**https://www.ibtimes.co.uk/god-underworld-microbes-hadesarchaea-discovered-living-toxic-gas-deep-below-yellowstone-hot-1543919**

**God of the underworld microbes Hadesarchaea discovered living on toxic gas deep below Yellowstone hot springs**

**(and the White Oak River in NC)**

* **

*By*[*Matt Atherton*](https://www.ibtimes.co.uk/reporters/matt-atherton)

*February 15, 2016 16:22 GMT*

*



The new species lives in the extreme environments found at Yellowstone National Park MARK RALSTON/Getty Images

A new species of microorganism has been discovered, and it lives several kilometres underneath the surface of Yellowstone hot springs, at 70 degrees Celsius. The new species – Hadesarchaea – survives without any light or oxygen, and lives off the toxic gas to humans, carbon monoxide.

The microbes were first discovered in Yellowstone National Park, and in a North Carolina estuary, just 1.5m underwater; showcasing the broad range of different environments they can thrive in. The scientists report in the journal [Nature Microbiology](http://nature.com/articles/doi%3A10.1038/nmicrobiol.2016.2), that Hadesarchaea survive by changing carbon monoxide and water into carbon dioxide and hydrogen, to act as an energy source.

[Why advertise with us](https://www.ibtimes.co.uk/corporate/advertise-with-us)

"This new class of microbes are specialised for survival beneath the surface, so we called them 'Hadesarchaea', after the ancient Greek god of the underworld," said Brett Baker, lead author of the study from Uppsala University in Sweden.

Baker's colleague, Thijs Ettema, added: "The discovery of the Hadesarchaea will help us increase our understanding of the biology and lifestyle of archaea that thrive in the deep biosphere."

To find their results, the research team collected samples from two sites in 2010; White Oak River in North Carolina, and the Lower Culex Basin in Yellowstone National Park.



Samples of Hadesarchaea were collected from this hot spring, in Yellowstone National Park, WY Dan Coleman/University of Montana

Genomic analyses of both samples led the researchers to identify the organisms as members of the Archaeal kingdom – a relatively un-researched group of organisms. Archaea can be extremely broad in terms of their characteristics – even wider than the difference between humans and a rose bush.

The study team discovered that these microorganisms managed to survive in Yellowstone – without oxygen, in 70 degrees Celsius heat, and in a very alkaline environment – by converting the oxygen from carbon monoxide to carbon dioxide. This is unique to any archaea researched previously, and therefore justified a new species being created – Hadesarchaea.

"Before this essentially nothing was known about the Hadesarchaea's ecological role and what makes them so prominent throughout the world," said Jimmy Saw, co-author of the research. "The new discovery expands our knowledge of how these organisms may have adapted to the extreme conditions of the deep biosphere."

Archaea were only officially discovered in 1977, by Carl Woese. They are generally found in extreme environments, although they are a wide branching group of organisms and can be found almost anywhere.

<https://www.nature.com/articles/nmicrobiol20162>

Genomic inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea

* [Brett J. Baker](https://www.nature.com/articles/nmicrobiol20162#auth-1)
* , [Jimmy H. Saw](https://www.nature.com/articles/nmicrobiol20162#auth-2)
* , [Anders E. Lind](https://www.nature.com/articles/nmicrobiol20162#auth-3)
* , [Cassandre Sara Lazar](https://www.nature.com/articles/nmicrobiol20162%22%20%5Cl%20%22auth-4)
* , [Kai-Uwe Hinrichs](https://www.nature.com/articles/nmicrobiol20162#auth-5)
* , [Andreas P. Teske](https://www.nature.com/articles/nmicrobiol20162#auth-6)
* & [Thijs J. G. Ettema](https://www.nature.com/articles/nmicrobiol20162#auth-7)
* *Nature Microbiology* **volume1**, Article number: 16002 (2016)
* doi:10.1038/nmicrobiol.2016.2
* [Download Citation](https://www.nature.com/articles/nmicrobiol20162.ris)
	+ [Archaeal genomics](https://www.nature.com/subjects/archaea-genomics)
	+ [Ecology](https://www.nature.com/subjects/ecology)
	+ [Microbiology](https://www.nature.com/subjects/microbiology)

Received:

01 October 2015

Accepted:

08 January 2016

Published online:

15 February 2016

* [Corrigendum: 07 March 2016](https://www.nature.com/articles/nmicrobiol201632)
* [Erratum: 06 June 2016](https://www.nature.com/articles/nmicrobiol201695)

**Abstract**

The subsurface biosphere is largely unexplored and contains a broad diversity of uncultured microbes[1](https://www.nature.com/articles/nmicrobiol20162#ref1). Despite being one of the few prokaryotic lineages that is cosmopolitan in both the terrestrial and marine subsurface[2](https://www.nature.com/articles/nmicrobiol20162#ref2),​[3](https://www.nature.com/articles/nmicrobiol20162#ref3),​[4](https://www.nature.com/articles/nmicrobiol20162#ref4), the physiological and ecological roles of SAGMEG (South-African Gold Mine Miscellaneous Euryarchaeal Group) Archaea are unknown. Here, we report the metabolic capabilities of this enigmatic group as inferred from genomic reconstructions. Four high-quality (63–90% complete) genomes were obtained from White Oak River estuary and Yellowstone National Park hot spring sediment metagenomes. Phylogenomic analyses place SAGMEG Archaea as a deeply rooting sister clade of the *Thermococci*, leading us to propose the name Hadesarchaea for this new Archaeal class. With an estimated genome size of around 1.5 Mbp, the genomes of Hadesarchaea are distinctly streamlined, yet metabolically versatile. They share several physiological mechanisms with strict anaerobic Euryarchaeota. Several metabolic characteristics make them successful in the subsurface, including genes involved in CO and H2 oxidation (or H2 production), with potential coupling to nitrite reduction to ammonia (DNRA). This first glimpse into the metabolic capabilities of these cosmopolitan Archaea suggests they are mediating key geochemical processes and are specialized for survival in the subsurface biosphere.

South-African Gold Mine Miscellaneous Euryarchaeal Group (SAGMEG) Archaea thrive in a broad range of anoxic subsurface environments and temperature regimes (4 to 80 **°**C)[5](https://www.nature.com/articles/nmicrobiol20162#ref5). They were first identified in the terrestrial subsurface and have since been shown to be broadly distributed in marine sediments[2](https://www.nature.com/articles/nmicrobiol20162#ref2),[3](https://www.nature.com/articles/nmicrobiol20162#ref3). To begin to resolve their ecological roles and sample a broad diversity within the group, we reconstructed genomic bins from the White Oak River (WOR, North Carolina) estuary and from Yellowstone National Park (YNP) hot spring sediments ([Supplementary Fig. 1](https://www.nature.com/articles/nmicrobiol20162#s1)). Two genomic bins were reconstructed (using emergent self-organizing map (ESOM) mapping of tetranucleotide signatures, see [Supplementary Information](https://www.nature.com/articles/nmicrobiol20162#s1) for additional details) from the WOR estuary sediment (DG-33 and DG-33–1, ∼63% complete) from the anoxic methane-rich layer (53–54 cmbsf (centimetres below surface))[6](https://www.nature.com/articles/nmicrobiol20162#ref6). Two additional genomic bins (YNP\_N21 and YNP\_45, 80% and 90% complete, respectively) were obtained from a microaerophilic (0.45 mg per litre O2) YNP hot spring (68.8 **°**C, pH 8.6). This spring is relatively high in sulphate (25 ppm) and low in sulphide (0.27 ppm) and nitrogen species (nitrate and nitrite were below detection, ammonium was 1.25 × 10−4 ppm). All four genomic bins were manually curated (see Methods) to remove potential contaminant sequences ([Table 1](https://www.nature.com/articles/nmicrobiol20162#t1)). Based on the size and completeness of the bins we estimated the average size of all these genomes to be around 1.5 Mbp. The average gene size is relatively small (716 (DG-33-1), 688 (DG-33), 821 (bin YNP\_45) and 820 (YNP\_N21) bp) compared to the average prokaryote (∼920 bp[7](https://www.nature.com/articles/nmicrobiol20162#ref7)). These observations indicate that SAGMEG genomes have been subjected to genome streamlining, possibly as a result of nutrient limitation[8](https://www.nature.com/articles/nmicrobiol20162#ref8).

**Table 1: General characteristics of genomic bins in this study.**

[**Full size table**](https://www.nature.com/articles/nmicrobiol20162/tables/1)

Phylogenetic identification of the genomic bins based on 16S rRNA genes (excluding DG-33-1, for which no such gene was recovered) revealed that they represent distinct lineages within the SAGMEG clade ([Fig. 1](https://www.nature.com/articles/nmicrobiol20162#f1)). Both of the YNP bins are closely related to clones that were previously recovered from Obsidian Pool at YNP[5](https://www.nature.com/articles/nmicrobiol20162#ref5), while WOR bin DG-33 is similar to rRNA sequences from subseafloor sediments ([Fig. 1](https://www.nature.com/articles/nmicrobiol20162#f1)). rRNA gene phylogeny places SAGMEG Archaea as a basal lineage within the Euryarchaeota. We were not able to resolve the branching position of SAGMEG within the Euryarchaeota solely on the basis of rRNA phylogenies. The placement of the group is also inconsistent in the literature[4](https://www.nature.com/articles/nmicrobiol20162#ref4),[9](https://www.nature.com/articles/nmicrobiol20162#ref9). Therefore, to investigate the phylogenetic position of SAGMEG within the Archaea in more detail, we constructed phylogenetic trees based on concatenated ribosomal proteins ([Supplementary Fig. 2](https://www.nature.com/articles/nmicrobiol20162#s1)), and more comprehensively using 48 phylogenetically conserved markers ([Fig. 2](https://www.nature.com/articles/nmicrobiol20162#f2) and [Supplementary Fig. 3](https://www.nature.com/articles/nmicrobiol20162#s1)). These analyses revealed that SAGMEG Archaea form a distinct clade within the Euryarchaeota that represents a sister group to the class*Thermococci*. To reflect their affiliation with the subsurface biosphere, we propose that they be given a new class name, ‘Hadesarchaea’.

**Figure 1: Phylogenetic analyses of 16S rRNA genes from three of the genomic bins in relation to those from a variety of other environments.**



Bin DG-33-1 lacked a gene long enough to accurately resolve its phylogeny. Taxa labelled in blue are from the terrestrial subsurface, orange are from the oceanic subsurface, green are from deep-sea sediments, purple are from shallow aquatic sediments, and red are from terrestrial hot springs. The phylogeny was generated based on 1,682 unambiguous characters using the RAxML (maximum likelihood) method with gamma rate distribution model, in the ARB software package[47](https://www.nature.com/articles/nmicrobiol20162#ref47). Bootstrap values were generated using UFBoot[48](https://www.nature.com/articles/nmicrobiol20162#ref48) with 10,000 replicates; those >95 are designated with filled circles and those >70 are shown as open circles. The numbers of sequences present in the collapsed groups are labelled on the wedges.

[**Full size image**](https://www.nature.com/articles/nmicrobiol20162/figures/1)

**Figure 2: Phylogenomic analysis of Hadesarchaea in the Archaea.**



Bayesian phylogeny of Hadesarchaea (SAGMEG) and other representative Archaeal phyla using 48 concatenated arCOGs (a total of 9,982 unambiguous characters). The tree was constructed from two independent Phylobayes chains after running for more than 4,000 generations until convergence (maxdiff of 0.15) and the consensus tree was obtained after burn-in of the first 2,500 generations and subsampling every tenth tree. Both maximum likelihood bootstrap (top) and Bayesian posterior probability (below) values are shown. The numbers of sequences present in the collapsed groups are labelled on the wedges. Bacteria were used as an outgroup (not shown).

[**Full size image**](https://www.nature.com/articles/nmicrobiol20162/figures/2)

The first hints into the metabolic capabilities of the Hadesarchaea came from analyses of the stable isotopic composition of Archaeal cells (Bathyarchaeota, Thermoplasmatales, SAGMEG and Lokiarchaeota) in deep-sea sediments, which indicate consistently that all members of this Archaeal community incorporate organic carbon, suggesting a heterotrophic lifestyle[3](https://www.nature.com/articles/nmicrobiol20162#ref3). Comparative genomic analyses of the obtained genome bins reveal that Hadesarchaea share several carbon pathways with other Euryarchaeota members. We could not identify genes encoding methyl-CoM reductase, indicating that Hadesarchaea are probably not involved in methanogenesis in a conventional manner. However, they do have a variety of genes for carbon metabolism present in methanogens and other Euryarchaeal lineages. All four Hadesarchaeal genome bins lack most of the genes for the forward and reverse tricarboxylic acid (TCA) cycle (or the glycoxylate bypass) for the production of acetyl-CoA. Instead, they contain genes for the C1pathway (the carbon monoxide dehydrogenase pathway), which is typically used by methanogens and numerous other anaerobes for CO2fixation or production[10](https://www.nature.com/articles/nmicrobiol20162#ref10).

The YNP genomic bins (YNP\_45 and YNP\_N21) lack several genes for the CO dehydrogenase pathway, including those encoding formylmethanofuran-tetrahydromethanopterin *N*-formyltransferase (*ftr*), methenyltetrahydromethanopterin cyclohydrolase (*mch*) and methylenetetrahydromethanopterin dehydrogenase (*mtd*), while the WOR bin DG-33 encodes all three ([Fig. 3](https://www.nature.com/articles/nmicrobiol20162#f3)). However, bin YNP\_N21 has a *mch* gene, suggesting that the absence of these genes may be a result of the incomplete draft assemblies that were obtained. Alternatively, both YNP genomic bins have D-3-phosphoglycerate dehydrogenase, phosphoserine phosphatase and glycine hydroxymethyltransferase, which may be used to generate 5,10-methylene-THMPT from glycerate 3-phosphate (as illustrated for bin YNP\_45 in [Fig. 3](https://www.nature.com/articles/nmicrobiol20162#f3)). This is referred to as the glycine pathway, and is thought to be an ancestral form of carbon fixation[11](https://www.nature.com/articles/nmicrobiol20162#ref11). However, it is possible that these C1-pathway genes are used in reverse for CO2 production[12](https://www.nature.com/articles/nmicrobiol20162#ref12).

**Figure 3: Diagram of metabolic capabilities of Hadesarchaea.**



Black arrows represent genes that were identified in bin YNP\_45 and YNP\_N21. Grey arrows are those that were not found in the YNP bins, but are present in WOR bin DG-33-1. DNRA, Denitrification and dissimilatory nitrite reduction to ammonia; FDH, format dehydrogenase; Hdr, membrane-type heterodisulphide hydrogenases; F420, coenzyme F420; V, complex V ATPase; COX, carbon-monoxide dehydrogenase; Ni,FeIII, Ni,Fe-dependent hydrogenases; Nit, nitrite reductase; Hyd, cytoplasmic hydrogenases; Nuo, NADH dehydrogenases; Sdh, succinate dehydrogenases; Znu, zinc transporter; Liv, a branched amino-acid transporter; Mal and Msm, monosaccharide transporters; PST, phosphate transporter.

[**Full size image**](https://www.nature.com/articles/nmicrobiol20162/figures/3)

The most complete genome bins YNP\_45 and YNP\_N21 lack the key glycolytic hexokinase, glucokinase (*pfkC*) and pyruvate kinase (*pyk*) genes, resulting in an incomplete glycolytic pathway. They do have genes encoding pyruvate:ferredoxin oxidoreductases (*porA*) and fructose-1,6-bisphosphatases (*fbp*), indicating that these Archaea contain a complete pathway for gluconeogenesis. With the exception of bin DG-33, all of the genomic bins contain a gene encoding a ribulose biphosphate carboxylase (RubisCO) homologue ([Fig. 3](https://www.nature.com/articles/nmicrobiol20162#f3)). To confirm that the genes are related to known RubisCOs, we generated a phylogenetic tree of these proteins ([Supplementary Fig. 4](https://www.nature.com/articles/nmicrobiol20162#s1)). This analysis revealed that the Hadesarchaeal RubisCOs form a distinct clade within Archaeal RubisCO type III proteins, which were recently shown to be involved in AMP metabolism in other Euryarchaea[13](https://www.nature.com/articles/nmicrobiol20162#ref13). However, the genomes of previously described Archaea lacked Calvin–Benson–Bassham (CBB) cycles, which is the pathway for carbon fixation using RubisCO[13](https://www.nature.com/articles/nmicrobiol20162#ref13). Interestingly, YNP\_45 and YNP\_N21 have near-complete CBB pathways, only lacking two CBB genes: phosphoribulokinase (*prk*) and fructose-1,6-bisphosphatase II (*glpX*). The ability to fix inorganic carbon opens up a new source of carbon in the deep subsurface, where dissolved organic carbon has been shown to be extremely low[14](https://www.nature.com/articles/nmicrobiol20162#ref14),[15](https://www.nature.com/articles/nmicrobiol20162#ref15) but inorganic carbon is readily available[16](https://www.nature.com/articles/nmicrobiol20162#ref16).

Acetate and hydrogen constitute an important link between respiration and fermentative processes in the subsurface[17](https://www.nature.com/articles/nmicrobiol20162#ref17). Hydrogen is believed to be a primary substrate for microbial growth in the subsurface and in hot springs[5](https://www.nature.com/articles/nmicrobiol20162#ref5),[18](https://www.nature.com/articles/nmicrobiol20162#ref18),[19](https://www.nature.com/articles/nmicrobiol20162#ref19). Hadesarchaea constitute a significant proportion of microbial communities in hydrogen-rich hot springs at YNP[5](https://www.nature.com/articles/nmicrobiol20162#ref5), and all genome bins contain genes encoding acetyl-CoA synthetase for the conversion of acetate into acetyl-CoA. In addition, these bins also contain genes encoding several hydrogenases, including Ni-dependent hydrogenase, Ni,Fe hydrogenase and coenzyme F420 hydrogenase. Ni,Fe hydrogenases are capable of both H2 consumption and production, and all the Hadesarchaea genomic bins contain the genes encoding the small and large subunits, G-components, as well as maturation factors. It is possible that the Ni,Fe hydrogenases, F420-hydrogen dehydrogenase and cytochrome b5 genes (also present in all four genomic bins) are used to derive energy by electron transfer from H2, as demonstrated in other Euryarchaeota[20](https://www.nature.com/articles/nmicrobiol20162#ref20). We did not identify any Fe-hydrogenase encoding genes, which are generally assumed to be associated with H2 production and have only been identified in bacteria[20](https://www.nature.com/articles/nmicrobiol20162#ref20), in the Hadesarchaeal genome bins.

Interestingly, some of these hydrogenases have been demonstrated to be involved in sulphur reduction (sulphidogenesis)[21](https://www.nature.com/articles/nmicrobiol20162#ref21) and sulphide oxidation[22](https://www.nature.com/articles/nmicrobiol20162#ref22). The YNP genomic bins YNP\_45 and YNP\_N21 contain phylogenetically distinct genes that are related to sulphohydrogenase ([Supplementary Fig. 5](https://www.nature.com/articles/nmicrobiol20162#s1)). They also encode the *nuoB*, *nuoH* and *ndhI*genes of the NADH:quinone oxidoreductase, which are essential for transferring electrons from NADH to S0 (ref. [21](https://www.nature.com/articles/nmicrobiol20162#ref21)). As these two Archaeal bins are associated with a hot sulphur spring environment, it seems likely that these genes have a role in sulphur cycling. Cultured species in the phylogenetically related sister group *Thermococci* are commonly shown to be reducing S0 to sulphide[23](https://www.nature.com/articles/nmicrobiol20162#ref23).

Carbon monoxide is a common gas in hydrothermal environments[24](https://www.nature.com/articles/nmicrobiol20162#ref24) and is produced by thermal decomposition of organic matter or as a side product of some anaerobes[25](https://www.nature.com/articles/nmicrobiol20162#ref25). It is therefore likely to occur widely in sediments and the subsurface. It has been estimated that after sulphate reduction, CO oxidation is the most energetically favourable anaerobic reaction in the deep terrestrial subsurface[15](https://www.nature.com/articles/nmicrobiol20162#ref15). In addition to carbon monoxide (*cox*) genes, the YNP\_45 Hadesarchaea bin also has distinct genes for catalytic nickel-containing CO dehydrogenase (CooS) and the iron sulphur subunits (CooF), suggesting that they are able to oxidize CO and H2O to CO2 and H2 as an energy source. The other bins are missing *cooS*, but have genes with sequence similarity to *cooF* present in *Thermococcus* spp. Phylogenetic analyses of these proteins show they are affiliated with ‘clade E’ ([Supplementary Fig. 6](https://www.nature.com/articles/nmicrobiol20162#s1))[12](https://www.nature.com/articles/nmicrobiol20162#ref12) and related to other methanogenic Euryarchaeota. The CooF sequences of *Thermococcus*spp, which also belong to this clade, have the ability to couple anaerobic CO oxidation to hydrogen production using *cooS* genes[26](https://www.nature.com/articles/nmicrobiol20162#ref26). If the Hadesarchaea are deriving energy from the oxidation of CO, then the C1pathway and hydrogenases are probably producing CO2 and H2[27](https://www.nature.com/articles/nmicrobiol20162#ref27).

CO oxidation coupled to anaerobic nitrate reduction is an energetically favourable reaction[28](https://www.nature.com/articles/nmicrobiol20162#ref28) that has previously been demonstrated in bacteria[29](https://www.nature.com/articles/nmicrobiol20162#ref29). Potentially, Hadesarchaea may be coupling CO oxidation to the reduction of nitrite to ammonia. We identified a distinct nitrite reductase in bin YNP\_N21 ([Supplementary Figs 7](https://www.nature.com/articles/nmicrobiol20162#s1)). Electron acceptors such as nitrite are generally thought to be depleted in the subsurface[16](https://www.nature.com/articles/nmicrobiol20162#ref16), but it has been shown that the deep terrestrial subsurface contains considerable concentrations of nitrate that may be derived from radiolytic sources[30](https://www.nature.com/articles/nmicrobiol20162#ref30). Thus, it seems feasible that these nitrite reductases can be used to derive energy.

The current understanding of the diversity, biology and ecology of the Archaea is very limited, when considering how few of the known phyla have been cultured or genomically explored, especially considering that many of these enigmatic lineages are predominant in nature. In the current study, we have obtained the first genomic sequences of a widespread group related to Euryarchaeota[4](https://www.nature.com/articles/nmicrobiol20162#ref4), here named Hadesarchaea. Inference of their metabolic capabilities indicates that they are probably deriving energy from the oxidation of carbon monoxide, which may be coupled to H2O or nitrite reduction in the terrestrial subsurface. They contain a variety of central carbon metabolic (C1 pathway) genes found in methanogens, which may be used for carbon fixation. The reconstruction of these novel genomes expands our knowledge of the evolutionary histories and metabolic potential of these widespread Archaea. Their metabolic features encoded in relatively streamlined genomes make them prominent members of the deep dark subsurface biosphere.

**Methods**

**Sample collection and processing**

The estuary genomic libraries were obtained as described previously[31](https://www.nature.com/articles/nmicrobiol20162#ref31). Six 1 m plunger cores were collected from a water depth of ∼1.5 m in three mid-estuary locations (two cores per site) of WOR, North Carolina, in October 2010 (site 1 at 34°44.592 N, 77°07.435 W; site 2 at 34°44.482 N, 77°07.404 W; site 3 at 34°44.141 N, 77°07298 W). Cores were stored at 4 °C overnight and processed 24 h after sampling[6](https://www.nature.com/articles/nmicrobiol20162#ref6). Each core was sectioned into 2 cm intervals. From each site, one core was subsampled for geochemical analyses, and the other core was subsampled for DNA extractions. The sulphate, sulphide and methane profiles of these cores indicated a distinct sulphur and methane-cycling zonation. The sulphate-reducing zone (SRZ) started in sediment layers where sulphate was abundant and sulphide just started to accumulate (8–12 cm). The methane-producing zone (MPZ) was characterized by accumulating methane (a single sample from site 1 of this zone was taken at 52–54 cm (ref. 6). DNA was extracted from these sediments using the UltraClean Mega Soil DNA Isolation Kit (MoBio), using 6 g of sediment, and stored at −80 °C until use.

Hot spring sediment samples from YNP were collected from a hot spring in Lower Culex Basin (from a site named as 10Y13 at 44°34.23 N, 110°47.41 W) as described previously[32](https://www.nature.com/articles/nmicrobiol20162#ref32). This uncharacterized hot spring has a temperature around 70 °C and pH of 8.6, and the hot spring sediments appeared light brown in colour. Sediment samples were taken using 50 ml falcon tubes, preserved in sucrose lysis buffer and kept at −20 °C until further use. DNA used for the metagenomic library was extracted from ∼500 mg of sediment using cetyltrimethyl ammonium bromide (CTAB)/phenol/chloroform method[33](https://www.nature.com/articles/nmicrobiol20162#ref33).

**Single-cell genomics**

Cell fractions were obtained from the Yellowstone hot spring sediment sample using Nycodenz density gradient centrifugation[34](https://www.nature.com/articles/nmicrobiol20162#ref34), and sent to the Bigelow Laboratory for Ocean Sciences for flow sorting and amplification to generate single-cell amplified genomes (SAGs) as described previously[32](https://www.nature.com/articles/nmicrobiol20162#ref32). A 384-well microtitre plate was then screened for Archaea using a LightCycler 480 qPCR instrument (Roche) with SYBR Green I Master kit and utilizing an Archaea-specific primer pair, A519F (5′-CAGCMGCCGCGGTAA-3′) and A1041R (5′-GGCCATGCACCWCCTCTC-3′). A SAG positive for the Archaeal 16S rRNA gene sequence was selected (identified by sequencing its 16S rDNA using the Sanger sequencing method), a sequencing library was prepared using NexteraXT library preparation kit (Illumina), and the library was paired-end sequenced (2 × 300 bp) on a MiSeq instrument (Illumina). 16S rRNA amplification and Sanger sequencing of the multiple displacement amplication (MDA)-amplified DNA from the cell yielded one phylotype belonging to SAGMEG. After sequencing, no contaminants were detected based on BLASTn comparison of contigs with the NCBI nt database. The number of high-quality sequences generated from the MiSeq run was 3,260,839, representing a total of 0.5 Gbp of sequence data, which were then assembled with SPAdes version 3.5.0[35](https://www.nature.com/articles/nmicrobiol20162#ref35) using the following parameters: --sc --careful -k 21,33,55,75,101,125.

**Genomic assembly, binning and annotation**

Illumina sequencing adapters from the read pairs (2 × 150 bp) were removed from both YNP and WOR libraries using Scythe (<https://github.com/vsbuffalo/scythe>), and low-quality reads (*Q* < 30) were trimmed with Sickle (<https://github.com/najoshi/sickle>). Reads with three or more Ns or with average quality score of less than *Q*20 and a length <50 bps were removed. Trimmed, screened, paired-end Illumina reads from WOR were assembled using IDBA-UD[36](https://www.nature.com/articles/nmicrobiol20162#ref36) with the following parameters (--pre\_correction --mink 55 --maxk 95 --step 10 --seed\_kmer 55). To maximize assembly, reads from different sites were co-assembled. The WOR assembly was generated from high-quality reads (378,027,948, average read length 124 bp and average insert 284 bp) of site 1 (52–54 cm) from WOR as described previously[31](https://www.nature.com/articles/nmicrobiol20162#ref31).

The YNP hot spring assembly consisted of 666,428,020 high-quality paired-end Illumina reads representing a total of 86.4 Gbp of sequence data with an estimated insert size of 337 bp. Assembly was performed using the IDBA-UD assembler[36](https://www.nature.com/articles/nmicrobiol20162#ref36) with the following parameter: --maxk 124. Contigs of genes of particular interest were checked for chimaeras by looking for dips in coverage within read mappings. Initial binning of the assembled fragments was performed using tetra-nucleotide frequency signatures using 5 kbp fragments of the contigs[37](https://www.nature.com/articles/nmicrobiol20162#ref37). The ESOMs were manually delineated and curated based on clusters within the map (as shown in [Supplementary Figs 8 and 9](https://www.nature.com/articles/nmicrobiol20162#s1)). Coverage was enhanced by recruiting reads (from each individual library/sample) to scaffolds by BLASTN (bitscore >75), which was then normalized to the number of reads from each library.

The accuracy of the binning was then assessed by checking the genomic bins to which each of the 5 kbp subportions of the contigs were assigned. If the contig was >15 kbp then the contig was assigned to the bin that contained the majority (>50%) of its 5 kbp subportions. The completeness, contamination and strain heterogeneity of the genomes within bins were then estimated using CheckM[38](https://www.nature.com/articles/nmicrobiol20162#ref38). Some of the fragments were identified as contaminants based on their phylogenetic placement, and were manually removed. Closely related variants of a lineage were retained in a bin. Binning was also manually curated based on GC content, top blast hits and mate-pairings.

Examination of fragments identified as contaminants in the YNP\_N21 bin revealed that CheckM[38](https://www.nature.com/articles/nmicrobiol20162#ref38) might be inflating contamination levels. CheckM identifies multiple copies of genes and then calls the fragment they are on as contaminants. We believe in the case of bin YNP\_N21 that gene duplications are indeed apparent in the genome. For example, ornithine carbamoyltransferase was identified to be present on four different large (12–71 kb) contigs that are all clearly Archaeal and have tetranucleotide signature binning, blast taxa matching and GC content consistent with other Hadesarchaea. This suggests that these are multiple copies of that gene in the genome, and not contaminants. CheckM also identified seven copies in the bin as a gene on five different contigs; four of these hypothetical are present on two contigs (two on each of the contigs). These sequences are considerably different (only 26–35% similarity at the protein level) and have unique hits to genes in GenBank, including hypotheticals, diguanylate cyclases and dihydroorotate dehydrogenases. The estimate that this bin has 10% contamination is thus potentially overestimated. Removal of four relatively low-coverage contigs (NODE\_10\_length\_48481\_cov\_4.44764\_ID\_19, NODE\_22\_length\_22948\_cov\_6.32857\_ID\_43, NODE\_32\_length\_12588\_cov\_4.53992\_ID\_63 and NODE\_35\_length\_10923\_cov\_3.69392\_ID\_69) resulted in the reduction of the double copy markers from six to one. However, examination of these contigs revealed they are definitely Archaeal and related to other Hadesarchaea bins present in this study. We therefore decided to leave them in the bin and have flagged them as potential contaminants in the public databases.

Co-assembly of the Yellowstone SAGMEG (YNP\_N21) SAG and metagenomic bin was performed by first extracting metagenomic reads mapped against the YNP\_N21 bin contigs using bwa[39](https://www.nature.com/articles/nmicrobiol20162#ref39) and samtools[40](https://www.nature.com/articles/nmicrobiol20162#ref40). This resulted in an improved assembly compared with that obtained just from the SAG reads ([Supplementary Table 1 and Supplementary Fig. 10](https://www.nature.com/articles/nmicrobiol20162#s1)). The read coverage in contigs was calculated using bedtools[41](https://www.nature.com/articles/nmicrobiol20162#ref41). Extracted reads were then co-assembled with SAG reads using SPAdes (version 3.5.0)[35](https://www.nature.com/articles/nmicrobiol20162#ref35) with the same parameters as the single-cell genome assembly. Resulting contigs larger than 5 kbp were manually checked for contamination and removed.

Genes were called using Prodigal[42](https://www.nature.com/articles/nmicrobiol20162#ref42). Initially gene functions were assigned by comparison to the KEGG database[43](https://www.nature.com/articles/nmicrobiol20162#ref43). The called functions were finalized by comparisons with functional databases, including PFAM[44](https://www.nature.com/articles/nmicrobiol20162#ref44) and arCOG[45](https://www.nature.com/articles/nmicrobiol20162#ref45). Proteins with important ecological or geochemical relevance were selected for protein alignments using MUSCLE[46](https://www.nature.com/articles/nmicrobiol20162#ref46) and phylogenetic comparisons with top BLAST hits in NCBI.

**Phylogenetic analyses**

The 16S rRNA gene trees were generated by running the RAxML (v2.2.1) tool implemented within the ARB software package (v2.5b)[47](https://www.nature.com/articles/nmicrobiol20162#ref47). Bootstrap values were generated using UFBoot[48](https://www.nature.com/articles/nmicrobiol20162#ref48) with 10,000 replicates. The concatenated ribosomal protein tree was generated using 16 syntenic genes that have been shown to undergo limited lateral gene transfer (rpL2, 3, 4, 5, 6, 14, 15, 16, 18, 22, 24 and rpS3, 8, 10, 17, 19)[49](https://www.nature.com/articles/nmicrobiol20162#ref49). Amino acid alignments of the individual ribosomal proteins were generated using MUSCLE[46](https://www.nature.com/articles/nmicrobiol20162#ref46) and trimmed using BMGE[50](https://www.nature.com/articles/nmicrobiol20162#ref50) (with the following settings: -m BLOSUM30 –g 0.5). The curated alignments were then concatenated for phylogenetic analyses, generated by running the RAxML (v2.2.1) tool implemented within the ARB software package (v2.5b)[47](https://www.nature.com/articles/nmicrobiol20162#ref47).

A total of 48 arCOGs (Archaeal cluster of orthologous genes[45](https://www.nature.com/articles/nmicrobiol20162#ref45)) previously determined to be universally present in single copies[51](https://www.nature.com/articles/nmicrobiol20162#ref51) were identified and extracted from the four SAGMEG genome bins using PSI-BLAST[51](https://www.nature.com/articles/nmicrobiol20162#ref51). The 48 reference arCOGs and identified orthologues from the genome bins were aligned using MAFFT (mafft-linsi option)[52](https://www.nature.com/articles/nmicrobiol20162#ref52). Alignments were trimmed for poorly aligned regions and gaps using the BMGE tool[50](https://www.nature.com/articles/nmicrobiol20162#ref50) (with the following settings: -m BLOSUM30 –g 0.5) and concatenated before phylogenomic inferences were perfomed using both maximum likelihood and Bayesian methods. Maximum likelihood analysis was performed using IQ-Tree[53](https://www.nature.com/articles/nmicrobiol20162#ref53) with the C60+LG mixture model for 1,000 ultra-rapid bootstraps and Bayesian inference was carried out using Phylobayes[54](https://www.nature.com/articles/nmicrobiol20162#ref54) with the CAT+GTR+Г mixture model of amino acid evolution for about 4,000 generations until convergence (maxdiff of 0.15, burnin of 2,500 generations and subsampling every tenth tree) for two of the independent chains out of the four.