

Supplementary Information

***De Novo* Metagenomic Assembly Reveals Abundant Novel Major Lineage of Archaea in Hypersaline Microbial Communities**

Priya Narasingarao, Sheila Podell, Juan A. Ugalde, Céline Brochier-Armanet, Joanne B. Emerson, Jochen J. Brocks, Karla B. Heidelberg, Jillian F. Banfield, Eric E. Allen

Contents

Supplementary Methods	2
Supplementary Tables S1-S8	5
Supplementary Figures S1-S2.....	16
References.....	19

Supplementary Methods

Metagenome assembly methods

Details of genomic DNA sequence libraries are provided in **Supplementary Table S1**. Sanger end sequencing yielded average read lengths of 800 nucleotides. One 0.1 μm filter library was also pyrosequenced using 454 Titanium chemistry, with an average read length of 400 nucleotides. Assembly software parameters were systematically varied in preliminary assemblies to empirically identify settings producing maximum scaffold length, minimum gap length, and sharpest peak resolution in a length-weighted histogram of scaffold percent GC. Final assembly parameters were:

```
utgErrorRate=0.10
ovlErrorRate=0.10
cnsErrorRate=0.10
cgwErrorRate=0.12
utgBubblePopping=0
utgGenomeSize=500000
merSize=15
doFragmentCorrection=0
doExtendClearRanges=1
doResolveSurrogates=1
Unitigger parameter -j = -20.
```

Composite assemblies including all available libraries were used to classify scaffolds into one of four categories: Nanohaloarchaea, non-Nanohaloarchaea, ambiguous, or unclassified. Scaffold classifications were based on manual assessment of multiple independent parameters. Matches to known microbial 16S rRNA sequences were identified by BLASTN search against the GreenGenes reference database (DeSantis *et al.*, 2006), requiring a minimum alignment length of 200 nucleotides and e-value of $1\text{e-}7$ or better. For each scaffold, number and alignment quality of predicted protein matches to GenBank sequences associated with previously sequenced halophilic archaeal genomes were assessed based on a BLASTX search against the GenBank nr database with an e-value cutoff of $1\text{e-}5$ or better. For scaffolds longer than 3,000 nucleotides, distribution of reads derived from the 0.1 μm and 0.8 μm filter libraries and percent G+C nucleotide composition were also considered in determining scaffold origin.

A second, iterative round of assembly was performed combining all reads previously associated with scaffolds identified as possible Nanohaloarchaea, ambiguous or unclassified. Unassembled singletons reads were also included in the second round, but reads associated with scaffolds identified as non-Nanohaloarchaea were excluded. Scaffolds from the second round of assembly were classified according to the same procedures as the first round.

Phylogenetic analysis methods

Sixty-six individual ribosomal protein datasets from Euryarchaeotal genomes (**Supplementary Table S8**), including J07AB43 and J07ABA56, were aligned using MAFFT version 6.833 (Philippe, 1993). All alignments were visually inspected and manually corrected when necessary. Regions where the alignment was doubtful were removed prior to further phylogenetic analyses. Datasets for nine ribosomal proteins were discarded because of insufficient taxonomic sampling (L13e, L14e, L20a, L30e, L34e, L35ae, S25e, S26e, S30e), since these proteins were absent in more than 10 euryarchaeal genomes. The remaining 57 ribosomal proteins were combined to construct a single alignment containing 5858 amino acid positions. This alignment was subsequently recoded using Dayhoff 4 categories (ASTGP, DNEQ, RHK and MVILFWY, with cysteines being considered as missing characters) (Hrdy *et al.*, 2004; Susko and Roger, 2007). A subset of 43 of the original 68 archaeal genomes were selected to represent the diversity of Euryarchaeota, eliminating multiple closely related representatives of the same order (e.g. *Pyrococcus*, *Thermococcus*, *Methanosarcina*, and *Methanococcus maripaludis*).

Maximum likelihood (ML) phylogenetic analyses were performed using both TreeFinder v.10.08 (Jobb *et al.*, 2004) and PhyML version 3.0 (Guindon and Gascuel, 2003). The evolutionary model was chosen in agreement with the ‘propose model’ tool implemented in TreeFinder (AICc criterion). For the euryarchaeal 16S rRNA and the ribosomal protein recoded alignment analyses we used the General Time Reversible model, with empirical nucleic acids estimated frequencies, a gamma correction (4 discrete categories and an estimated alpha-parameter) to take into account evolutionary rate variation across sites. For the euryarchaeal 16S rRNA analysis we have also considered that a proportion of the sites were invariant (this proportion was estimated by ML). For the alignment of the non-recoded ribosomal protein concatenation analysis, the Le and Gascuel model with empirical amino acid estimated frequencies, a gamma correction (4 discrete categories and an estimated alpha-parameter), and an estimated frequency of invariant sites was employed. Additional ML phylogenetic analyses of these three datasets were performed using

PhyML with the same parameters. Finally, nanohaloarchaeal diversity (Figure 6) was inferred by ML phylogenetic analysis via TreeFinder using the J1 model and gamma correction. The robustness of each resulting ML phylogenetic tree was estimated using non parametric bootstrap approaches implemented in TreeFinder and PhyML (100 replicates of the original dataset).

Supplementary Table S1. Summary of Lake Tyrrell metagenomic sequencing libraries used in this study. Average read length is shown \pm standard deviation.

Library name	Retention Filter	Collection Date	Library type	Sequencing platform	Number of reads	Avg. read length
ABM	0.8 μ m	1/23/2007	8-10 kb plasmid	Sanger	81,197	776 \pm 157
ABL	0.8 μ m	1/23/2007	40 kb fosmid	Sanger	43,490	599 \pm 230
CBM	0.8 μ m	1/25/2007	8-10 kb plasmid	Sanger	81,032	781 \pm 155
CBL	0.8 μ m	1/25/2007	40 kb fosmid	Sanger	33,029	678 \pm 191
AAM	0.1 μ m	1/23/2007	8-10 kb plasmid	Sanger	95,988	699 \pm 198
CAM	0.1 μ m	1/25/2007	8-10 kb plasmid	Sanger	92,242	747 \pm 179
CAT	0.1 μ m	1/25/2007	454-Titanium	Pyrosequence	205,925	401 \pm 120
Total					632,903	622 \pm 229

Supplementary Table S2. Assembly properties of J07AB43 and J07AB56 group scaffolds. Total percentages were calculated as length-weighted averages.

Group	Scaffold ID	Scaffold length	Gap length	Num. gaps	Read depth	% GC	% 0.1 μm reads
J07NFR43	Total	1,227,157	26,201	210	8.3	43%	97%
	7180000030744	798,418	8,083	144	8.8	43%	97%
	7180000030739	112,863	310	9	9.7	43%	97%
	7180000030726	111,825	3,412	19	5.9	43%	99%
	7180000030734	65,032	370	12	7.9	43%	96%
	7180000030724	54,503	2,312	8	5.4	43%	98%
	7180000030742	52,428	8,276	10	8.1	43%	98%
	7180000030737	32,088	3,438	8	6.0	44%	97%
J07NFR56	Total	1,215,802	121,708	259	5.1	56%	92%
	7180000039101	959,093	100,058	208	5.1	56%	92%
	7180000039072	196,424	8,804	40	5.3	55%	93%
	7180000039097	60,285	12,846	11	4.5	55%	97%

Supplementary Table S3. Estimation of draft genome completeness by detection of 53 highly conserved archaeal proteins.

Func_id	Func_name	J07AB56	J07AB43
COG0013	Alanyl-tRNA synthetase	+	+
COG0018	Arginyl-tRNA synthetase	+	+
COG0455	ATPases involved in chromosome partitioning	+	+
COG0459	Chaperonin GroEL (HSP60 family)	+	+
COG0085	DNA-directed RNA polymerase, beta subunit/140 kD subunit	+	+
COG0086	DNA-directed RNA polymerase, beta' subunit/160 kD subunit	+	+
COG0008	Glutamyl- and glutamyl-tRNA synthetases	+	+
COG0124	Histidyl-tRNA synthetase	+	+
COG0060	Isoleucyl-tRNA synthetase	+	+
COG0495	Leucyl-tRNA synthetase	+	+
COG0024	Methionine aminopeptidase	+	+
COG0143	Methionyl-tRNA synthetase	+	+
COG0071	Molecular chaperone (small heat shock protein)	+	+
COG0016	Phenylalanyl-tRNA synthetase alpha subunit	+	+
COG0177	Predicted EndoIII-related endonuclease	+	+
COG0012	Predicted GTPase, probable translation factor	+	+
COG0201	Preprotein translocase subunit SecY	+	+
COG0130	Pseudouridine synthase	+	+
COG0468	RecA/RadA recombinase	+	+
COG0164	Ribonuclease HII	+	+
COG0081	Ribosomal protein L1	+	+
COG0080	Ribosomal protein L11	+	+
COG0102	Ribosomal protein L13	+	+
COG0093	Ribosomal protein L14	+	-
COG0200	Ribosomal protein L15	+	+
COG0197	Ribosomal protein L16/L10E	+	+
COG0256	Ribosomal protein L18	+	+

Func_id	Func_name	J07AB56	J07AB43
COG0090	Ribosomal protein L2	+	-
COG0091	Ribosomal protein L22	+	-
COG0087	Ribosomal protein L3	+	-
COG0088	Ribosomal protein L4	+	-
COG0094	Ribosomal protein L5	+	+
COG0097	Ribosomal protein L6P/L9E	+	+
COG0051	Ribosomal protein S10	+	+
COG0100	Ribosomal protein S11	+	+
COG0048	Ribosomal protein S12	+	+
COG0099	Ribosomal protein S13	+	+
COG0186	Ribosomal protein S17	+	-
COG0185	Ribosomal protein S19	+	-
COG0052	Ribosomal protein S2	+	+
COG0092	Ribosomal protein S3	+	-
COG0522	Ribosomal protein S4 and related proteins	+	+
COG0098	Ribosomal protein S5	+	+
COG0049	Ribosomal protein S7	+	+
COG0096	Ribosomal protein S8	+	+
COG0103	Ribosomal protein S9	+	+
COG0441	Threonyl-tRNA synthetase	+	+
COG0250	Transcription antiterminator	+	+
COG0480	Translation elongation factors (GTPases)	+	+
COG0361	Translation initiation factor 1 (IF-1)	+	-
COG0532	Translation initiation factor 2 (IF-2; GTPase)	+	+
COG0180	Tryptophanyl-tRNA synthetase	+	+
COG0525	Valyl-tRNA synthetase	+	+

Total		53	44
percent		100%	83%

Supplementary Table S4. Cell counts from fluorescence in situ hybridization images of Lake Tyrrell Australia and South Bay Salt Works, Chula Vista, California.

(a) Lake Tyrrell, Australia (LT)

	Detection	Number (Cells/mL)	Pct.
Total	DAPI	$1.6 \pm 0.3 \times 10^7$	100
Archaea	Arc915 FISH probe	$1.4 \pm 0.2 \times 10^7$	90
Nanohaloarchaea	Narc_1214 FISH probe	$2 \pm 0.5 \times 10^6$	14
Bacteria	Eubmix FISH probe	$1.9 \pm 0.5 \times 10^6$	10

(b) South Bay Salt Works, Chula Vista, California (CV)

	Detection	November, 2009		June, 2010	
		Number (Cells/mL)	Pct.	Number (Cells/mL)	Pct.
Total	DAPI	$3.2 \pm 0.3 \times 10^7$	100	$3.4 \pm 0.3 \times 10^7$	100
Archaea	Arc915 FISH probe	$2.6 \pm 0.3 \times 10^7$	83	$2.8 \pm 0.2 \times 10^7$	77
Nanohaloarchaea	NArc_1214 FISH probe	$2.5 \pm 0.4 \times 10^6$	8	$3.8 \pm 0.7 \times 10^6$	11
Bacteria	Eubmix FISH probe	$5.5 \pm 0.3 \times 10^6$	17	$7.9 \pm 0.4 \times 10^6$	23

Supplementary Table S5. DNA-distance matrix comparison of 16S rRNA genes.

	J07AB43	J0756AB	<i>Halorubrum lacusprofundi</i>	<i>Haloquadratum walsbyi</i>	<i>Halogeometricum borinquense</i>	<i>Halomicrobium mukohataei</i>	<i>Natrialba magadii</i>	<i>Haloterrigena turkmenica</i>	<i>Halobacterium salinarum</i>	<i>Natronomonas pharaonis</i>	<i>Haloarcula marismortui</i>	<i>Halorhabdus utahensis</i>
J07AB43 (<i>Nanosalina</i> sp.)												
J0756AB (<i>Nanosalinarum</i> sp.)	16%											
<i>Halorubrum lacusprofundi</i>	34%	35%										
<i>Haloquadratum walsbyi</i>	37%	35%	15%									
<i>Halogeometricum borinquense</i>	35%	37%	14%	9%								
<i>Halomicrobium mukohataei</i>	34%	34%	14%	16%	16%							
<i>Natrialba magadii</i>	35%	36%	14%	14%	11%	15%						
<i>Haloterrigena turkmenica</i>	35%	35%	15%	13%	12%	16%	4%					
<i>Halobacterium salinarum</i>	33%	34%	13%	14%	13%	14%	11%	12%				
<i>Natronomonas pharaonis</i>	34%	32%	13%	14%	12%	13%	12%	12%	10%			
<i>Haloarcula marismortui</i>	34%	36%	16%	15%	14%	11%	11%	12%	11%	10%		
<i>Halorhabdus utahensis</i>	32%	34%	15%	16%	13%	14%	10%	10%	9%	9%	8%	

Supplementary Table S6. Complete amino acid compositions of microbial genomes shown in Figure 5, ordered by taxonomic group. Key numbers are based on genome percent G+C ranked in ascending order.

key	Species	pet GC	Glu	Asp	Gln	Asn	Lys	Arg	His	Ser	Thr	Tyr	Gly	Ala	Val	Leu	Ile	Pro	Phe	Trp	Met	Cys	Taxonomic Group
19	<i>J07AB43 (Nanosalina sp.)</i>	43	10.1	7.3	3.7	4.6	5.8	4.7	1.7	7.6	5.2	3.3	6.8	5.9	6.9	8.1	6.3	3.4	3.8	0.9	2.3	0.7	Nanohaloarchaea
39	<i>J07AB56 (Nanosalinarum sp.)</i>	56	10.0	7.3	3.6	3.4	3.8	6.5	1.8	7.6	5.3	2.7	7.8	7.3	8.2	8.9	4.4	3.8	3.5	0.8	2.0	0.7	Nanohaloarchaea
43	<i>Haloarcula marismortui</i>	61	8.1	8.3	3.1	2.6	2.0	6.1	2.0	6.0	6.9	2.7	8.3	10.4	8.6	8.8	4.4	4.6	3.3	1.1	1.9	0.8	Halobacteriales
50	<i>Halobacterium salinarum R1</i>	65	7.1	8.9	2.8	2.2	1.8	6.5	2.2	5.4	6.8	2.6	8.2	12.4	9.3	8.6	3.8	4.6	3.1	1.1	1.7	0.8	Halobacteriales
41	<i>Halogeometricum borinquense</i>	61	8.3	8.2	2.7	2.7	2.3	6.3	2.0	6.2	6.7	2.7	8.1	10.2	8.9	8.7	4.3	4.5	3.5	1.1	1.8	0.7	Halobacteriales
51	<i>Halomicrobium mukohataei</i>	66	8.3	8.6	2.7	2.2	1.7	6.7	2.0	5.6	6.7	2.7	8.5	11.1	9.2	8.9	3.9	4.6	3.2	1.2	1.7	0.7	Halobacteriales
28	<i>Haloquadratum walsbyi</i>	48	7.4	7.8	3.4	3.4	2.3	5.8	2.1	6.8	7.9	2.8	7.5	9.4	7.8	8.2	6.2	4.4	3.2	1.0	1.8	0.8	Halobacteriales
46	<i>Halorhabdus utahensis</i>	63	8.6	8.5	2.7	2.3	1.8	6.4	2.0	5.6	6.7	2.6	8.5	10.6	8.9	8.9	4.5	4.6	3.3	1.2	1.7	0.6	Halobacteriales
48	<i>Halorubrum lacusprofundi</i>	64	8.5	8.8	2.4	2.5	2.0	6.7	2.0	5.5	6.3	2.7	8.5	11.0	8.8	8.7	4.2	4.6	3.3	1.1	1.7	0.7	Halobacteriales
49	<i>Haloterrigena turkmenica</i>	65	9.2	8.6	2.4	2.4	1.7	6.7	2.0	5.7	6.5	2.7	8.3	10.5	8.6	8.8	4.4	4.7	3.2	1.2	1.7	0.8	Halobacteriales
42	<i>Natrialba magadii</i>	61	9.2	8.9	2.8	2.5	1.8	6.0	2.1	6.0	6.9	2.7	8.1	10.4	8.3	8.7	4.5	4.6	3.2	1.1	1.7	0.7	Halobacteriales
45	<i>Natronomonas pharaonis</i>	63	8.8	8.7	2.6	2.2	1.9	6.4	2.0	5.3	6.4	2.6	8.4	11.4	8.9	8.8	4.2	4.6	3.3	1.1	1.7	0.8	Halobacteriales
4	<i>Methanobrevibacter smithii</i>	31	7.0	6.5	2.3	6.8	8.0	3.0	1.6	6.5	5.3	4.0	6.6	6.0	6.8	8.6	9.4	3.2	4.3	0.7	2.3	1.3	Methanobacteriales
1	<i>Methanosphaera stadtmanae</i>	29	6.8	6.1	2.6	7.5	8.2	3.0	1.7	6.3	6.7	4.3	5.9	5.2	6.3	8.4	10.3	3.2	3.6	0.6	2.3	1.2	Methanobacteriales
29	<i>Methanobacterium thermoautotrophicum</i>	49	8.1	5.9	1.9	3.3	4.6	6.8	1.9	6.1	5.0	3.2	8.0	7.3	7.7	9.4	7.7	4.3	3.6	0.8	3.1	1.2	Methanobacteriales
7	<i>Methanothermobacter feravidus</i>	32	7.4	5.1	1.6	4.9	10.6	4.1	1.7	5.3	4.6	3.7	6.7	5.9	7.1	8.9	10.0	4.1	3.9	0.7	2.4	1.3	Methanobacteriales
3	<i>Methanococcus jannaschii</i>	31	8.6	5.5	1.5	5.3	10.4	3.8	1.4	4.5	4.0	4.4	6.3	5.5	6.8	9.5	10.5	3.4	4.3	0.7	2.3	1.3	Methanococcales
2	<i>Methanococcus aeolicus Nankai-3</i>	31	7.5	5.6	1.8	7.1	9.6	3.0	1.6	5.4	4.8	4.1	6.7	5.4	6.1	8.7	10.8	3.5	3.7	0.6	2.6	1.4	Methanococcales
5	<i>Methanococcus vannielii SB</i>	32	7.9	5.2	1.8	6.0	9.4	3.2	1.5	6.4	4.9	3.9	6.6	5.3	6.7	9.3	9.9	3.4	4.3	0.6	2.5	1.3	Methanococcales
6	<i>Methanococcus voltae A3</i>	32	7.7	5.8	2.1	7.4	9.3	2.9	1.4	6.1	5.1	4.3	6.2	5.4	6.3	9.0	9.5	3.1	3.7	0.6	2.6	1.4	Methanococcales
8	<i>Methanococcus maripaludis C5</i>	33	8.3	5.6	1.9	5.6	8.9	3.2	1.5	6.2	5.0	3.9	6.6	5.6	6.8	9.2	9.4	3.3	4.3	0.6	2.6	1.3	Methanococcales
34	<i>Methanocella paludicola</i>	54	6.2	5.8	2.4	3.6	6.1	5.2	1.7	6.1	5.2	3.6	8.0	8.6	7.7	9.2	7.1	4.5	3.7	1.0	3.2	1.3	Methanocellales
38	<i>Candidatus Methanosphaerula palustris</i>	56	6.1	5.6	3.4	3.4	3.9	5.6	2.0	6.1	6.8	3.0	8.2	8.5	7.8	9.4	7.0	4.8	3.7	1.0	2.5	1.3	Methanomicrobiales
36	<i>Candidatus Methanoregula boonei</i>	55	6.1	5.3	2.9	3.3	4.7	5.5	2.0	6.0	6.2	3.0	8.0	9.5	7.5	9.2	7.2	4.9	3.9	1.0	2.4	1.4	Methanomicrobiales

key	Species	pct GC	Glu	Asp	Gln	Asn	Lys	Arg	His	Ser	Thr	Tyr	Gly	Ala	Val	Leu	Ile	Pro	Phe	Trp	Met	Cys	Taxonomic Group
31	<i>Methanocorpusculum labreanum</i>	50	6.7	5.5	2.5	3.6	5.5	4.7	1.8	6.0	6.0	3.4	7.8	8.5	7.5	9.2	7.6	4.4	3.9	0.9	2.9	1.5	Methanomicrobiales
44	<i>Methanoculleus marisnigri</i>	62	7.2	5.7	2.3	2.7	3.4	7.0	2.0	5.1	5.5	3.0	8.5	10.2	8.3	9.5	6.1	5.0	3.6	1.0	2.4	1.3	Methanomicrobiales
30	<i>Methanoplanus petrolearius</i>	50	7.5	6.0	2.2	4.1	5.7	4.5	1.6	6.9	5.3	3.6	7.9	7.2	7.0	8.7	8.4	4.1	4.3	1.0	2.7	1.4	Methanomicrobiales
25	<i>Methanospirillum hungatei</i>	45	6.8	5.8	3.2	3.6	5.0	5.2	2.2	6.7	5.8	3.4	7.3	7.0	6.5	9.2	8.5	4.7	4.0	1.1	2.6	1.4	Methanomicrobiales
35	<i>Methanosaeta thermophila</i>	54	7.3	5.8	2.1	2.9	4.0	7.5	1.8	6.7	4.6	3.1	7.9	8.3	7.8	9.5	7.3	4.5	3.3	1.0	3.1	1.3	Methanosarcinales
16	<i>Methanococcoides burtonii</i>	41	7.3	6.3	2.5	4.6	6.5	4.2	1.9	6.7	5.4	3.4	7.1	6.8	7.3	8.9	8.4	3.6	4.0	0.9	3.0	1.2	Methanosarcinales
21	<i>Methanohalophilus mahii</i>	43	7.7	6.2	2.8	4.5	6.1	4.5	1.9	6.4	5.3	3.4	7.4	7.1	7.2	8.8	8.0	3.9	3.8	0.8	2.9	1.3	Methanosarcinales
11	<i>Methanosalsum zhilinae</i>	38	7.4	6.3	2.7	4.6	6.2	4.7	2.0	6.9	5.1	3.5	7.0	6.5	6.9	8.6	9.0	3.8	3.9	0.9	2.9	1.2	Methanosarcinales
15	<i>Methanosarcina barkeri fusaro</i>	41	7.6	5.2	2.6	4.8	7.0	4.2	1.7	7.0	5.5	3.7	7.1	6.8	6.8	9.4	7.8	3.9	4.3	1.0	2.4	1.2	Methanosarcinales
20	<i>Methanosarcina mazei</i>	43	8.2	5.3	2.5	4.4	6.8	4.7	1.7	6.8	5.1	3.5	7.2	7.1	6.9	9.4	7.7	4.0	4.3	0.9	2.5	1.2	Methanosarcinales
22	<i>Methanosarcina acetivorans</i>	44	8.0	5.3	2.5	4.5	6.6	4.5	1.7	6.9	5.4	3.7	7.2	6.9	6.8	9.4	7.4	4.0	4.4	1.1	2.5	1.3	Methanosarcinales
40	<i>Methanopyrus kandleri</i>	60	10.0	5.8	1.4	1.9	4.0	8.3	1.9	4.6	4.6	2.8	8.1	8.3	10.4	10.1	4.8	5.5	2.9	1.2	1.9	1.3	Methanopyrales
18	<i>Archaeoglobus profundus</i> 1	43	8.6	5.3	1.6	3.5	8.2	5.7	1.5	5.0	4.1	3.8	6.6	6.5	8.7	9.8	8.3	3.7	4.5	1.0	2.3	1.3	Archaeoglobales
27	<i>Archaeoglobus fulgidus</i>	48	8.9	4.9	1.8	3.2	6.9	5.8	1.5	5.5	4.2	3.6	7.2	7.8	8.6	9.5	7.2	3.9	4.6	1.0	2.6	1.2	Archaeoglobales
23	<i>Ferroglobus placidus</i>	45	9.5	4.7	1.5	3.4	8.4	5.3	1.5	5.5	4.0	3.8	6.7	7.1	8.6	9.5	7.7	3.7	4.8	1.0	2.2	1.1	Archaeoglobales
37	<i>Palaeococcus ferrophilus</i>	55	8.9	4.6	1.7	3.2	6.3	6.2	1.7	5.1	4.6	3.8	7.9	7.5	8.5	10.5	6.6	4.3	4.3	1.2	2.5	0.5	Thermococcales
13	<i>Pyrococcus furiosus</i>	40	8.9	4.4	1.8	3.5	8.1	5.3	1.5	4.9	4.4	4.0	7.1	6.6	7.9	10.1	8.7	4.3	4.4	1.2	2.3	0.6	Thermococcales
17	<i>Pyrococcus horikoshii</i>	41	8.7	4.4	1.7	3.5	8.0	5.6	1.5	5.2	4.3	3.9	7.1	6.4	7.7	10.2	8.8	4.3	4.4	1.2	2.4	0.6	Thermococcales
24	<i>Pyrococcus abyssi</i>	45	8.8	4.6	1.7	3.3	7.8	5.7	1.5	5.0	4.2	3.8	7.3	6.7	8.1	10.2	8.5	4.2	4.4	1.2	2.4	0.6	Thermococcales
32	<i>Thermococcus onnurineus</i>	51	8.6	4.7	1.8	3.2	6.8	5.7	1.6	4.8	4.7	3.9	7.6	7.6	8.1	10.4	7.6	4.3	4.4	1.2	2.5	0.6	Thermococcales
33	<i>Thermococcus kodakaraensis</i>	52	8.8	4.7	1.8	3.2	6.9	5.8	1.6	5.0	4.6	3.8	7.6	7.4	8.3	10.5	7.0	4.4	4.3	1.3	2.3	0.5	Thermococcales
12	<i>Ferroplasma acidarmanus Fer1</i>	38	6.2	5.4	1.9	6.0	7.4	3.9	1.6	7.3	5.1	4.9	7.0	6.0	6.1	8.6	10.0	3.5	4.7	0.7	3.1	0.6	Thermoplasmatales
9	<i>Picrophilus torridus</i>	36	5.7	5.9	1.5	6.6	6.9	4.4	1.5	7.4	4.4	5.4	6.5	5.6	5.7	8.6	11.1	3.5	4.9	0.7	3.1	0.6	Thermoplasmatales
14	<i>Thermoplasma volcanium</i>	40	6.4	5.5	2.1	4.7	6.9	4.7	1.5	7.5	4.8	4.7	7.0	6.4	7.1	8.7	9.2	3.8	4.7	0.8	2.8	0.6	Thermoplasmatales
26	<i>Thermoplasma acidophilum</i>	46	6.0	5.8	2.2	4.3	5.7	5.5	1.6	7.5	4.8	4.6	7.3	7.0	7.2	8.4	9.0	3.9	4.7	0.9	3.2	0.6	Thermoplasmatales
52	<i>Salinibacter ruber</i>	66	7.0	7.0	3.5	2.5	2.1	7.7	2.2	5.8	6.2	2.6	8.4	10.7	7.9	9.7	3.7	5.7	3.4	1.3	1.9	0.7	Sphingobacteriales
10	<i>Halothermothrix orenii</i>	38	7.5	5.5	2.7	5.2	7.9	4.7	1.6	5.4	4.8	3.9	7.4	5.8	7.2	9.9	8.8	3.7	4.1	0.9	2.3	0.7	Halanaerobiales
53	<i>Halorhodospira halophila</i>	67	7.5	6.1	3.9	2.0	1.8	8.7	2.4	4.5	4.8	2.4	8.9	12.0	7.5	10.7	4.1	5.4	3.1	1.4	2.0	0.9	Chromatiales
47	<i>Chromohalobacter salexigens</i>	64	6.2	6.1	3.7	2.4	2.5	7.5	2.6	5.4	5.2	2.4	8.1	11.5	7.2	11.3	4.6	4.9	3.4	1.5	2.5	1.0	Oceanospirillales

Supplementary Table S7. Environmental sampling and habitat characteristics of database matches to Nanohaloarchaea 16S rRNA gene sequences.

Clone group	Location	Site characteristics	Sample processing	PCR primers	Reference
LT	Lake Tyrrell, Australia	Thalassohaline, neutral pH, 29% salinity, crystallizer pond	0.1 μm (0.8 μm pre-filter)	Arc21F Arc529R	This study
CV	Chula Vista San Diego, USA	Thalassohaline, neutral pH, 27% salinity, crystallizer pond	0.1 μm (0.8 μm pre-filter)	Arc21F LT_1215R	This study
Cry/Baj/LDS	Australia	Neutral pH, 34% salinity	Direct Centrifugation	ArcF1 Arc1492R	(Oh <i>et al.</i> , 2010)
MSP	Magadi Salt Pond, Kenya, East Africa	Soda lake, pH 12, crystallizer pond	0.22 μm (0.8 μm pre-filter)	Arc27Fa Arc1492R	(Grant <i>et al.</i> , 1999)
SFH1E041	Tunisia	Neutral pH, Solar saltern	Direct Centrifugation	Arc21F Univ1390R	(Baati <i>et al.</i> , 2008)
ARDARCSS13	Sidi Ameur Salt Lake, Algerian Sahara	Salt Lake soil	ND	ND	EU869371.1 (Unpublished)
Sua-GRP6	Makgadikgadi salt pans, Botswana	Alkaline evaporator pond	Centrifugation	Arc751F UA1406R	(Gareeb and Setani, 2009)
9cA6	Rio Mesquite, Cuatro Cienegas Basin, Mexico	Neutral pH Shallow pond	Direct to 0.45 μm filter	Univ515F Univ1492R	(Escalante <i>et al.</i> , 2008)
EN	Lake Erliannor, China	pH 8, >30% salinity	Direct to 0.45 μm filter	Arc27Fa Arc1492R	(Pagaling <i>et al.</i> , 2009)
SH	Lake Shangmata, China	pH 8.5, >30% salinity	Direct to 0.45 μm filter	Arc27Fa Arc1492R	(Pagaling <i>et al.</i> , 2009)
XH	Unknown lake, Inner Mongolia China	pH 8.5, >30% salinity	Direct to 0.45 μm filter	Arc27Fa Arc1492R	(Pagaling <i>et al.</i> , 2009)
XA	Unknown pool, Inner Mongolia China	pH 9.5, 19% salinity	Direct to 0.45 μm filter	Arc27Fa Arc1492R	(Pagaling <i>et al.</i> , 2009)
AGA	Salar Guayatayoc, Argentina	pH neutral	Direct to 0.45 μm filter	Arc27Fa Arc1492R	(Pagaling <i>et al.</i> , 2009)

Supplementary Table S8. NCBI accession numbers and 16S rRNA gene coordinates of taxa used in the construction of phylogenetic trees shown in Figure 3 and Supplementary Figure S2A.

Genome	NCBI accession number	DNA strand	Coordinates
<i>Archaeoglobus fulgidus</i> DSM 4304	NC_000917.1	(-) strand	1790478-1788987
<i>Archaeoglobus profundus</i> Av18, DSM 5631	NC_000917.1	(-) strand	1051439-1049930
Candidatus <i>Methanoregula boonei</i> 6A8	NC_009712.1	(-) strand	2251978-2250513
Candidatus <i>Methanosphaerula palustris</i> E1-9c	NC_011832.1	(+) strand	754054-755515
<i>Ferroplasma acidarmanus</i> Fer1, unfinished sequence	gbAABC05000010.1	(+) strand	30596-32036
<i>Haloarcula marismortui</i> ATCC 43049 chromosome I	AY596297.1	(+) strand	2632355-2633826
<i>Halobacterium salinarum</i> R1	NC_010364.1	(+) strand	1863623-1865094
<i>Halogeometricum borinquense</i> PR3, DSM 11551	ABTX01000001.1	(+) strand	1509934-1511411
<i>Halomicrobium mukohataei</i> arg-2, DSM 12286	NC_013202.1	(-) strand	1954831-1956302
<i>Haloquadratum walsbyi</i> DSM 16790	NC_008212.1	(+) strand	67537-69008
<i>Halorhabdus utahensis</i> AX-2, DSM 12940	NC_013158.1	(+) strand	1665958 -1667429
<i>Halorubrum lacusprofundi</i> ATCC 49239 chromosome 2	NC_012028.1	(+) strand	161307-162776
<i>Haloterrigena turkmenica</i> VKM, DSM 5511	NC_013743.1	(+) strand	513463-514937
<i>Methanobacterium thermoautotrophicum</i> str. Delta H	NC_000916.1	(+) strand	1602210-1603688
<i>Methanobrevibacter smithii</i> ATCC 35061	NC_009515.1	(-) strand	334957-333483
<i>Methanocaldococcus jannaschii</i> DSM 2661	NC_000909.1	(-) strand	159459-157985
<i>Methanococcoides burtonii</i> DSM 6242	NC_007955.1	(-) strand	1023598-1022125
<i>Methanococcus aeolicus</i> Nankai-3	NC_009635.1	(-) strand	1196101-1194635
<i>Methanococcus maripaludis</i> C5	NC_009135.1	(+) strand	1594066-1595530
<i>Methanococcus vannieli</i> SB	NC_009634.1	(+) strand	155 -1619
<i>Methanococcus voltae</i> A3, unfinished sequence	NC_014222.1	(+) strand	1536115-1534651
<i>Methanocorpusculum labreanum</i> Z	NC_008942.1	(+) strand	403632-405096
<i>Methanoculleus marisnigri</i> JR1	NC_009051.1	(-) strand	1824115-1822651
<i>Methanohalophilus mahii</i> DSM 05219	NC_014002.1	(+) strand	561038-562512
<i>Methanoplanus petrolearius</i> SEBR 4847, DSM 11571	NC_014507.1	(+) strand	740158 -741625
<i>Methanopyrus kandleri</i> AV19	NC_003551.1	(-) strand	518289-516896
<i>Methanosaeta thermophila</i> PT	NC_008553.1	(+) strand	845527-846999
<i>Methanosalsum zhilinae</i> WeN5, DSM 4017	FJ224366.1	(+) strand	Direct sequence
<i>Methanosarcina acetivorans</i> str. C2A	NC_003552.1	(+) strand	1073041-1074470
<i>Methanosarcina barkeri</i> str. fusaro chromosome I	NC_007355	(-) strand	4677310-4678787
<i>Methanosarcina mazei</i> strain Goel	refNC_003901.1	(-) strand	236779-235306
<i>Methanosphaera stadtmanae</i> DSM 3091	NC_007681.1	(+) strand	408655-410144
<i>Methanospirillum hungatei</i> JF-1	NC_007796.1	(+) strand	39814-41279
<i>Methanothermus fervidus</i> DSM 2088	NC_014658.1	(-) strand	28362-26991
<i>Natrialba magadii</i> ATCC 43099, unfinished sequence	NC_013922.1	(+) strand	26876-28343
<i>Natronomonas pharaonis</i> DSM 2160	NC_007426.1	(-) strand	215928-214463
<i>Palaeococcus ferrophilus</i> DMJ, DSM 13482	NR_028149.1	(+) strand	Direct sequence
<i>Picrophilus torridus</i> DSM 9790	NC_005877.1	(-) strand	472195-470727

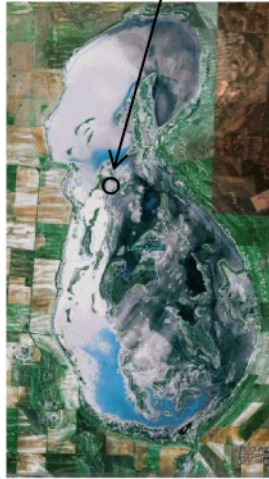
Genome	NCBI accession number	DNA strand	Coordinates
<i>Pyrococcus abyssi</i> GE5	NC_000868.1	(+) strand	205128-206630
<i>Pyrococcus furiosus</i> DSM 3638	NC_003413.1	(+) strand	136930-138426
<i>Pyrococcus horikoshii</i> OT3 DNA	NC_000961.1	(+) strand	190975-192469
<i>Thermococcus onnurineus</i> NAI	NC_011529.1	(+) strand	1279595-1278100
<i>Thermoplasma volcanium</i> GSS1 DNA	NC_002689.2	(-) strand	1520246-1518773
uncultured halophilic archaeon J0756AB	This study	(+) strand	This study
uncultured halophilic archaeon J07AB43	This Study	(-) strand	This Study
Uncultured methanogenic archaeon RC-I	AM114193.2	(+) strand	2083318-2084792
<i>Ferroplasma placidus</i> AEDIII2DO, DSM 10642	AF220166.1	(+) strand	1404007-1405474
<i>Thermococcus kodakarensis</i> KOD1	NC_006624.1	(+) strand	2022864-2024361
<i>Thermoplasma acidophilum</i> 122-1B2	NC_002578.1	(-) strand	1475770-1474300

Supplementary Figure Legends

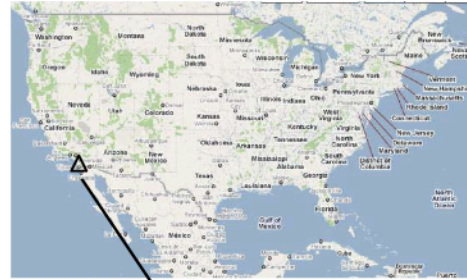
Supplementary Figure S1. Location of sampling sites and sample characteristics.

Supplementary Figure S2. Unrooted maximum-likelihood phylogenetic trees for the Euryarchaeota. Trees are based on (A) 16S rRNA sequences (48 sequences, 1275 positions; NCBI accession numbers and coordinates provided in Supplementary Table S9) and (B) concatenation of 57 ribosomal proteins (43 sequences, 5858 positions). Names of euryarchaeal sub-lineages are shown in the box. Numbers at nodes represent bootstrap values (A) inferred by TreeFinder/PhyML; (B) inferred by TreeFinder based on amino acid alignments and the 4 state recoded amino acid alignments, shown in parentheses. Bootstrap values of less than 50% are indicated by "-" sign. Scale bar represents an average of 0.1 substitutions per site.

Supplementary Figure S1



Lake Tyrrell (LT), Victoria, Australia



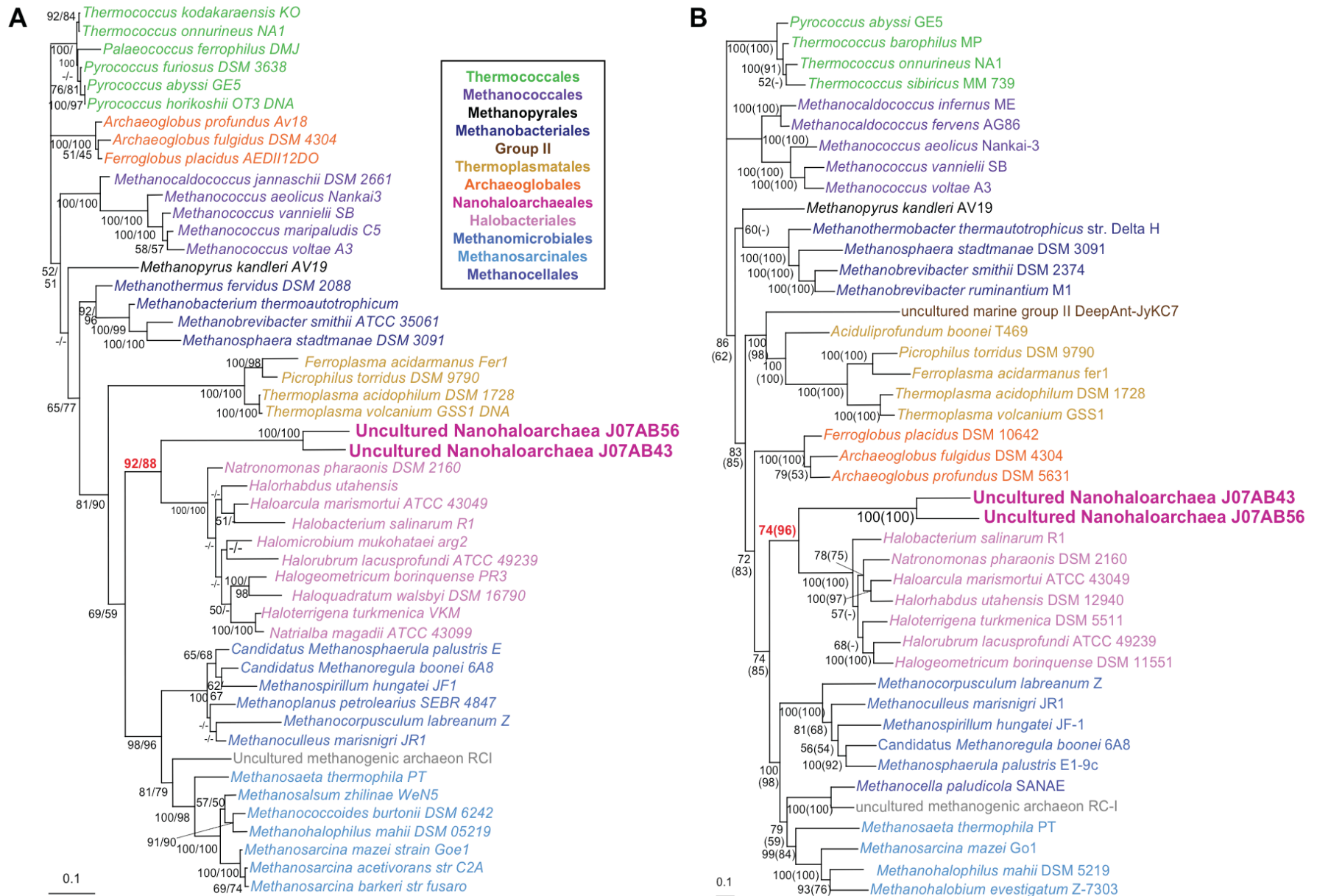
South Bay Salt Works, Chula Vista (CV), California USA

Images:

©2010 Google - Imagery ©2010 TerraMetrics,
Map data ©2010 Google, Whereis(R), Sensis Pty Ltd
©California Department of Fish and Game

Site	Sampling Date	Temp (°C)	pH	Total Dissolved Solids - Salinity (%)
LT	1/23/2007	21.6	7.30	29
LT	1/25/2007	27.9	7.09	29
CV	10/27/2009	22.0	7.40	27
CV	06/22/2009	28.0	7.50	27

Supplementary Figure S2



References

- Baati H, Guerhazi S, Amdouni R, Gharsallah N, Sghir A, Ammar E (2008). Prokaryotic diversity of a Tunisian multipond solar saltern. *Extremophiles* **12**: 505-18.
- DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069-72.
- Escalante AE, Eguiarte LE, Espinosa-Asuar L, Forney LJ, Noguez AM, Souza Saldivar V (2008). Diversity of aquatic prokaryotic communities in the Cuatro Ciénegas basin. *FEMS Microbiol Ecol* **65**: 50-60.
- Gareeb A, Setani M (2009). Assessment of alkaliphilic haloarchaeal diversity in Sua pan evaporator ponds in Botswana. *Afr. J. Biotechnol.* **8**: 259-267.
- Grant S, Grant WD, Jones BE, Kato C, Li L (1999). Novel archaeal phylotypes from an East African alkaline saltern. *Extremophiles* **3**: 139-45.
- Guindon S, Gascuel O (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696-704.
- Hrdy I, Hirt RP, Dolezal P, Bardonova L, Foster PG, Tachezy J *et al* (2004). Trichomonas hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. *Nature* **432**: 618-22.
- Jobb G, von Haeseler A, Strimmer K (2004). TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* **4**: 18.
- Oh D, Porter K, Russ B, Burns D, Dyal-Smith M (2010). Diversity of Haloquadratum and other haloarchaea in three, geographically distant, Australian saltern crystallizer ponds. *Extremophiles* **14**: 161-9.
- Pagaling E, Wang H, Venables M, Wallace A, Grant WD, Cowan DA *et al* (2009). Microbial biogeography of six salt lakes in Inner Mongolia, China, and a salt lake in Argentina. *Appl Environ Microbiol* **75**: 5750-60.

Philippe H (1993). MUST, a computer package of Management Utilities for Sequences and Trees. *Nucleic Acids Res* **21**: 5264-72.

Susko E, Roger AJ (2007). On reduced amino acid alphabets for phylogenetic inference. *Mol Biol Evol* **24**: 2139-50.