The Korarchaeota: Archaeal orphans representing an ancestral lineage of life

James G. Elkins[1,2*], Victor Kunin[3], Iain Anderson[3], Kerrie Barry[3], Eugene Goltsman[3], Alla Lapidus[3], Brian Hedlund[1], Phil Hugenholtz[3], Nikos Kyrpides[3], David Graham[4], Martin Keller[5], Gerhard Wanner[6], Paul Richardson[3], & Karl O. Stetter[1,7]

[1]Lehrstuhl für Mikrobiologie und Archaeenzentrum, Universität Regensburg, D-93053, Regensburg, Germany. [2]Diversa Corporation, San Diego, California 92121, USA. [3]Joint Genome Institute, Walnut Creek, California 94598, USA. [4]Department of Chemistry and Biochemistry, The University of Texas at Austin, Austin, Texas 78712, USA. [5]Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA. [6]Institute of Botany, LM University of Munich, D-80638, Munich, Germany. [7]Institute of Geophysics and Planetary Science, University of California, Los Angeles, California 90095, USA.

* Present address: Department of Bioengineering, University of California, San Diego, La Jolla, California, 92093, USA

Based on conserved cellular properties, all life on Earth can be grouped into different phyla which belong to the primary domains Bacteria, Archaea, and Eukarya[1]. However, tracing back their evolutionary relationships has been impeded by horizontal gene transfer and gene loss. Within the Archaea, the kingdoms Crenarchaeota and Euryarchaeota exhibit a profound divergence[2]. In order to elucidate the evolution of these two major kingdoms, representatives of more deeply diverged lineages would be required. Based on their environmental small subunit ribosomal (ss RNA) sequences, the *Korarchaeota* had been originally suggested to have an ancestral relationship to all known Archaea although this assessment has been refuted[3-5]. Here we describe the cultivation and initial

characterization of the first member of the *Korarchaeota*, highly unusual, ultra-thin filamentous cells about 0.16 µm in diameter. A complete genome sequence obtained from enrichment cultures revealed an unprecedented combination of signature genes which were thought to be characteristic of either the Crenarchaeota, Euryarchaeota, or Eukarya. Cell division appears to be mediated through a FtsZ-dependent mechanism which is highly conserved throughout the Bacteria and Euryarchaeota[6]. An rpb8 subunit of the DNA-dependent RNA polymerase was identified which is absent from other Archaea and has been described as a eukaryotic signature gene[7]. In addition, the representative organism possesses a ribosome structure typical for members of the Crenarchaeota. Based on its gene complement, this lineage likely diverged near the separation of the two major kingdoms of Archaea. Further investigations of these unique organisms may shed additional light onto the evolution of extant life.

Several environmental ss rRNA sequences have been recovered from globally dispersed hydrothermal environments which, by their initial phylogenetic analysis, were thought to comprise a novel lineage of Archaea which potentially branched before the Crenarchaeota/Euryarchaeota bifurcation[8,3]. This group was provisionally designated the *Korarchaeota* and little has been revealed regarding their gene content or physiology[4]. In order to gain a better understanding of this novel lineage, we attempted to sequence the entire genome of a member of the *Korarchaeota* from a community of hyperthermophilic microorganisms. Continuous enrichment cultures were established using a dilute organic medium and sediment samples from Obsidian Pool, Yellowstone National Park, Wyoming, USA, as a source of inoculum[9,10]. The 85° C, strictly anaerobic cultivation system supported the stable growth of several hyperthermophilic Archaea and Bacteria (Supplementary Fig. S1). Analysis of ss rRNA gene sequences amplified from the enrichment culture revealed an organism containing a sequence with 99% identity to the *korarchaeal* environmental clone, pJP27[3]. The organism harbouring

this sequence was designated OPF8 and was identified using fluorescence *in-situ* hybridization analysis[11,12]. *Korarchaeota* specific probes hybridized to ultra-thin, filamentous cells which occurred at a density ca. $1.0x10^7$ cells/ml. The filamentous organisms were approximately 0.16 μm in diameter with an average length of 15 μm although cells were observed with lengths up to 100 μm. Surprisingly, the cellular morphology reported here (Fig. 1) was unlike the rod-shaped, thicker morphotype that has been described for the pJP27 associated organism[10]. It was observed that in the presence of increasing concentrations of sodium dodecyl sulfate (SDS) [up to 1% (w/v) at 46^o C for 3 hours], the integrity of cells hybridizing to *Korarchaeota* specific probes was much greater than that of the other background organisms. We exposed planktonic, non-fixed cells from the enrichment culture to SDS [up to 1% (w/v)] and found that a significant number of filamentous organisms were remarkably resistant to lysis. The structural stability of the korarchaeal OPF8 cells in the presence of SDS allowed homogeneous cell preparations to be made by exposing the Obsidian Pool enrichment culture to 0.2% (w/v) SDS followed by several washing steps and filtration through 0.45 μm syringe filters. Analysis of PCR amplified ss rRNA genes from physically enriched cell preparations showed that over 99% of the clones sequenced (n=180) were identical to the ss rRNA gene sequence of OPF8.

DNA libraries produced from the highly enriched OPF8 cells were shotgun sequenced to obtain an intact genome at ca. 10X depth using standard finishing techniques. Genomic heterogeneity was encountered at a rate of approximately 2 single nucleotide polymorphisms per 1000 bases of sequence. The genome of OPF8 is a single, circular molecule 1590757 bp in size with an average G+C content of 49%. Our initial analysis identified 1617 protein-coding gene models of which 1098 (66%) could be assigned a putative function. Other general features of the genome are summarized in Table 1.

To reveal the potential metabolic capabilities of OPF8, we reconstructed a potential *in-silico* metabolism based on the predicted gene functions. OPF8 potentially gains carbon and energy through the fermentation of peptides and amino acids (Supplementary Tab. S1). In support of the predicted metabolism, OPF8 could be cultivated and transferred in serum bottles using a medium containing only casein as a carbon and energy source. Genes involved in the dissimilatory reduction of SO_3^{2-}, SO_4^{2-}, NO_3^-, NO_2^-, Fe^{3+}, $HCOO^-$, or O_2 could not be readily identified. However, since a large portion of the predicted protein coding genes could not be assigned a function, the potential to utilize these electron acceptors cannot be ruled out until more detailed cultivation studies are completed.

Analysis of the OPF8 genome revealed a mosaic of conserved features which are generally thought to be exclusive among the Crenarchaeota, Euryarchaeota, or Eukarya. A coding region of 110 amino acids in length was identified with a predicted pfam domain corresponding to subunit Rpb8 (pfam03870; E-value = 9×10^{-3}) of the eukaryotic DNA-dependent RNA polymerases. The putative Rpb8 homologue resides in a two-gene operon with the eukaryotic-like transcription factor, TFIIB. The Rpb8 subunit is shared among all three eukaryotic nuclear RNAP enzymes and is essential for transcription; however, no Rpb8 homologues have been identified in any archaeal genomes[13]. The putative *korarchaeal* Rpb8 displayed the greatest sequence conservation in the C-terminal portion of the peptide and retained all but one residue in the highly conserved GGLLM signature motif (Fig. 2). This motif recognizes a region of the Rpb1 subunit defined by the P.I.KP..LW.GKQ motif which is conserved in both eukaryotic and archaeal Rpb1 homologues[14]. Although it is well known that the archaeal transcriptional system is similar to that of the Eukarya[15], the Rpb8 subunit was thought to be exclusive among eukaryotes and considered a eukaryotic signature protein[7]. Functional validation of the Rpb8 homologue from OPF8 is in progress (Michael Thomm, personal communication).

A number of conserved genes are considered to be hallmark features of either of the two archaeal kingdoms. For example, the FtsZ protein, which is involved in (non-eukaryotic) cell division and structurally homologous to eukaryotic tubulins, is highly conserved throughout the Euryarchaeota[6] In contrast, *ftsZ* structural genes appear to be absent in members of the Crenarchaeota. The OPF8 genome encodes a total of 7 *ftsZ* homologues which represents the highest number of these genes in any organism. The ftsZ genes are encoded by two separate orthologous sequences and, in a 7-gene operon containing 5 *ftsZ* sequences which appear to be recently diverged paralogues. One of the FtsZ orthologues is included in a gene cluster containing *secE*, *nusG*, and several ribosomal protein genes (Fig. 4). This gene cluster is highly conserved within the euryarchaeal kingdom[17,18]. In addition to *ftsZ*, other features which are generally characteristic of the Euryarchaeota were also identified in the OPF8 strain including large and small subunits (COGs 1933 and 1911 respectively) of a DNA polymerase II (D family) and 2 histones (COG2036) containing proline tetrad motifs[19,20].

The distribution of ribosomal proteins in OPF8 is similar to that found in members of the Crenarchaeota. This was apparent by the identification of several eukaryotic/crenarchaeal ribosomal proteins which are absent in all members of the Euryarchaeota[21]. These ribosomal proteins included S30e, S25e, S26e, and L13e. Other crenarchaeal specific genes were also identified which encode an uncharacterized zinc ribbon containing protein (COG 4888), an uncharacterized conserved protein containing a coiled-coil domain (COG 5493), and actin-related proteins (COG 5277).

Since we were able to reconstruct an intact genome from the OP enrichment culture, a more in-depth phylogenetic analysis was possible based on whole-genome methods as well as conserved sets of informational genes. A suite of single-copy COGs which resist horizontal transfer[22] were initially selected for constructing phylogenetic trees. The dataset was primarily composed of ribosomal proteins and other orthologues

involved in translation (Supplementary Table S2). The tree topology generated from alignments of the concatenated COGs showed the *korarchaeal* OPF8 sequences forming a deep archaeal lineage with association to the Nanoarchaeota[23] and Crenarchaeota (Fig. 4a). Analysis of trees based on genome conservation, which combines both sequence similarity and gene content in a single measure[24], placed the OPF8 genome as an ancestor to the Crenarchaeota and thermoacidophilic Euryarchaeota (Supplementary Fig. 2). Also, the clearly separate position of OPF8 on whole-genome trees suggests that this lineage has been evolving independently for a considerable period of time, and apparently has not experienced rampant horizontal transfer with other known organisms. Informational genes involved in transcription have also been used to trace the evolutionary history of the archaeal domain[25]. A separate maximum-likelihood analysis based on concatenated sequences of archaeal, eukaryotic, and bacterial DNA-dependent RNAP subunits suggested that the *korarchaeote* OPF8, from a transcriptional perspective, bears an ancestral relationship to members of the Euryarchaeota and Crenarchaeota with 100% bootstrap support (Fig. 4b). When bacterial RNAP subunits *β, β',* and *α*, were included as an outgroup, the OPF8 RNAP sequences displayed a monophyletic relationship with eukaryotic RNAP sequences relative to other Archaea (Supplementary Fig. 3).

Improvements in cell identification and cultivation techniques, combined with robust DNA sequencing and assembly capabilities, allowed the first genome to be completely sequenced from the elusive *Korarchaeota*. The *korarchaeal* genome represented by OPF8 reflects a pattern of conserved orthologues that would be expected in a lineage that diverged early in the evolution of the Archaea. The presence of multiple Crenarchaeota- and Euryarchaeota-specific signature genes in a single

organism narrows the divergence between these two archaeal kingdoms. Also, the presence of an additional eukaryotic subunit of the RNAP further twines the evolutionary relationship between the Archaea and Eukarya. Previously, environmental ss rRNA sequences comprised the sole basis for taxonomic classification of the *Korarchaeota*. Although the complete genome sequence of the first *korarchaeote* has greatly expanded our knowledge of this group; high resolution phylogenetic analyses are still impeded by the limited availability of archaeal genomic data relative to the diverse abundance of Archaea known to exist in nature[26]. Obtaining genomic information from additional members of the *Korarchaeota*, as well as other deep-branching archaeal lineages[27], should aid in grounding a consensus taxonomic classification for these organisms.

METHODS

Cultivation and cell purification. The cultivation conditions for the Obsidian Pool enrichment culture were similar to those described previously[10,28]. For separation of *korarchaeal* cells, culture effluent from the Obsidian Pool fermentor was collected anaerobically in sterile 2 l glass bottles. The cells were washed twice with phosphate buffered saline (pH 7.2) and resuspended in a total volume of 200 ml of PBS (pH 7.2). SDS was added to a final concentration of 0.2% (w/v) and the samples were immediately mixed by inverting several times. The samples were then centrifuged at 6000 rpm for 20 min in a Beckman JA-12 rotor at 25^{o}C. Each pellet was washed 3 times with 50 ml of PBS (pH 7.2) and centrifuging as described above. After the final wash, the SDS-treated cells were heated to 85^{o}C and filtered through MILLEX HV 0.45 μm syringe filters (Millipore) in 25 ml aliquots. The filtrate was centrifuged at 6000 rpm for 30 min in a Beckman JA-12 rotor at 4^{o}C. Small white pellets were resuspended

in a total volume of 1 ml of PBS (pH 7.2). Cells were then used for microscopic analysis or DNA isolation.

Light and electron microscopy. Fluorescence *in-situ* hybridization analysis was performed as previously described[12] with the addition of fish sperm DNA (1 μg/hybridization) to reduce background fluorescence. The following Cy3 labeled probes provided optimal results: KR515-CCAGCCTTACCCTCCCCT; KR565-AGTATGCGTGGGAACCCCTC. Images were captured and processed with a Zeiss Axioplan 2 microscope and Axiocam digital imaging system and software. For EM analysis, *Korarchaeota* cell pellets were fixed in a solution containing 2.5% glutaraldehyde (EM grade) in 20 mM sodium cacodylate buffer (pH 6.5). For scanning electron microscopy drops of the fixed sample were placed onto glass slides, covered with a cover slip and rapidly frozen with liquid nitrogen. The cover slip was removed with a razor blade and the slide was immediately transferred into fixative buffer, postfixed with osmium tetroxide, dehydrated in a graded series of acetone solutions and critical-point dried from liquid CO_2, mounted on stubs, and coated with 3 nm platinum with a magnetron sputter coater. The specimens were examined with a Hitachi S-4100 field emission scanning electron microscope. For negative staining a drop of the sample at appropriate dilution was placed an a 400 mesh carbon-coated copper grid, freshly hydrophilisized by glow discharge. After incubation for 2 min, the drop was quickly removed with a pasteur pipette and the grid was air dried. The grid was stained with 2% uranium acetate and 0.01% glucose. Micrographs were taken with an EM 912 electron microscope (Zeiss) equipped with an integrated OMEGA energy filter operated in the zero loss mode.

DNA sequencing and assembly. DNA for sequencing was extracted as previously described with the elimination of freeze/thaw cycles to avoid shearing[3]. Library construction, DNA sequencing, and genome assembly were performed at the Joint

Genome Institute, Walnut Creek, CA, USA. Fosmid libraries were constructed from enrichment culture DNA using the pCC1FOS vector (Epicentre Biotechnologies). Small-insert libraries (3-5 kb inserts) based on pUC18 were generated from DNA originating from raw, filtered, and SDS treated cells. DNA sequencing was performed with BigDye Terminators v3.1 and resolved with ABI PRISM 3730 DNA sequencers (PE Applied Biosystems). The sequencing reads were base called using phred version 0.990722.g, vector trimmed using crossmatch SPS-3.57, and assembled into contigs using parallel phrap (www.phrap.org). The contigs were aligned with the OPF8 ss rRNA gene, and the ss rRNA genes from *Thermofilum pendens* and *Thermosphaera aggregans*. The resulting assembly was binned by scaffolding and depth, and by aligning binned scaffolds against the 16S rRNA sequences from the predominant archaeal genomes in the sample.

Genome analysis. An automated annotation was provided by the Computational Biology Group at Oak Ridge National Laboratory (ORNL), Oak Ridge, TN, USA. The predicted open reading frame dataset is based on a consensus between three gene modeling programs including Generation (Genome Informatics Corp.), Glimmer, and Critica (v1.05). An additional automated analysis was performed by loading the complete OPF8 genome into the Integrated Microbial Genomes (IMG) analysis tool[29]. To identify putative functions for the predicted open reading frames, the predicted gene models were compared against all non-redundant (NR) sequences deposited in the National Center for Biotechnology Information (NCBI) database using BLASTP. The gene models were also compared against the protein families (Pfam), Clusters of Orthologous Groups (COGs), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) databases for further functional information. Phylogenetic patterns and metabolic reconstruction were deduced from the functional analysis provided by the ORNL and IMG genome annotations.

Alignments and phylogenetic analysis. Construction of the concatenated ortholog tree was constructed using universally conserved single copy genes that do not participate in horizontal gene transfer as previously described[22]. We removed several COG groups since they were present in multiple copies in several genomes, or alternatively were not detected in several archaeal genomes by the gene calling and annotation procedure used in the IMG database. Concatenated RNA polymerase subunit trees were constructed by retrieving RpoA/A'/A'', RpoB/B'/B'', and RpoD orthologues from the NCBI and IMG databases using BLASTP. Sequence alignments were generated with CLUSTALW and edited manually using BioEdit v7.0.5 (Ibis Therapeutics). Poorly aligned regions were masked with 1864 amino acid positions remaining. Trees were constructed from the masked alignments using PhyML with a Jones-Taylor-Thornton model and 100 bootstrap replicates. Eight substitution rate categories were selected and the gamma distribution parameter was estimated by the model. The final trees were edited using TreeExplorer from the MEGA2 package and Adobe Illustrator CS2. A complete list of NCBI accession numbers, sequence datasets, and sequence alignments are available upon request from J. G. E.

1. Woese, C. R., Kandler, O., & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 4576-4579 (1990).

2. Gribaldo, S. & Brochier-Armanet, C. The origin and evolution of Archaea: a state of the art. *Philosophical transactions of the Royal Society of London* **361**, 1007-1022 (2006).

3. Barns, S. M., Fundyga, R. E., Jeffries, M. W., & Pace, N. R. Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 1609-1613 (1994).

4. Barns, S. M., Delwiche, C. F., Palmer, J. D., & Pace, N. R. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 9188-9193 (1996).

5. Robertson, C. E., Harris, J. K., Spear, J. R., & Pace, N. R. Phylogenetic diversity and ecology of environmental Archaea. *Current opinion in microbiology* **8**, 638-642 (2005).

6. Vaughan, S., Wickstead, B., Gull, K., & Addinall, S. G. Molecular evolution of FtsZ protein sequences encoded within the genomes of archaea, bacteria, and eukaryota. *Journal of molecular evolution* **58**, 19-29 (2004).

7. Hartman, H. & Fedorov, A. The origin of the eukaryotic cell: a genomic investigation. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 1420-1425 (2002).

8. Auchtung, T. A., Takacs-Vesbach, C. D., & Cavanaugh, C. M. 16S rRNA phylogenetic investigation of the candidate division "Korarchaeota". *Applied and environmental microbiology* **72**, 5077-5082 (2006).

9. Allen, M. B. Studies with Cyanidium caldarium, an anomalously pigmented chlorophyte. *Archiv fur Mikrobiologie* **32**, 270-277 (1959).

10. Burggraf, S., Heyder, P., & Eis, N. A pivotal Archaea group. *Nature* **385**, 780 (1997).

11. Amann, R. I., Krumholz, L., & Stahl, D. A. Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *Journal of bacteriology* **172**, 762-770 (1990).

12. Burggraf, S. et al. Identifying members of the domain Archaea with rRNA-targeted oligonucleotide probes. *Applied and environmental microbiology* **60**, 3112-3119 (1994).

13. Woychik, N. A., Liao, S. M., Kolodziej, P. A., & Young, R. A. Subunits shared by eukaryotic nuclear RNA polymerases. *Genes & development* **4**, 313-323 (1990).

14. Briand, J. F. et al. Partners of Rpb8p, a small subunit shared by yeast RNA polymerases I, II and III. *Molecular and cellular biology* **21**, 6056-6065 (2001).

15. Langer, D., Hain, J., Thuriaux, P., & Zillig, W. Transcription in archaea: similarity to that in eucarya. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 5768-5772 (1995).

16. Futterer, O. et al. Genome sequence of *Picrophilus torridus* and its implications for life around pH 0. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9091-9096 (2004).

17. Faguy, D. M. & Doolittle, W. F. Cytoskeletal proteins: the evolution of cell division. *Curr Biol* **8**, R338-341 (1998).

18. Poplawski, A., Gullbrand, B., & Bernander, R. The ftsZ gene of Haloferax mediterranei: sequence, conserved gene order, and visualization of the FtsZ ring. *Gene* **242**, 357-367 (2000).

19. Makarova, K. S. & Koonin, E. V. Evolutionary and functional genomics of the Archaea. *Current opinion in microbiology* **8**, 586-594 (2005).

20. Decanniere, K. et al. Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon *Methanothermus fervidus*. *Journal of molecular biology* **303**, 35-47 (2000).

21. Lecompte, O. et al. Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic acids research* **30**, 5382-5390 (2002).

22. Ciccarelli, F. D. et al. Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**, 1283-1287 (2006).

23. Huber, H. et al. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**, 63-67 (2002).

24. Kunin, V. et al. Measuring genome conservation across taxa: divided strains and united kingdoms. *Nucleic acids research* **33**, 616-621 (2005).

25. Brochier, C., Forterre, P., and Gribaldo, S. Archaeal phylogeny based on proteins of the transcription and translation machineries: tackling the *Methanopyrus kandleri* paradox. *Genome biology* **5**, R17 (2004).

26. Schleper, C., Jurgens, G., and Jonuscheit, M. Genomic studies of uncultivated archaea. Nature reviews 3, 479-488 (2005).

27. Takai, K. & Horikoshi, K. Genetic diversity of archaea in deep-sea hydrothermal vent environments. *Genetics* **152**, 1285-1297 (1999).

28. Haring, M. et al. Morphology and genome organization of the virus PSV of the hyperthermophilic archaeal genera Pyrobaculum and Thermoproteus: a novel virus family, the Globuloviridae. *Virology* **323**, 233-242 (2004).

29. Markowitz, V. M. et al. The integrated microbial genomes (IMG) system. *Nucleic Acids Res.* **34,** D344-348 (2006)

30. Guindon, S., Lethiec, F., Duroux, P., and Gascuel, O. PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic acids research* **33**, W557-559 (2005).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements Funding was provided to J.G.E and K.O.S. by Diversa Corporation, San Diego, California, and the The Deutsche Forschungsgemeinschaft. Support for genome sequencing and assembly was provided by the US Dept. of Energy and the JGI Community Sequencing Program. This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.

Author Information The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.O.S. (karl.stetter@biologie.uni-regensburg.de).

Figure 1 Images of *Korarchaeum sp.* OPF8 cells. a, Phase contrast image of OPF8 filaments following physical enrichment. Scale bar represents 5 μm. **b,** FISH analysis with Cy3-labeled oligonucleotide probes targeting *Korarchaeota* spp. The undulated cell shape results from drying of the specimen on agar coated slides prior to hybridization. Scale bar represents 5 μm. **c,** Scanning electron micrograph of OPF8 cells. **d,** Transmission electron micrograph of OPF8 cells displaying crystalline S-layer. Cells are flattened which increases their apparent thickness.

Figure 2 Amino acid alignment of eukaryotic Rpb8 sequences with OPF8 homologue. Pink shading indicates conservation in at least 3 out of 4 residues. Blue shading indicates similar residues. The GGLLM signature motif is underlined.

Figure 3 Conserved gene order within *ftsZ* operon in Euryarchaeota. Gene order conservation between OPF8 *ftsZ* operon and homologous regions from the *Methanothermobacter thermautotrophicus* (*M. t.*) and *Pyrococcus furiosus* (*P. f.*) genomes (rp ribosomal protein).

Figure 4 Phylogenetic analysis of *Korarchaeum* sp. OPF8 relative to members of the Archaea and Eukarya. Maximum-likelihood (ML) trees were constructed from alignments of **(a)** concatenated universally conserved single-copy COGs or, **(b)** concatenated DNA-dependent RNA polymerase subunits A/A'/A'', B/B'/B'', and D. ML Bootstrap support is shown at the corresponding node (100 replicates). The scale bar represents the number of changes per

aligned amino acid position per unit branch length. Eukaya outgroup for COG tree includes *Aspergillus fumigatus*, *Schizosaccharomyces pombe*, and *Cryptococcus neoformans*.

Table 1. General features of the OPF8 genome

Total number of bases	1590757
DNA coding number of bases	1442371
DNA G+C number of bases	779482
Total number of predicted genes	1665
Protein coding genes	1617
rRNA genes*	3
tRNA genes	45
Genes with function prediction	1098
Genes without function prediction	519
Genes w/o function with similarity	505
Genes w/o function w/o similarity	14
Genes in ortholog clusters	1524
Genes in paralog clusters	167

*16S, 23S, and 5S rRNA

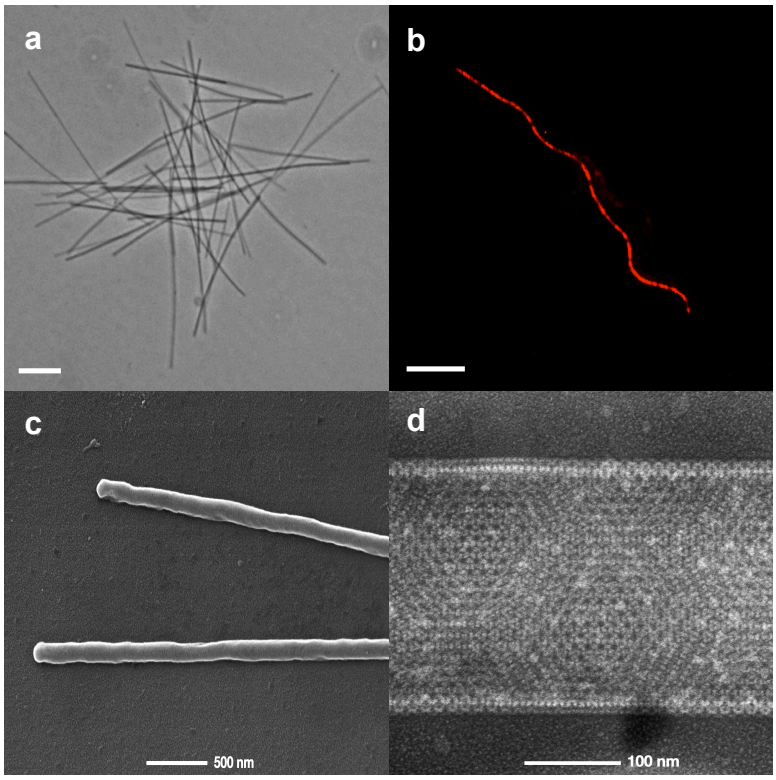

Figure 1

```
OPF8          --------LSKVRLSSVRSDPIFDINILSFESLDRSFEATLEFPKGLIDLKEGMEADLTLGE----------------
S. pombe      -MSESVLLDEIFTVTSVDKQKYQRVSRITAVSGQNDMNLTLDINSQIYPLEKDATFSLQITSNLNS-------------
P. falcipa    -MASNILFEDRFVISSVDNSKFEKVSRIKAKSTGYDAELILDVHSELFKVEEKKAIYLALQDKLLNRNDEKGWEQNENV
T. parva      MTTASCVFEDRFLVRSIDNSKFERVSRLSAKSTGFDAELLLDFNSDLLQVYNKQVLHILITNSLLPSGSDVNLSESVDI

OPF8          --EGEGDIIMKGVVYKFDDSS---RTIEISFHGLLMKLTYSKAAPFKLSQGEEVYLTIKFA-------
S. pombe      DLKEAADYIMYGKVYRVEEAKDEKVSVYVSFGGLLMAIEGSHRKLYRLSLD-HVYLLLRR--------
P. falcipa    --LNNIEYIMSGRIFKFEELSSERRTVYASFGGLMMALTTDKQFIGDLEIDMKIYLLTKNIDIERTRE
T. parva      SLLGDYEYAMYGKIFKFEEVSSETRTIYASFGGLLMSLTADKQVVADLDLDMKIYLLVRRISSNR---
```

Figure 2

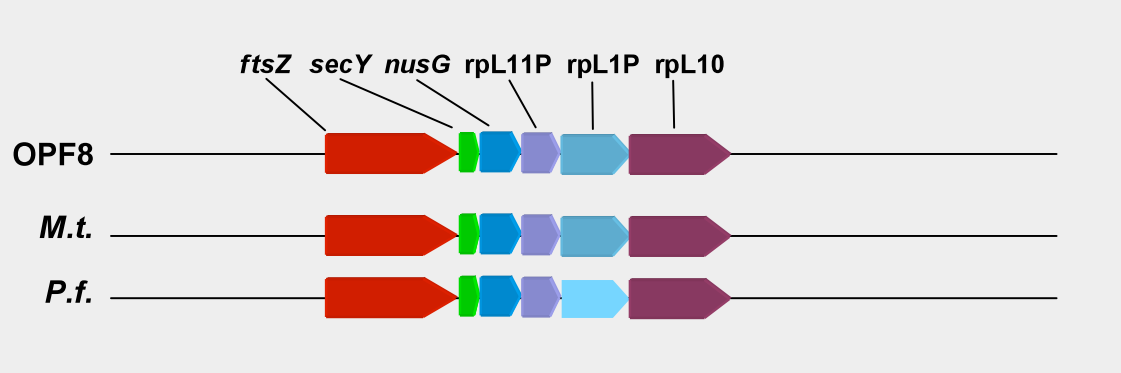

Figure 3

.

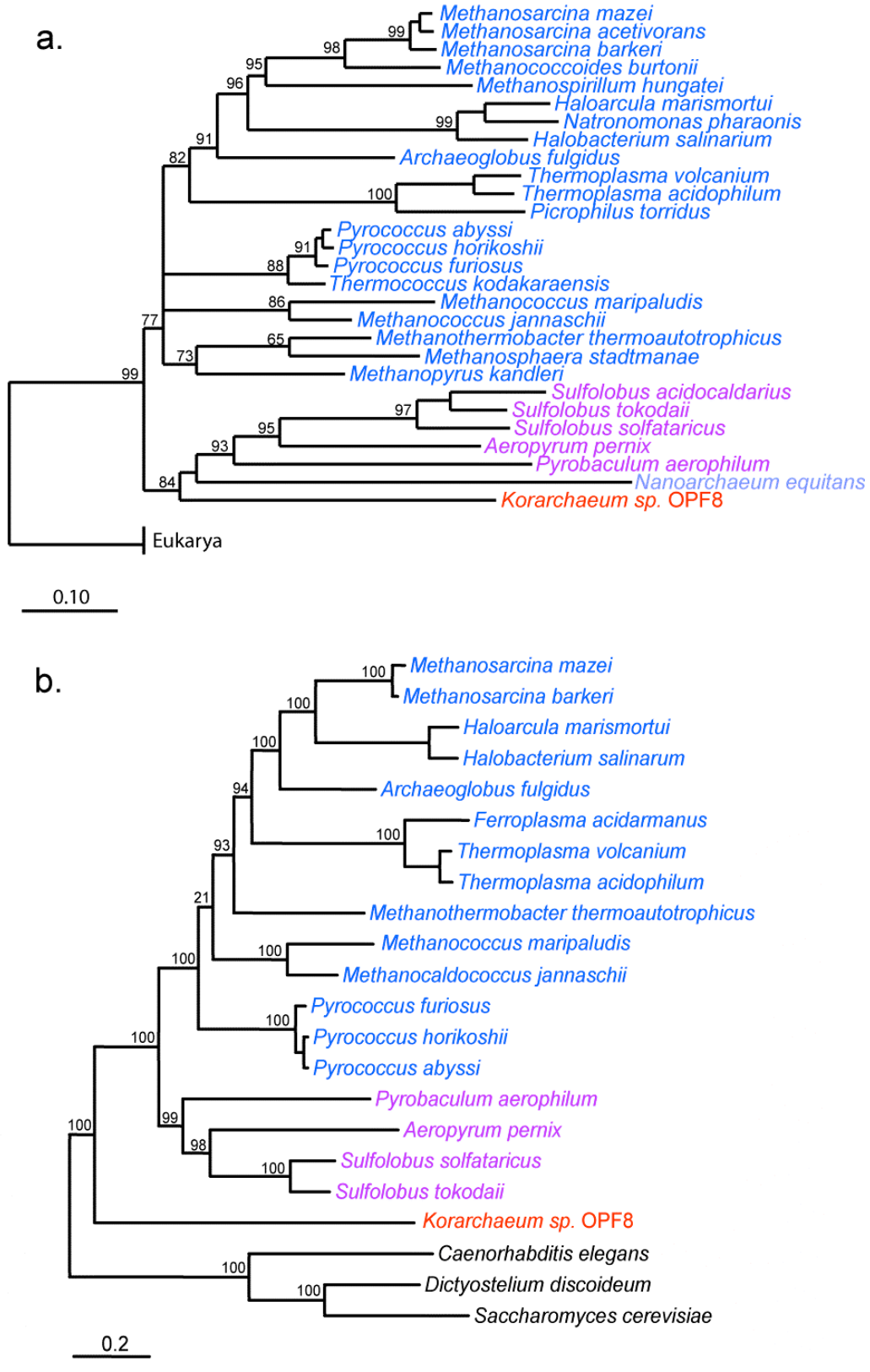
a.
Methanosarcina mazei
Methanosarcina acetivorans
Methanosarcina barkeri
Methanococcoides burtonii
Methanospirillum hungatei
Haloarcula marismortui
Natronomonas pharaonis
Halobacterium salinarium
Archaeoglobus fulgidus
Thermoplasma volcanium
Thermoplasma acidophilum
Picrophilus torridus
Pyrococcus abyssi
Pyrococcus horikoshii
Pyrococcus furiosus
Thermococcus kodakaraensis
Methanococcus maripaludis
Methanococcus jannaschii
Methanothermobacter thermoautotrophicus
Methanosphaera stadtmanae
Methanopyrus kandleri
Sulfolobus acidocaldarius
Sulfolobus tokodaii
Sulfolobus solfataricus
Aeropyrum pernix
Pyrobaculum aerophilum
Nanoarchaeum equitans
Korarchaeum sp. OPF8
Eukarya
0.10
b.
Methanosarcina mazei
Methanosarcina barkeri
Haloarcula marismortui
Halobacterium salinarum
Archaeoglobus fulgidus
Ferroplasma acidarmanus
Thermoplasma volcanium
Thermoplasma acidophilum
Methanothermobacter thermoautotrophicus
Methanococcus maripaludis
Methanocaldococcus jannaschii
Pyrococcus furiosus
Pyrococcus horikoshii
Pyrococcus abyssi
Pyrobaculum aerophilum
Aeropyrum pernix
Sulfolobus solfataricus
Sulfolobus tokodaii
Korarchaeum sp. OPF8
Caenorhabditis elegans
Dictyostelium discoideum
Saccharomyces cerevisiae
0.2

Figure 4

Supplementary data:

Table. S1. Predicted functions identified in the OPF8 genome that are involved in peptide degradation.

Step	Symbol	Activity	E.C. Number	Predicted Function	Accession Number	Organism	Expect value
1	**PRO/PEP**	**Protease/ Peptidase**	3.4.-.-	hypothetical protein Peptidase family M28	XP_383290	*Gibberella zeae*	2.0E-18
				predicted Zn-dependent proteases	AAB98269	*Methanocaldococcus jannaschii*	2.0E-26
				thermostable carboxypeptidase	AAL63098	*Pyrobaculum aerophilum*	2.0E-135
				protease	YP_255674	*Sulfolobus acidocaldarius*	7.0E-57
				hpothetical protein Peptidase_M20	ZP_01121911	*Robiginitalea biformata*	4.0E-102
				acylamino acid-releasing enzyme	BAD84941	*Thermococcus kodakarensis*	3.0E-129
				Xaa-Pro aminopeptidase	BAD85644	*Thermococcus kodakarensis*	7.0E-109

peptidase T2, asparaginase 2	YP_589629	*Acidobacteria bacterium*	4.0E-56
endo-1,4-beta-glucanase/peptidase	AAB84943	*Methanothermobacter thermautotrophicus*	6.0E-68
intracellular protease	AAL62669	*Pyrobaculum aerophilum*	4.0E-63
hypothetical X-Pro dipeptidase	BAA30249	*Pyrococcus horikoshii*	4.0E-42
proteasome alpha subunit	O29760	*Archaeoglobus fulgidus*	3.0E-53
proteasome beta subunit precursor	Q9P996	*Archaeoglobus fulgidus*	8.0E-29
membrane-associated metalloprotease	BAD86009	*Thermococcus kodakarensis*	4.0E-28
proteasome protease subunit	YP_255339	*Sulfolobus acidocaldarius*	2.0E-43
o-sialoglycoprotein endopeptidase	AAL80296	*Pyrococcus furiosus*	2.0E-78
methionyl aminopeptidase	BAD85372	*Thermococcus kodakarensis*	1.0E-55
peptidase M42	YP_643519	*Rubrobacter xylanophilus*	4.0E-47

2	**FOR**	**Formaldehyde oxidoreductase**	1.2.7.5	aldehyde oxidoreductase, ferredoxin-dependent	CAF18474	*Thermoproteus tenax*	0.0E+00
3	**AT**	**Aminotransferases**	2.6.1.-	probable asparate aminotransferase	ZP_00909884	*Clostridium beijerincki*	2.0E-28
				hypothetical 4-aminobutyrate aminotransferase	AAL81545	*Pyrococcus furiosus*	2.0E-180
				aminotransferase, class I and II	ZP_00672559	*Trichodesmium erythraeum*	1.0E-99
				aspartate aminotransferase	AAC07746	*Aquifex aeolicus*	4.0E-46
				4-aminobutyrate aminotransferase	YP_462494	*Syntrophus aciditrophicus*	3.0E-73
				aminotransferase, class V	YP_430497	*Moorella thermoacetica*	7.0E-84
				acetylornithine aminotransferase	AAK40506	*Sulfolobus solfataricus*	2.0E-96
				hypothetical protein	BAA29276	*Pyrococcus horikoshii*	8.0E-53
4	**GDH**	**Aminoacid**		probable glutamate dehydrogenase	CAC11907	*Thermoplasma acidophilum*	6.0E-139

		(glutamate) dehydrogenase					
5	**POR**	**pyruvate ferredoxin oxidoreductase/ 2-ketovalerate ferredoxin oxidoreductase**	1.2.7.1				
			α	pyruvate flavodoxin/ferredoxin oxidoreductase	ZP_01189632	*Halothermothrix orenii*	5.0E-109
			β	thiamine pyrophosphate enzyme, porB like	ZP_01189631	*Halothermothrix orenii*	1.0E-84
			γ	pyruvic-ferredoxin oxidoreductase, gamma subunit	AAW40022	*Dehalococcoides ethenogenes*	1.0E-40
			δ	ferredoxin oxidoreductase delta subunit	BAA29773	*Pyrococcus horikoshii*	7.0E-17

6	**VOR**	**2-ketoisovalerate ferredoxin oxidoreductase**	1.2.7.1				
			α	pyruvate ferredoxin oxidoreductasse [POR] alpha subunit	BAB59991	*Thermoplasma volcanium*	5.0E-104
			β	ferrodoxin oxidoreductase beta subunit	YP_256832	*Sulfolobus acidocaldarius*	5.0E-90
			γ	pyruvate ferredoxin oxidoreductase/2-ketovalerate ferredoxin oxidoreductase gamma subunit	AAL81095	*Pyrococcus furiosus*	2.0E-45
			δ	pyruvic-ferredoxin oxidoreductase delta chain	AAK41455	*Sulfolobus solfataricus*	8.0E-26
7	**KGOR**	**2-ketoglutarate ferredoxin**	1.2.7.3				

		oxidoreductase					
			α	ferredoxin oxidoreductase alpha subunit	BAA30774	*Pyrococcus horikoshii*	2.0E-100
			β	2-ketoglutarate ferredoxin oxidoreductase subunit beta (korB)	AAB98531	*Methanocaldococcus jannaschii*	4.0E-71
			γ	korG-2 2-ketoglutarate ferredoxin oxidoreductase, subunit gamma	CAB49431	*Pyrococcus abyssi*	5.0E-22
			δ	korD 2-ketoglutarate ferredoxin oxidoreductase, subunit delta	CAB49425	*Pyrococcus abyssi*	8.0E-10
8	**IOR**	**indolepyruvate ferredoxin oxidoreductase**	1.2.7.8				
			α	pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases, alpha subunit	GAA00295	*Pelotomaculum thermopropionicum*	2.0E-54

9	FNOR	Ferredoxin NADP oxidoreductase	1.18.1.3	H-II gamma (hydrogenase subunit gamma) Oxidoreductase NAD-binding domain	AAL81454	*Pyrococcus furiosus*	8.0E-78
10	**ACS**	**Acyl-CoA synthetase**	6.2.1.13				
			α	acdA-2 acetate--coA ligase (ADP-forming) alpha chain	CAB49565	*Pyrococcus abyssi*	1.0E-64
			β	acetyl-CoA synthetase I (NDP forming), beta subunit	BAD84654	*Thermococcus kodakarensis*	2.0E-68
11	**AOR**	**Aldehyde ferredoxin oxidoreductase**	1.2.7.5	aldehyde oxidoreductase, ferredoxin-dependent	CAF18474	*Thermoproteus tenax*	0.0E+00
				aldehyde ferredoxin oxidoreductase	ZP_0077761	*Thermoanaerobacter ethanolicus*	3.0E-116

				aor-1 tungsten-containing aldehyde ferredoxin oxidoreductase	CAB49318	*Pyrococcus abyssi*	0.0E+00
				aldehyde ferredoxin oxidoreductase, tungsten-containing	ABB14938	*Carboxydothermus hydrogenoformans*	1.0E-171
				aldehyde ferredoxin oxidoreductase (aor)	AAL63910	*Pyrobaculum aerophilum*	1.0E-139
				aldehyde ferredoxin oxidoreductase (aor-2)	NP_068918	*Archaeoglobus fulgidus*	0.0E+00
12	**H_2ase**	**Hydrogenase**	1.12.99.6				
			α	hydA-1 cytochrome-c3 hydrogenase	CAB49779	*Pyrococcus abyssi*	5.0E-94
			β	sulfhydrogenase II, beta chain	CAB49860	*Pyrococcus abyssi*	3.0E-91
			γ	H-II gamma	AAL81454	*Pyrococcus furiosus*	8.0E-78
			δ	sulfhydrogenase II, delta chain	CAB49862	*Pyrococcus abyssi*	4.0E-51

Table S2. COG groups used for concatenated gene phylogeny.

COG	Function
COG0012*	Predicted GTPase, probable translation factor
COG0016	Phenylalanine-tRNA synthethase alpha subunit
COG0048	Ribosomal protein S12
COG0049	Ribosomal protein S7
COG0052	Ribosomal protein S2
COG0080	Ribosomal protein L11
COG0081	Ribosomal protein L1
COG0087	Ribosomal protein L3
COG0091	Ribosomal protein L22
COG0092	Ribosomal protein S3
COG0093	Ribosomal protein L14
COG0094	Ribosomal protein L5
COG0096	Ribosomal protein S8
COG0097	Ribosomal protein L6P/L9E
COG0098	Ribosomal protein S5
COG0099	Ribosomal protein S13
COG0100*	Ribosomal protein S11
COG0102-	Ribosomal protein L13
COG0103	Ribosomal protein S9
COG0172*	Seryl-tRNA synthetase

COG0184	Ribosomal protein S15P/S13E
COG0186*	Ribosomal protein S17
COG0197	Ribosomal protein L16/L10E
COG0200-	Ribosomal protein L15
COG0201	Preprotein translocase subunit SecY
COG0202	DNA-directed RNA polymerase, alpha subunit
COG0256	Ribosomal protein L18
COG0495*	Leucyl-tRNA synthetase
COG0522*	Ribosomal protein S4 and related proteins
COG0533	Metal-dependent proteases with chaperone activity

* Denotes that the gene was called multiple times in the genome and - denotes that the gene was not identified in some genomes and therefore omitted. Both these categories were excluded from the analysis.

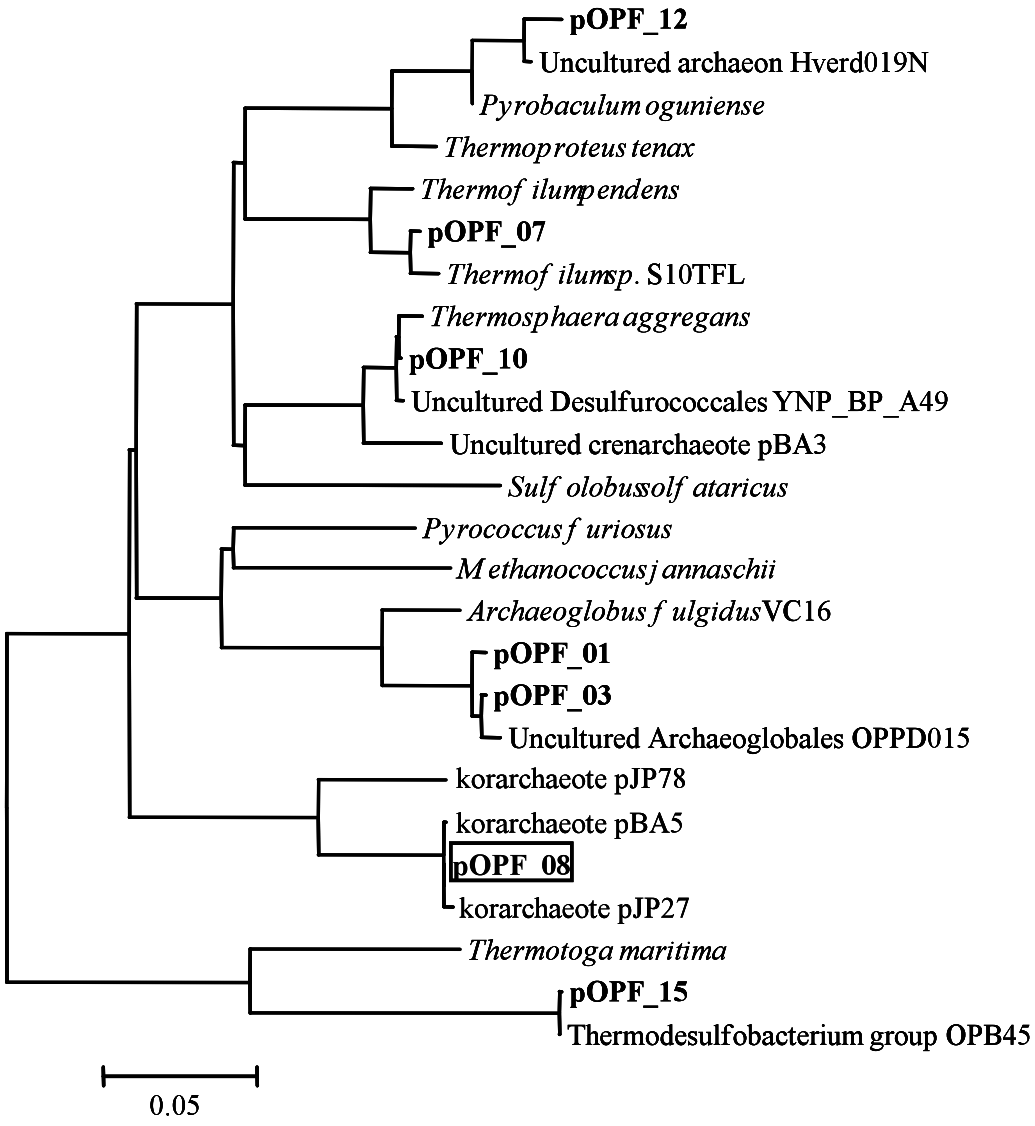

Figure S1. A. Neighbour-Joining tree constructed from distance matrix analysis of 16S rDNA sequences obtained from the Obsidian Pool enrichment culture. The *Korarchaeum* sp. represented by pOPF_08 is outlined.

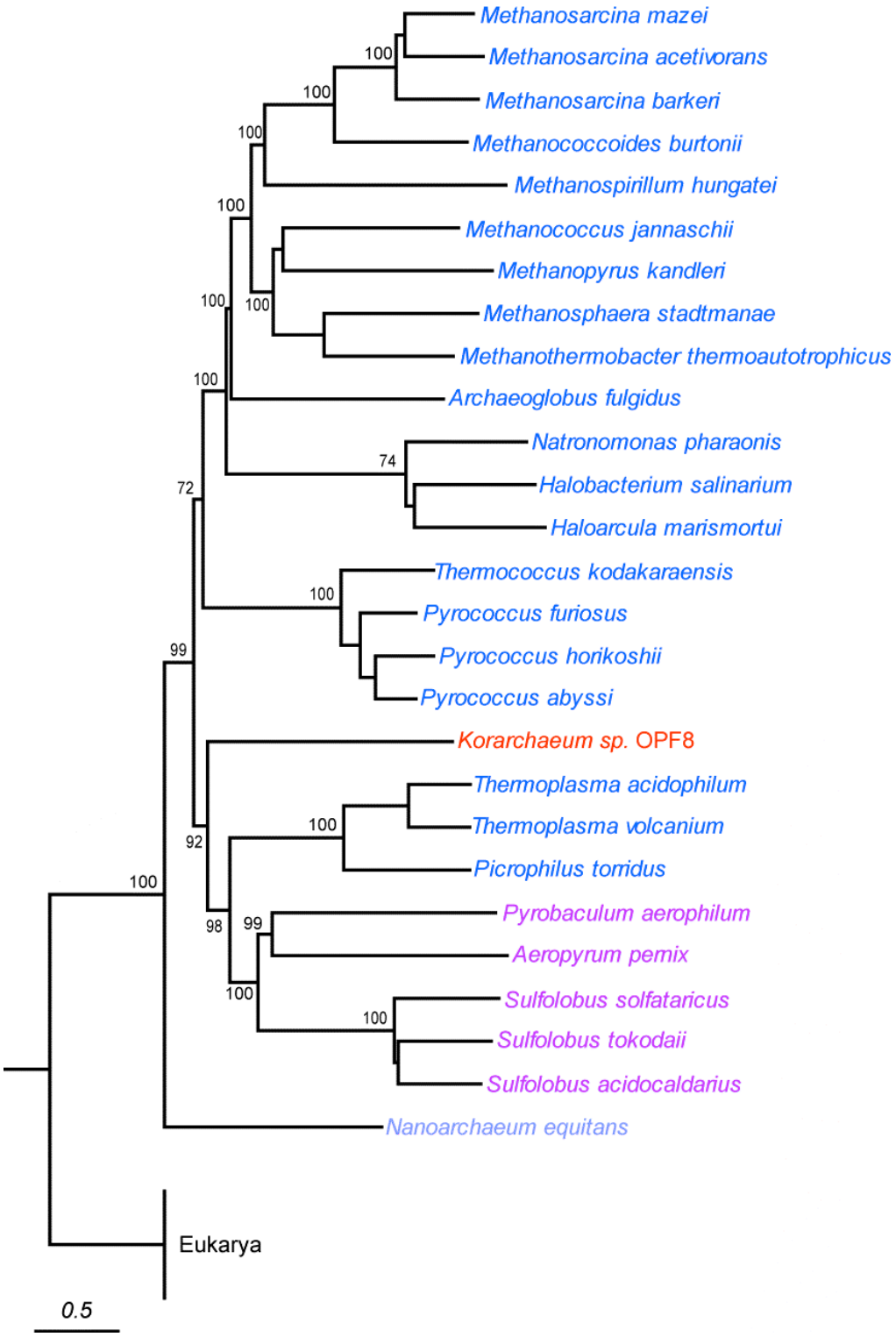

Figure S2. Figure 4 Phylogenetic tree showing placement of *Korarchaeum sp.* OPF8 relative to other Archaea and Eukarya based on whole-genome conservation. Whole genome conservation method combines both sequence

similarity and gene content in a single measure (See reference 24). Maximum-likelihood bootstrap support is shown at the corresponding node (100 replicates). The scale bar represents the number of changes per aligned amino acid position per unit branch length. Eukarya outgroup for COG tree includes *Saccharomyces cerevisiae*, *Candida glabrata*, *Eremothecium gossypii*, *Kluyveromyces lactis*, *Debarymyces hansenii*, *Yarrowia lipolytica*, *Schizosaccharomyces pombe*, *Paramecium tetraurelia*, *Aspergillus fumigatus*, *Cryptococcus neoformans*, *Encephalitozoon cuniculi*,and *Guillardia theta.*

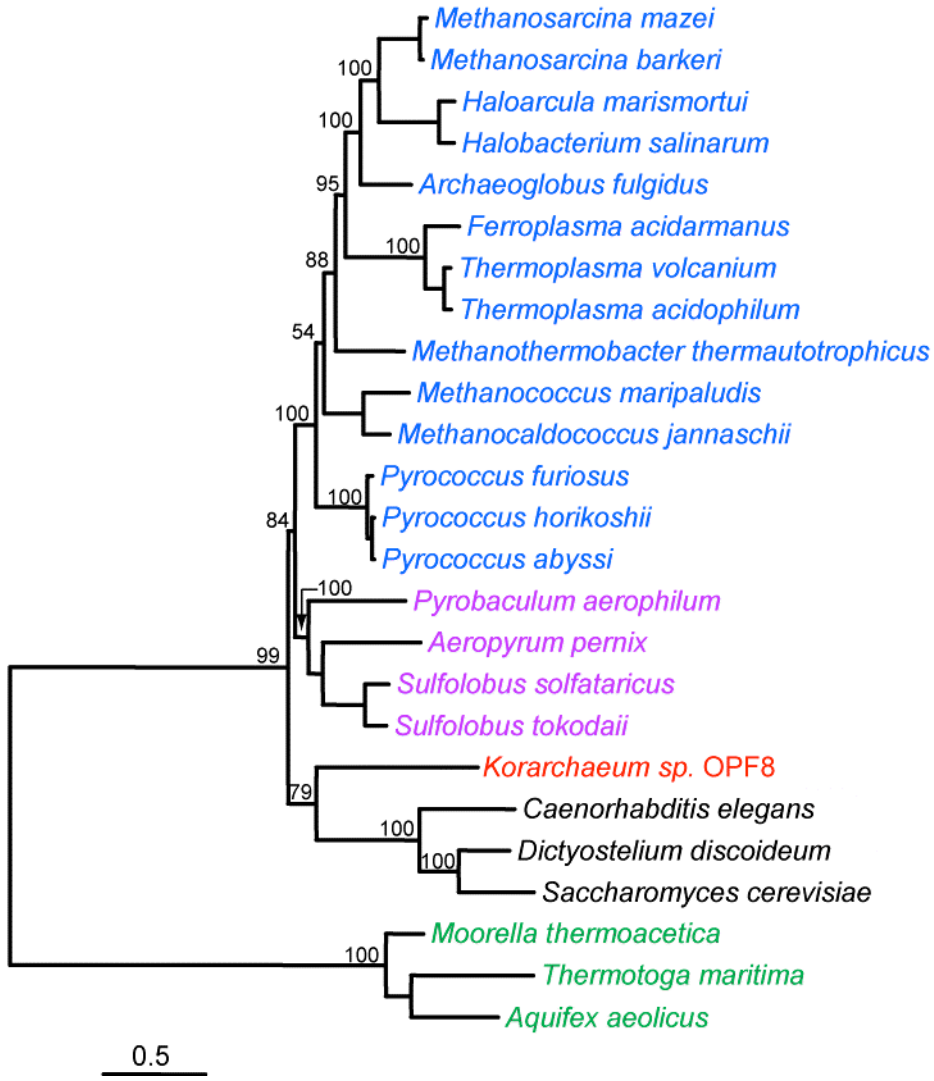

Figure S3. Maximum-likelihood (ML) phylogenetic trees constructed from concatenated DNA-dependent RNA polymerase (RNAP) subunits A/A'/A'', B/B'/B'', and D from Archaea and Eukarya. Corresponding β, β', and α subunits from several bacterial RNAPs were used as an outgroup. ML Bootstrap support is shown at the corresponding node (100 replicates). The scale bar represents the number of changes per aligned amino acid position per unit branch length.