

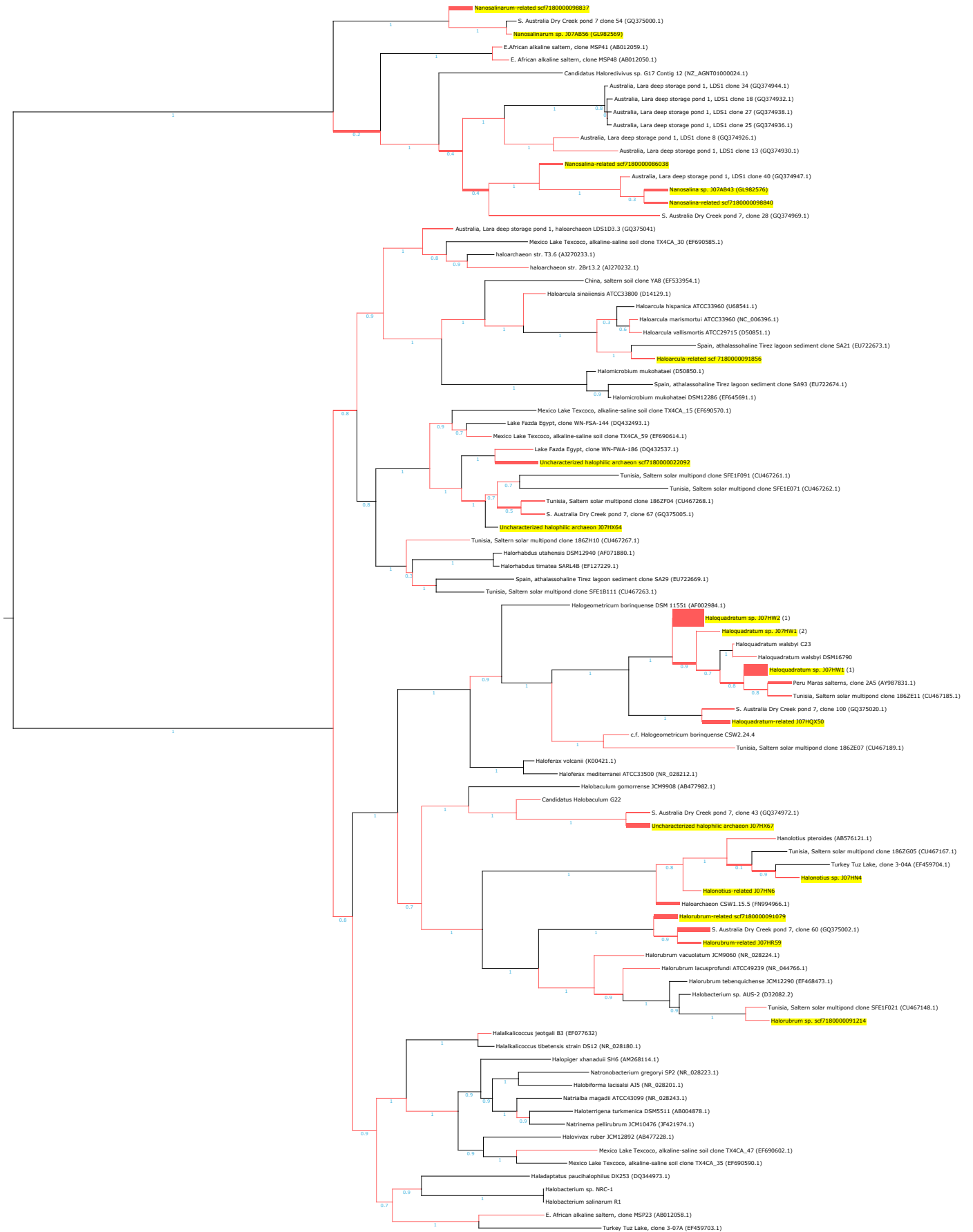
Supporting Figure S1. Bioinformatic analysis pipeline.

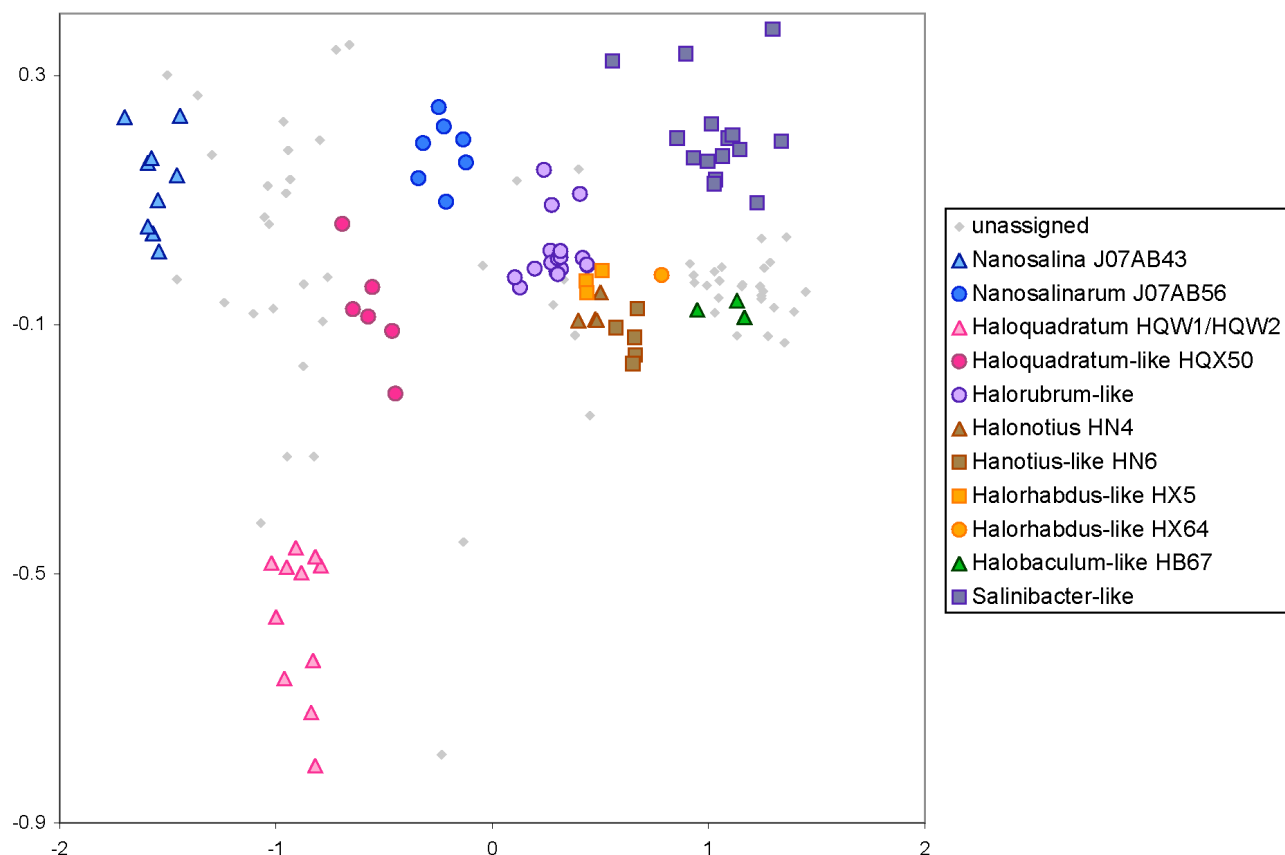
Supporting Figure S2. Phylogenetic trees showing abundance of clustered archaeal 16S rRNA sequences from (A) unassembled reads and (B) PCR-amplified clone libraries. A maximum likelihood archaeal reference tree was constructed using FastTree [1], based on full-length 16S genes from isolate genomes and environmental clones from Genbank nt, supplemented with sequences obtained from Lake Tyrrell assembled scaffolds. Unassembled 16S rRNA sequences from Lake Tyrrell were clustered and inserted into the reference tree using pplacer version v1.1 (model GTR, fig-eval-all) [2] and visualized using Archaeopteryx 0.968 [3]. Part A shows placement of un-amplified Lake Tyrrell raw reads containing 16S gene sequences > 350 nt. Part B shows placement of PCR-amplified 16S rRNA clones (> nt). Numbers at nodes indicate confidence values estimated by FastTree for the reference database. Red lines indicate branches where Lake Tyrrell sequences were observed. The thickness of each red line is proportional to the number of Lake Tyrrell sequences associated with that branch, ranging from one in the thinnest line to 74 in the thickest line.

References

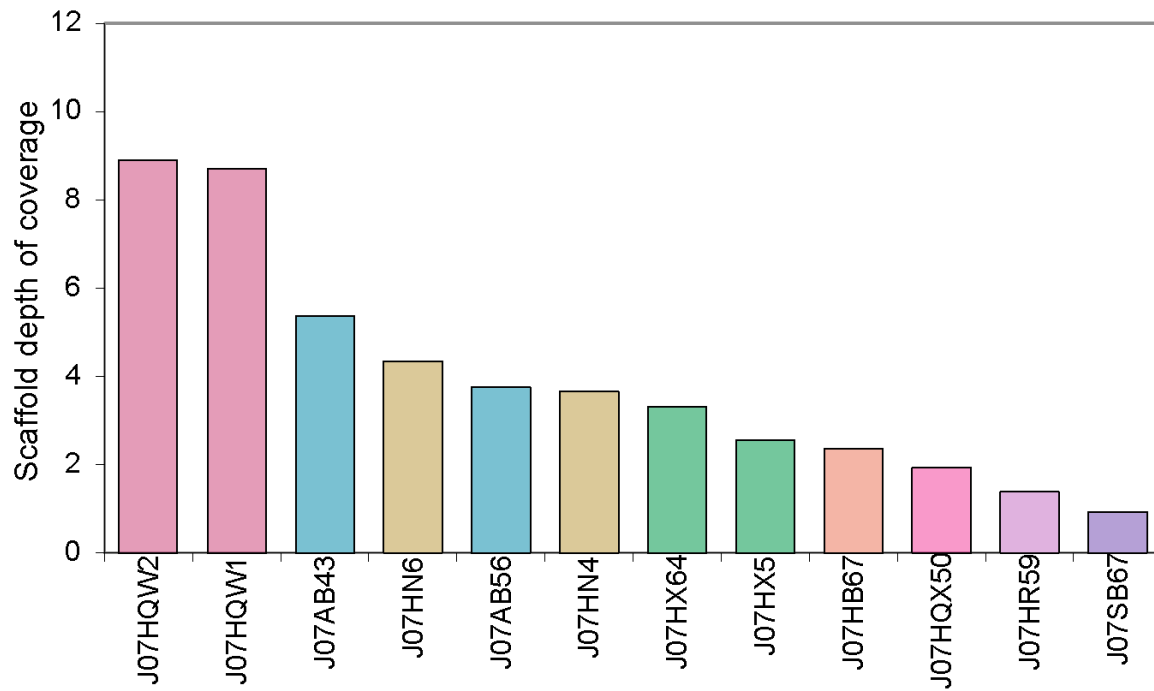
1. Price, M.N., P.S. Dehal, and A.P. Arkin, FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One, 2010. 5(3): p. e9490.
2. Matsen, F., A., Kodner, R., Armbrust, E.V. (2010) pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences on a fixed reference tree. BMC Bioinformatics, 11:538
3. Han M.V. and Zmasek C.M. (2009). phyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinformatics, 10:356.

A





Supplementary Figure S3. Non-metric multidimensional scaling plot illustrating distinctive scaffold groups. Scaffolds > 5Kb from the composite Sanger assembly were subjected to non-metric multidimensional scaling analysis using Primer 1.6, with Euclidean distance, 25 random starts, Krustal fit scheme 1, and minimum stress value 0.01 for the 13 parameters shown in Supplementary Table S5. Axes shown are arbitrary units of composite clustering, although the X axis appears to be dominated by nucleotide percent G+C. Scaffolds associated with major taxonomic groups are highlighted with colored symbols. Small grey dots indicate scaffolds that could not be unambiguously classified into major groups.



Supplementary Figure S4. Rank abundance of assembled microbial populations based on depth of coverage.

Supporting Table S1. Water chemistry of Lake Tyrrell sampling site, located at 35°19' 12.24S 142°48' 00.45E.

collection date	water temp (°C)	Salinity (ppt)	TDS (wt%)	pH
1/23/07	22	127	31	7.23
1/25/07	28	132	31	7.09

Supporting Table S2. Summary of 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
Sanger Total					426,978	727 \pm 180
ABT	0.8 μ m	1/23/2007	-	454 Titanium	994,337	362 \pm 96
CAT	0.1 μ m	1/25/2007	-	454 Titanium	205,925	401 \pm 119
454 total					1,200,262	368 \pm 100

Supporting Table S3. Assembly statistics for combined Sanger metagenomic libraries using Celera Assembler version 5.4. Assembly parameters used were as follows: 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.

total num nucleotides	286,258,541
num reads	426,864
num scaffolds	15,008
scaffold N50	29,346
max scaffold length	2,982,959
num predicted proteins	62,915
num placed reads	282,583
% bp placed	69.60%
% reads placed	66.20%

Supporting Table S4. Assembled 16S rRNA sequences and their closest database matches to environmental clones and cultured isolates. Matches were required to have BLAST alignments to previously identified 16S rRNA genes of 450 nt or longer, with e-value < 1e-7 and 80% or greater sequence identity between query and subject. Part A shows 16S rRNA gene sequences obtained in targeted genomic assemblies. Part B shows additional 16S rRNA gene sequences observed in scaffolds obtained by composite assembly of all Sanger reads.

A

Genome	Scaffold id	16S sequence length	Closest Environmental match	Environ match pct identity	Closest cultured isolate match	Isolate match pct identity
J07HQW1 copy 1	scf56329	1472	GQ374993 - Baj_clone41 (Australia)	99.8	Haloquadratum walsbyi DSM 16790	99.1
J07HQW1 copy 2	scf56329	1472	GQ374996 - Baj_clone47 (Australia)	99.0	Haloquadratum walsbyi DSM 16790	98.6
J07HQW2 copy 1	scf02443	1473	GQ374961 - Cry7_clone4 (Australia)	99.1	Haloquadratum walsbyi DSM 16790	97.5
J07HQW2 copy 2	scf02443	709	GQ374961 - Cry7_clone4 (Australia)	99.3	Haloquadratum walsbyi DSM 16790	97.6
J07HQX50	scf01540	1473	GQ375020 - Cry7_clone100 (Australia)	96.9	Haloquadratum walsbyi DSM 16790	93.5
J07AB43	scf30744	1479	GQ375009 - Cry7_clone76 (Australia)	99.4	none	-
J07AB56	scf39101	1486	GQ375000 - Cry7_clone54 (Australia)	99.5	none	-
J07HN4	scf01348	1469	GQ374992 - Baj_clone38 (Australia)	96.1	Halonotus pteroides 302698938	96.1
J07HN6	scf01851	1468	FN994966 - SFF1C121 (Tunisia)	97.9	Halonotus pteroides 302698938	95.4
J07HR59	scf01382	1470	GQ375002 - Cry7_clone60 (Australia)	97.4	Halonotus tebenquichense JCM12290	90.6
J07HX5	scf22092	1468	DQ432537 - WN-FWA-186 (Egypt)	96.0	Halosimplex carlsbadense	91.0
J07HX64	scf96483	1010	GQ375005 - Cry7_clone67	98.0	Halorhabdus utahensis AX-2	92.3

Genome	Scaffold id	16S sequence length	Closest Environmental match	Environ match pct identity	Closest cultured isolate match	Isolate match pct identity
			(Australia)			
J07HB67	scf98878	1360	GQ374972 - Cry7_clone43 (Australia)	99.0	Halobaculum gomorrense str JCM 9908	90.0
J07SB67	scf87461	1495	HQ157666 - TSHNAE18 (Tunisia)	98.2	Salinibacter ruber DSM 13855	97.3

B

Taxonomic group	Scaffold id	16S sequence length	Closest Environmental match	Environ match pct identity	Closest cultured isolate match	Isolate match pct identity
Haloquadratum	scf98861	1474	GQ374961 - Cry7_clone4 (Australia)	99.2	Haloquadratum walsbyi 640715938	98.0
J07AB43	scf98840	1448	GQ375009 - Cry7_clone76 (Australia)	99.7	none	-
J07AB56	scf98837	1383	GQ375000 - Cry7_clone54 (Australia)	99.9	none	-
J07HN4	scf98475	1352	GQ374992 - Baj_clone38 (Australia)	97.3	Halonotius pteroides 302698938	97.1
Haloarcula	scf91856	1295	EU722673 - SA21 (Spain)	97.3	Haloarcula hispanica str HLR2	96.6
J07AB43	scf86038	1039	GQ375009 - Cry7_clone76 (Australia)	95.6	none	-
J07HR59	scf91079	1038	GQ375002 - Cry7_clone60 (Australia)	98.1	Halorubrum tebenquichense JCM12290	92.2
Halorubrum	scf91214	839	GQ375031 - LDS1_clone58 (Australia)	98.5	Halorubrum tebenquichense	96.2
J07HN4	scf98871	796	GQ374959 - Baj_clone12 (Australia)	99.4	Halonotius pteroides 302698938	95.2
J07HX5	scf93563	793	GQ375002 - Cry7_clone60 (Australia)	94.4	Natronomonas pharaonis DSM 2160	90.6

Taxonomic group	Scaffold id	16S sequence length	Closest Environmental match	Environ match pct identity	Closest cultured isolate match	Isolate match pct identity
J07HX5	scf98852	780	GQ375002 - Cry7_clone60 (Australia)	94.0	Halomicrobium mukohataei DSM 12286	91.4
HQX50	scf98854	738	GQ375020 - Cry7_clone100 (Australia)	99.9	Haloquadratum walsbyi DSM 16790	92.0
Haloquadratum	scf98861	712	GQ374961 - Cry7_clone4 (Australia)	99.2	Haloquadratum walsbyi DSM 16790	97.5
HQX50	scf98854	653	GQ374945 - LDS1_clone37 (Australia)	99.7	Haloquadratum walsbyi DSM 16790	95.1
Unclassified	scf93739	650	GQ375041 - LDS1D3.3 (Australia)	100.0	Halorientalis regularis strain TBN19	96.0
Haloarcula	scf84445	631	JN714439 - XKL43 (China)	96.8	Haloarcula hispanica str HLR2	95.5
J07HN6	scf98809	574	FN994966 - SFF1C121 (Tunisia)	99.5	Halonotius pteroides 302698938	95.0
Haloquadratum	scf98859	560	GQ374990 - Baj_clone34 (Australia)	99.8	Haloquadratum walsbyi DSM 16790	99.1
Haloquadratum	scf98874	512	GQ374952 - Baj_clone1 (Australia)	99.6	Haloquadratum walsbyi DSM 16790	98.4
J07HXX64	scf98875	480	GQ375005 - Cry7_clone67 (Australia)	98.1	Haloferax zhejiangensis ZJ206	92.1

Supporting Table S5. Distinctive properties of major scaffold clusters. Percentages are based on taxonomic classifications of all predicted protein tophit matches to Genbank nr, as determined using the DarkHorse algorithm at a filter threshold setting of 0.05, including only alignments covering at least 70% of both query and target sequences with an e-value of 1e-5 or better.

group	pct G+C	pct reads 0.1 fit	avg read depth	pct Halo- quadrataum	pct Nano- haloarchaea	pct Halo- rabdus	pct Halo- ferax	pct Halo- rubrum	pct Salini- bacter	pct LYS	pct THR	pct ALA	pct GLU
Haloquadratum	48	20%	8.6	79%	0%	2%	1%	1%	0%	2.3	7.5	9.0	7.7
HQX50	50	20%	2.5	44%	1%	5%	7%	2%	0%	1.4	4.3	5.6	4.8
Halonotus HN4	61	59%	3.8	10%	1%	12%	15%	12%	0%	1.7	6.5	10.1	7.2
Halonotus HN6	63	81%	4.6	9%	0%	12%	17%	12%	0%	1.7	6.2	9.9	7.2
Nanosalina J07AB43	43	94%	6.4	1%	60%	3%	1%	2%	0%	6.0	5.0	5.8	10.5
Nanosalarium J07AB56	56	91%	4.7	1%	53%	4%	0%	1%	0%	4.0	5.3	7.1	10.2
HX5	60	21%	2.9	5%	1%	28%	2%	3%	0%	1.4	6.3	9.8	6.1
HX64	64	24%	3.6	5%	1%	27%	3%	4%	0%	1.3	6.6	9.4	7.9
HB67	67	33%	2.9	10%	1%	12%	14%	8%	0%	1.3	7.5	11.4	7.7
Halorubrum	60	25%	2.4	7%	0%	7%	14%	32%	0%	0.9	3.1	4.9	3.7
Salinibacter	66	44%	2.2	0%	0%	0%	0%	0%	92%	0.8	2.7	4.4	3.0

Function id	Function name	J07HQW1	J07HQW2	J07HQS50	J07HN4	J07HN6	J07HR59	J07HXS64	J07HXS	J07HBS67
	polymerase, beta' subunit/160 kD subunit									
COG0087	Ribosomal protein L3	+	+	+	+	+	-	+	-	+
COG0088	Ribosomal protein L4	+	+	+	+	+	+	+	-	+
COG0090	Ribosomal protein L2	+	+	+	+	+	+	+	-	+
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	+	+	+	+	+	+	+	+	+
COG0124	Histidyl-tRNA synthetase	+	+	+	+	+	+	+	+	+
COG0130	Pseudouridine synthase	+	+	+	+	+	+	+	+	+
COG0143	Methionyl-tRNA synthetase	+	+	+	+	+	+	+	+	+
COG0164	Ribonuclease HII	+	+	+	+	+	+	+	-	+
COG0177	Predicted EndoIII-related endonuclease	+	+	+	+	+	+	+	+	+
COG0180	Tryptophanyl-tRNA synthetase	+	+	+	+	+	+	+	+	+
COG0185	Ribosomal protein S19	+	+	-	+	+	+	+	-	-
COG0186	Ribosomal protein S17	+	+	+	+	+	+	-	-	-
COG0197	Ribosomal protein L16/L10E	+	+	+	+	+	-	+	+	-

Function id	Function name	J07HQW1	J07HQW2	J07HQS50	J07HN4	J07HN6	J07HR59	J07HXS64	J07HXS	J07HBS67
COG0200	Ribosomal protein L15	+	+	+	+	+	+	+	-	+
COG0201	Preprotein translocase subunit SecY	+	+	+	+	+	+	+	-	+
COG0250	Transcription antiterminator	+	+	-	+	+	+	+	+	+
COG0256	Ribosomal protein L18	+	+	-	+	+	+	+	-	+
COG0361	Translation initiation factor 1 (IF-1)	+	+	+	+	+	+	+	+	+
COG0441	Threonyl-tRNA synthetase	+	+	+	+	+	+	+	+	+
COG0455	ATPases involved in chromosome partitioning	+	-	+	+	+	-	+	-	+
COG0459	Chaperonin GroEL (HSP60 family)	+	+	+	+	+	+	+	+	+
COG0468	RecA/RadA recombinase	+	+	+	+	+	+	+	+	+
COG0480	Translation elongation factors (GTPases)	+	+	+	+	+	+	+	-	+
COG0495	Leucyl-tRNA synthetase	+	+	+	+	+	+	+	+	+
COG0522	Ribosomal protein S4 and related proteins	+	+	+	+	+	+	+	-	+
COG0525	Valyl-tRNA synthetase	+	+	+	+	+	+	+	+	+
COG0532	Translation initiation factor 2 (IF-2; GTPase)	+	+	+	+	+	-	-	+	+