data. *PeerJ* **3**, e1319 (2015).
54. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
55. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
56. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
57. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33–36 (2000).
58. Delmont, T. O. & Eren, A. M. Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ* **6**, e4320 (2018).
59. van Dongen, S. & Abreu-Goodger, C. Using MCL to extract clusters from networks. *Methods Mol. Biol.* **804**, 281–295 (2012).
60. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).
61. Finn, R. D. et al. Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–D230 (2014).
62. Richter, M. & Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl Acad. Sci. USA* **106**, 19126–19131 (2009).
63. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
64. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
65. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
66. Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104–2105 (2005).
67. Stamatakis, A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
68. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
69. Pruesse, E., Peplies, J. & Glockner, F. O. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**, 1823–1829 (2012).
70. Ludwig, W. et al. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).

Acknowledgements

The authors appreciate support from the NASA Postdoctoral Program through the NASA Astrobiology Institute (L.J.M.), the Montana Agricultural Experiment Station (project 911300; W.P.I.), the National Institutes of Health IDeA Program (COBRE grant GM110732; M.D.), the NSF Integrative Graduate Education and Research Traineeship Program (NSF DGE 0654336; Z.J.J.), the W. M. Keck Foundation (L.J.M., M.W.F., K.B.K. and W.P.I.), NSF 1736255 (L.J.M. and M.W.F.) and by the US Department of Energy—Ecosystems and Networks Integrated with Genes and Molecular Assemblies (contract number DE-AC02-05CH11231; M.W.F.). Metagenome sequencing of DNA from Washburn Hot Springs was conducted at the DOE-Joint Genome Institute under the Community Sequencing Program (CSP 701; W.P.I.). Computations were performed on the Hyalite High-Performance Computing System, operated and supported by MSU's Information Technology Center. We thank the NSF EarthCube ECOGEO RCN (1440066) for metagenomics analysis support and individuals involved in the 2016 ‘Omics Workshop’ at the University of Hawaii; V. Krukenberg, G. Borrel, S. Gribaldo, R. Hatzepichler, M. Lever and A. Teske for helpful discussions and J. Beam for assistance with sample collection and processing.

Author contributions

W.P.I. and L.J.M. designed the investigation. W.P.I. and Z.J.J. collected samples and performed initial metagenome processing. L.J.M. and M.D. performed clustering, coverage analyses and phylogenetic analyses. L.J.M., M.W.F. and W.P.I. built metabolic reconstructions. M.D. performed additional metagenome assemblies, phylogenomic, emergent self-organizing maps and sequence reconstruction analyses. T.O.D. and A.M.E. performed pangenomics and microdiversity analyses. M.D., K.B.K. and L.J.M. performed phylogenetic analyses of McrA and DsrAB protein sequences. K.B.K. and L.J.M. performed phylogenetic analysis of *mcrA* genes. D.B.R. and M.D. assisted with nucleotide frequency analyses. L.J.M., W.P.I. and M.D. wrote the manuscript and responses to reviewer comments. All authors contributed to manuscript editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41564-019-0362-4>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to L.J.M. or W.P.I.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☒ ☐ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☐ ☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- ☒ ☐ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

No software was used for data collection.

Data analysis

The t-stochastic neighbor-embedding algorithm for tetranucleotide frequency clustering was developed by Dr. Douglas B. Rusch at Indiana University.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Metagenomic sequences used in this study are available on IMG/M (DOE-Joint Genome Institute, Walnut Creek, CA) under the genome ID 3300005860. Metagenome-assembled genomes are available under NCBI BioProject ID PRJNA492148.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This study is based on a single metagenome. This metagenome was binned and two genome bins (i.e., populations) of interest were extracted for analysis.
Research sample	Sediments from Washburn Hot Spring.
Sampling strategy	Sediments were collected from the bank of Washburn Hot Spring via sterile scoop and placed in 50-mL sterile tubes. One sample was chosen randomly for downstream metagenomic sequencing.
Data collection	The Joint Genome Institute of the US Department of Energy sequenced the metagenomic DNA and reported the data on their online platform: Integrated Microbial Genomes & Microbiomes.
Timing and spatial scale	This sediment sample was collected on October 4th, 2012.
Data exclusions	This manuscript reports on two metagenome assembled genomes from a larger metagenome. Only the two genomes of interest are reported here as they belong to phylum Korarchaeota, which was the taxonomic focus of our research.
Reproducibility	Five metagenomic assemblies and binning strategies were performed to confirm that the genes of interest (i.e., sulfur and methane genes) repeatedly fell within the novel Korarchaeotal genome, Ca. Methanodesulfokores washburnensis.
Randomization	N/A: This manuscript is based on a single metagenome from one environmental sample.
Blinding	N/A: This manuscript is based on a single metagenome from one environmental sample.
Did the study involve field work?	☒ Yes ☐ No

Field work, collection and transport

Field conditions	Partly cloudy, 14C.
Location	The largest pool at Washburn Hot Springs in Yellowstone National Park. 44.376, -110.69. Elevation: 1,883 m
Access and import/export	Samples were collected in accordance with guidelines set forth the the US National Park Service under research permit #: YELL-2012-SCI-5068
Disturbance	No disturbance was caused.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involvement in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involvement in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about [availability of materials](#)

Obtaining unique materials	Sediments are readily available from Washburn Hot Springs but sampling requires permitted access from the US National Park Service.
----------------------------	---