Title

Complete genome sequence of Halogeometricum borinquense type strain (PR3T)

Journal

Environmental Microbiome, 1(2)

ISSN

2524-6372

Authors

Malfatti, Stephanie
Tindall, Brian J
Schneider, Susanne
et al.

Publication Date

2009-09-01

DOI

10.4056/sigs.23264

Peer reviewed

Complete genome sequence of *Halogeometricum borinquense* type strain (PR3^T)

Stephanie Malfatti[1,2], Brian J. Tindall[3], Susanne Schneider[3], Regine Fähnrich[3], Alla Lapidus[1], Kurt LaButtii[1], Alex Copeland[1], Tijana Glavina Del Rio[1], Matt Nolan[1], Feng Chen[1], Susan Lucas[1], Hope Tice[1], Jan-Fang Cheng[1], David Bruce[1,4], Lynne Goodwin[1,4], Sam Pitluck[1], Iain Anderson[1], Amrita Pati[1], Natalia Ivanova[1], Konstantinos Mavromatis[1], Amy Chen[5], Krishna Palaniappan[5], Patrik D'haeseleer[1,2], Markus Göker[3], Jim Bristow[1], Jonathan A. Eisen[1,6], Victor Markowitz[5], Philip Hugenholtz[1], Nikos C. Kyrpides[1], Hans-Peter Klenk[3], and Patrick Chain[1,2*]

[1] DOE Joint Genome Institute, Walnut Creek, California, USA
[2] Lawrence Livermore National Laboratory, Livermore, California, USA
[3] DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany
[4] Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA
[5] Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[6] University of California Davis Genome Center, Davis, California, USA

*__Corresponding author:__ Patrick Chain

Keywords: halophile, free-living, non-pathogenic, aerobic, pleomorphic cells, euryarchaeon

Halogeometricum borinquense Montalvo-Rodríguez *et al*. 1998 is the type species of the genus, and is of phylogenetic interest because of its distinct location between the *halobacterial* genera *Haloquadratum* and *Halosarcina*. *H. borinquense* requires extremely high salt (NaCl) concentrations for growth. It can not only grow aerobically but also anaerobically using nitrate as electron acceptor. The strain described in this report is a free-living, motile, pleomorphic, euryarchaeon, which was originally isolated from the solar salterns of Cabo Rojo, Puerto Rico. Here we describe the features of this organism, together with the complete genome sequence, and annotation. This is the first complete genome sequence of the halobacterial genus *Halogeometricum*, and this 3,944,467 bp long six replicon genome with its 3937 protein-coding and 57 RNA genes is part of the ***G**enomic **E**ncyclopedia of **B**acteria and **A**rchaea* project.

Introduction

Strain PR3^T (= DSM 11551 = ATCC 700274 = JCM 10706) is the type strain of *Halogeometricum borinquense*, representing the sole species of the genus *Halogeometricum* [1]. Strain PR3^T was first described by Montalvo-Rodríguez *et al*. in 1998 [1] as Gram-stain negative and motile. The organism is of interest because of its position in the tree of life, where it is located between members of the *Haloferax/Halorubrum* cluster within the large euryarchaeal family *Halobacteraceae*(Figure 1). Here we present a summary classification and a set of features for *H. geometricum* PR3^T together with the description of the complete genomic sequencing and annotation.

Classification and features

In addition to the solar salterns of Cabo Rojo, Puerto Rico, where the type strain PR3^T and two accompanying strains (PR7 and PR9) were initially isolated [1], the occurrence of strains or phylotypes closely related or belonging to *H. borinquense* have so far only been reported from high salt environments such as an Australian crystallizer pond [6], Maras Salterns in the Peruvian Andes [7], a salt field at Nie, Ishikawa Prefecture, Japan [8], the salterns of Tamilnadu, India (Kannan *et al*. unpublished), Exportadora del Sal, Guerro Negro, Mexico (FJ609942), a Taiwanese saltern soil (FJ348412), and a low-salt, sulfide- and sulfur-rich spring in southwestern Oklahoma, USA [9].

H. geometricum PR3^T cells are highly pleomorphic (short and long rods, squares, triangles and ovals) and motile by peritrichous flagella (Table 1 and Figure 2). Cells lyse in distilled water. Gas vesicles are present and are responsible for modifying the color of colonies or cell suspensions from red to pink. *H. geometricum* PR3^T is aerobic, but also capable of anaerobic growth with nitrate. No anaerobic growth on arginine (arginine dihydrolase is not present). At least 8% NaCl (w/v) is required for growth, reflecting the primary characteristic requirement for high salt concentrations of the *Halobacteriaceae* [18]. The optimal NaCl concentration range is 20-25**%** NaCl (w/v) at 40°C (optimal growth temperature). Nitrate is reduced to nitrite with the production of gas [1]. Spores or other resting stages have not been reported [1].

H. geometricum PR3^T is capable of degrading gelatin, but starch is not hydrolysed. A number of sugars and polyols are used as carbon sources, and acid is produced from some sugars [1].

Figure 1 shows the phylogenetic neighborhood of *H. borinquense* strain PR3^T in a 16S rRNA based tree. Analysis of the two 16S rRNA gene sequences in the genome of strain PR3^T indicated that the two genes differ by five nucleotides (nts) from each other, and by 3-5 nts from the previously published 16S rRNA sequence generated from DSM 11551 (AF002984). The slight differences between the genome data and the reported 16S rRNA gene sequence are most likely the result of sequencing errors in the previously reported sequence data.

Chemotaxonomy

The quinone composition of *H. borinquense* strain PR3^T has not been recorded, but based on reports from other members of the family *Halobacteriaceae* menaquinones with eight isoprenoid units are likely to be present. Typically both MK-8 and MK-8 (VIII-H_2) are predicted. The lipids are based on isoprenoid diether lipids, but the exact nature of the isoprenoid side chains remains to be investigated. The major phospholipids are the diether, isoprenoid analogs of phosphatidylglycerol and methyl-phosphatidylglycerophosphate (typical of all members of the family *Halobacteriaceae*), the diether analog of phosphatidyl-glycerol sulfate is absent [1]. A single glycolipid has been reported with an R_f value similar to that of the triglycosyl diether from *Haloarcula marismortui*, but its structure has not been determined [1]. The pigments responsible for the red color of the cells have not been determined, but it may be predicted that they are carotenoids, probably bacterioruberins. Outer cell layers are probably proteinaceous. The presence of peptidoglycan has not been investigated, but is generally absent from members of this family *Halobacteriaceae.*

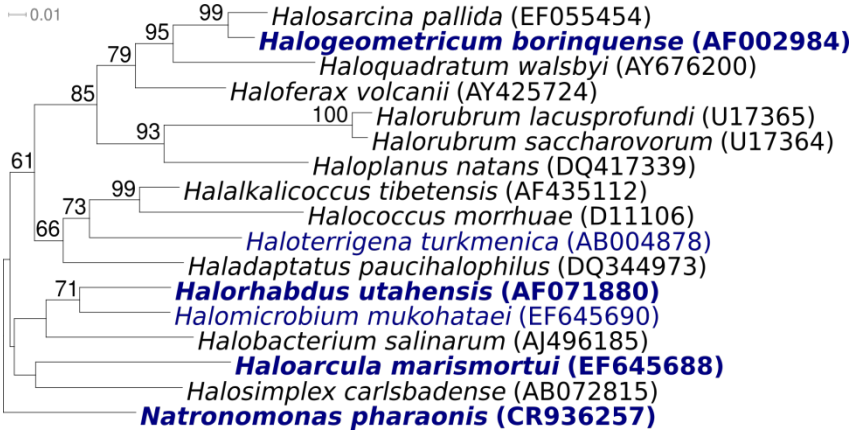

Figure 1. Phylogenetic tree of *H. borinquense* PR3^T with a selection of type strains of the family *Halobacteriaceae*, inferred from 1,433 aligned 16S rRNA characters [2] under the maximum likelihood criterion [3,4]. The tree was rooted with *Natronomonas pharaonsis*, the deepest branching member of the family *Halobacteriaceae*. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches are support values from 1,000 bootstrap replicates. Strains with a genome sequencing project registered in GOLD [5] are printed in blue; published genomes in bold.

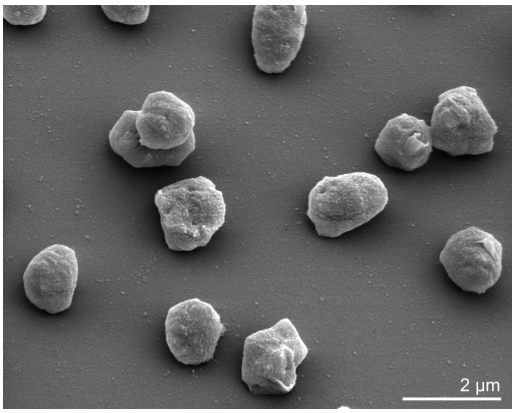

Figure 2. Scanning electron micrograph of *H. borinquense* PR3^T (Manfred Rohde, Helmholtz Centre for Infection Research, Braunschweig)

Table 1. Classification and general features of *H. borinquense* PR3^T according to the MIGS recommendations [10]

MIGS ID	Property	Term	Evidence code
	Current classification	Domain *Archaea*	TAS [11]
		Phylum *Euryarchaeota*	TAS [12]
		Class *Halobacteria*	TAS [13]
		Order *Halobacteriales*	TAS [14]
		Family *Halobacteriaceae*	TAS [15]
		Genus *Halogeometricum*	TAS [1]
		Species *Halogeometricum borquinense*	TAS [1]
		Type strain PR3	TAS [1]
	Gram stain	negative	TAS [1]
	Cell shape	highly pleomorphic	TAS [1]
	Motility	motile	TAS [1]
	Sporulation	non-sporulating	NAS
	Temperature range	mesophile, between 22°C and 50°C	TAS [1]
	Optimum temperature	40°C	TAS [1]
	Salinity	halophile, at least 8% (w/v) NaCl	TAS [1]
MIGS-22	Oxygen requirement	primarily aerobic; facultatively anaerobic growth *via* nitrate reduction	TAS [1]
	Carbon source	glucose, mannose, fructose, xylose, maltose, trehalose, cellobiose, raffinose, glycerol	TAS [1]
	Energy source	carbohydrates	TAS [1]
MIGS-6	Habitat	aquatic	TAS [1]
MIGS-15	Biotic relationship	free living	NAS
MIGS-14	Pathogenicity	none	NAS
	Biosafety level	1	TAS [16]
	Isolation	solar salterns of Cabo Rojo, Puerto Rico	TAS [1]
MIGS-4	Geographic location	Cabo Rojo, Puerto Rico	TAS [1]
MIGS-5	Sample collection time	1994	TAS [1]
MIGS-4.1	Latitude / Longitude	18,088 / -67,147	TAS [1]
MIGS-4.2			
MIGS-4.3	Depth	not reported	
MIGS-4.4	Altitude	sea level	NAS

Evidence codes - IDA: Inferred from Direct Assay (first time in publication); TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [17]. If the evidence code is IDA then the property was directly observed for a living isolate by one of the authors or an expert mentioned in the acknowledgements.

Genome sequencing and annotation

Genome project history

This organism was selected for sequencing on the basis of each phylogenetic position, and is part of the ***G**enomic **E**ncyclopedia of **B**acteria and **A**rchaea* project. The genome project is deposited in the Genome OnLine Database [5]. The complete genome sequence has not yet been released from GenBank. Sequencing, finishing and annotation were performed by the DOE Joint Genome Institute (JGI). A summary of the project information is shown in Table 2.

Table 2. Genome sequencing project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Finished
MIGS-28	Libraries used	Two genomic libraries: 8kb pMCL200 and fosmid pcc1Fos Sanger libraries. One 454 pyrosequence standard library.
MIGS-29	Sequencing platforms	ABI3730, 454 GS FLX
MIGS-31.2	Sequencing coverage	9.7× Sanger; 21.8× pyrosequencing
MIGS-30	Assemblers	Newbler, PGA
MIGS-32	Gene calling method	GeneMark 4.6b, tRNAScan-SE-1.23, infernal 0.81
	INSDC / Genbank ID	CP001688
	Genbank Date of Release	September 10, 2009
	GOLD ID	Gc01108
	NCBI project ID	20743
	Database: IMG-GEBA	2501416934
MIGS-13	Source material identifier	DSM 11551
	Project relevance	Tree of Life, GEBA

Growth conditions and DNA isolation

H. borinquense PR3^T, DSM 11551, was grown in DSMZ medium 372 (*Halobacteria* Medium) at 35°C [19]. DNA was isolated from 1-1.5 g of cell paste using a Qiagen Genomic 500 DNA Kit (Qiagen, Hilden, Germany) with a modified protocol for cell lysis, LALMP procedure according to Wu *et al*. [20]..

Genome sequencing and assembly

The genome was sequenced using a combination of Sanger and 454 sequencing platforms. All general aspects of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov/. 454 Pyrosequencing reads were assembled using the Newbler assembler version v 2.0.0 (Roche). Large Newbler contigs were broken into 4,435 overlapping fragments of 1,000 bp and entered into assembly as pseudo-reads. The sequences were assigned quality scores based on Newbler consensus q-scores with modifications to account for overlap redundancy and adjust inflated q-scores. A hybrid 454/Sanger assembly was made using the PGA assembler. Possible mis-assemblies were corrected and gaps between contigs were closed by custom primer walks from sub-clones or PCR products. A total of 2,826 Sanger finishing reads were produced. The error rate of the completed genome sequence is less than 1 in 100,000. Together all sequence types provided 31.5× coverage of the genome.

Genome annotation

Genes were identified using GeneMark [21] as part of the genome annotation pipeline in the Integrated Microbial Genomes Expert Review (IMG-ER) system [22], followed by a round of manual curation using the JGI GenePRIMP pipeline [23]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. The tRNAScanSE tool [24] was used to find tRNA genes, whereas ribosomal RNAs were found by using the tool RNAmmer [25]. Other non coding RNAs were identified by searching the ge-

nome for the Rfam profiles using INFERNAL (v0.81) [26]. Additional gene prediction analysis and manual functional annotation was performed within the Integrated Microbial Genomes (IMG) platform [27].

Metabolic network analysis

The metabolic Pathway/Genome Database (PGDB) was computationally generated using Pathway Tools software version 12.5 [28] and MetaCyc version 12.5 [29], based on annotated EC numbers and a customized enzyme name mapping file. It has undergone no subsequent manual curation and may contain errors, similar to a Tier 3 BioCyc PGDB [30].

Genome properties

The genome is 3,944,467 bp long and comprises one main circular chromosome with a 60% GC content and five plasmids. Of the 3,994 genes predicted, 3,937 were protein coding genes, and 57 RNAs. Thirty seven pseudogenes were also identified. A total of 62% of the genes were assigned a putative function while the remaining ones are annotated as hypothetical proteins. The properties and the statistics of the genome are summarized in Table 3. The distribution of genes into COGs functional categories is presented in Figure 3 and Table 4. A cellular overview diagram is presented in Figure 4, followed by a summary of metabolic network statistics shown in Table 5.

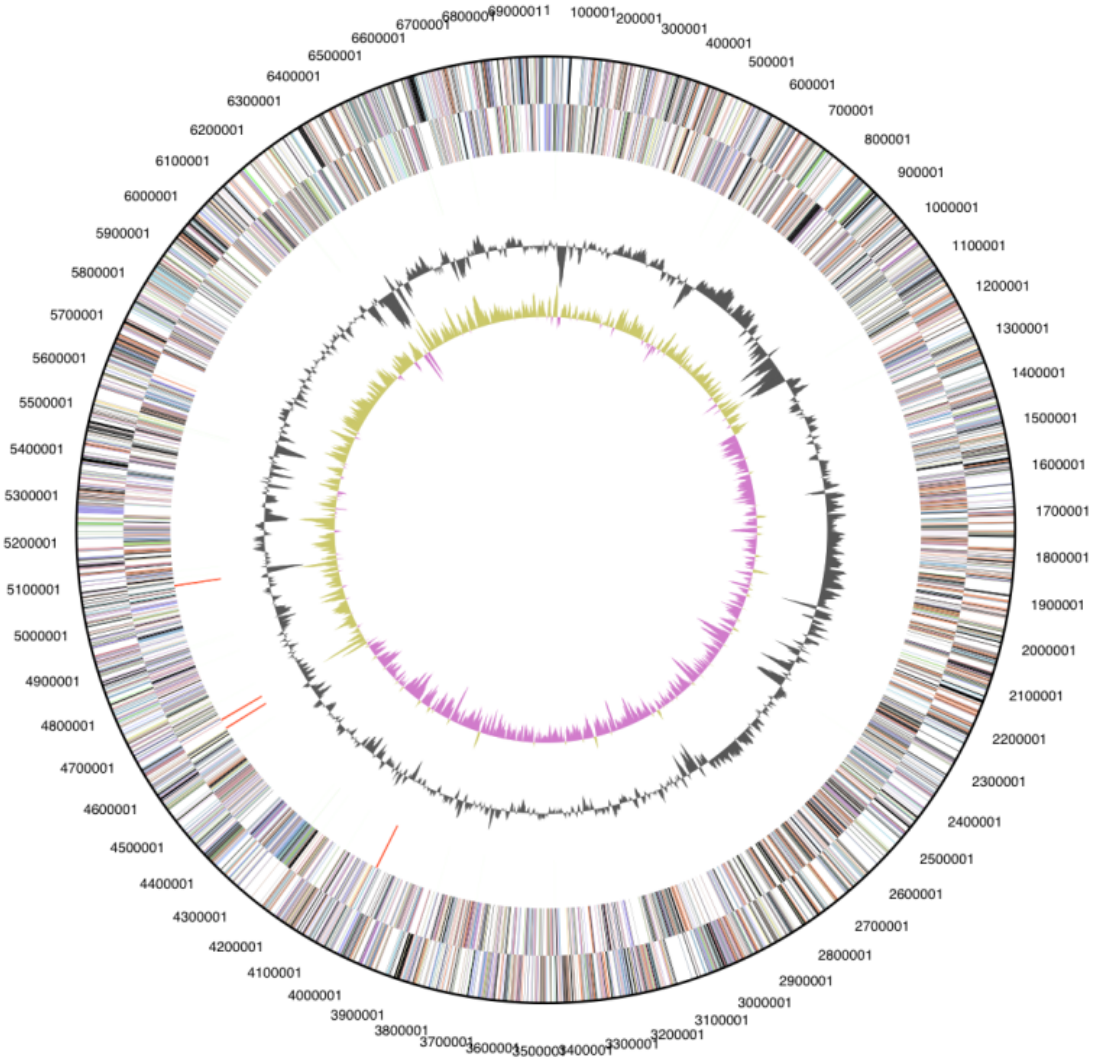

Figure 3. Graphical circular map of the genome. From outside to the center: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew.

Table 3. Genome Statistics

Attribute	Value	% of Total
Genome size (bp)	3,944,467	100.00%
DNA Coding region (bp)	3,441,571	87.25%
DNA G+C content (bp)	2,364,339	59.94%
Number of replicons	1	
Extrachromosomal elements	5	
Total genes	3994	100.00%
RNA genes	57	1.90%
rRNA operons	2	
Protein-coding genes	3937	98.57%
Pseudogenes	37	0.93%
Genes with function prediction	2486	62.24%
Genes in paralog clusters	741	18.55%
Genes assigned to COGs	2449	61.32%
Genes assigned Pfam domains	2385	59.71%
Genes with signal peptides	533	13.35%
Genes with transmembrane helices	971	24.31%
CRISPR repeats	1	

Table 4. Number of genes associated with the general COG functional categories

Code	Value	% age	Description
J	162	4.1	Translation, ribosomal structure and biogenesis
A	1	0.0	RNA processing and modification
K	140	3.6	Transcription
L	138	3.5	Replication, recombination and repair
B	3	0.0	Chromatin structure and dynamics
D	0	0.1	Cell cycle control, mitosis and meiosis
Y	0	0.0	Nuclear structure
V	46	1.2	Defense mechanisms
T	113	2.8	Signal transduction mechanisms
M	87	2.2	Cell wall/membrane biogenesis
N	38	0.1	Cell motility
Z	0	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	27	0.7	Intracellular trafficking and secretion
O	123	3.1	Posttranslational modification, protein turnover, chaperones
C	174	4.4	Energy production and conversion
G	124	3.1	Carbohydrate transport and metabolism
E	271	6.9	Amino acid transport and metabolism
F	77	1.9	Nucleotide transport and metabolism
H	140	3.5	Coenzyme transport and metabolism
I	98	2.5	Lipid transport and metabolism
P	178	4.5	Inorganic ion transport and metabolism
Q	60	1.5	Secondary metabolites biosynthesis, transport and catabolism
R	433	11.0	General function prediction only
S	227	5.8	Function unknown
-	1488	37.8	Not in COGs

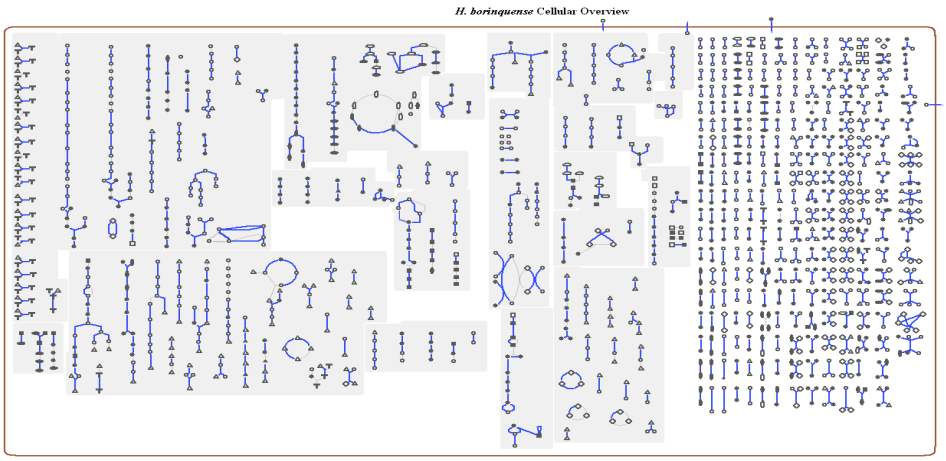

Figure 4. Schematic cellular overview diagram of all pathways of *H. borinquense* strain PR3[T]. Nodes represent metabolites, with shape indicating class of metabolite. Lines represent reactions.

Table 5. Metabolic Network Statistics

Attribute	**Valu**
Total genes	**3801**
Enzymes	578
Enzymatic reactions	687
Metabolic pathways	125
Metabolites	578

Acknowledgements

This work was performed under the auspices of the US Department of Energy 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, as well as German Research Foundation (DFG) INST 599/1-1.

References

1. Montalvo-Rodríguez R, Vreeland RH, Oren A, Kessel M, Betancourt C, López-Garriga J. *Halogeometricum borinquense* gen. nov., sp. nov., a novel halophilic archaeon from Puerto Rico. *Int J Syst Bacteriol* 1998; **48**:1305-1312. PubMed

2. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics* 2002; **18**:452-464. PubMed doi:10.1093/bioinformatics/18.3.452

3. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 1981; **17**:368-376. PubMed doi:10.1007/BF01734359

4. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web-servers. *Syst Biol* 2008; **57**:758-771. PubMed doi:10.1080/10635150802429642

5. Liolios K, Mavromatis K, Tavernarakis N, Kyrpides NC. The Genomes OnLine Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2008; **36**:D475-D479. PubMed doi:10.1093/nar/gkm884

6. Burns DG, Camakaris HM, Janssen PH, Dyall-Smith ML. Combined use of cultivation-dependent and cultivation-independent methods indicates that members of most haloarchaeal groups in an Australian crystallizer pond are cultivable. *Appl Environ Microbiol* 2004; **70**:5258-5265. PubMed PubMed doi:10.1128/AEM.70.9.5258-5265.2004

7. Maturrano L, Santos F, Rosselló-Mora R, Antón J. Microbial diversity in Maras Salterns, a hypersaline environment in the Preuvian Andes. *Appl Environ Microbiol* 2006; **72**:3887-3895. PubMed PubMed doi:10.1128/AEM.02214-05

8. Fukushima T, Usami R, Kamekura M. A traditional Japanese-style salt field is a nich for haloarchaeal strains that can survive in 0.5% salt solution. *Saline Systems* 2007; **72**:3887-3895.

9. Elshahed MS, Najar FZ, Roe BA, Oren A, Dewers TA, Krumholz LR. Survey of archaeal diversity reveals an abundance of halophilic archaea in a

low-salt, sulphide- and sulphur-rich spring. *Appl Environ Microbiol* 2004; **70**:2230-2239. PubMed doi:10.1128/AEM.70.4.2230-2239.2004

10. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV *et al*. Towards a richer description of our complete collection of genomes and metagenomes: the "Minimum Information about a Genome Sequence" (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. PubMed doi:10.1038/nbt1360

11. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea, Bacteria*, and *Eucarya*. *Proc Natl Acad Sci USA* 1990; **87**: 4576-4579. PubMed doi:10.1073/pnas.87.12.4576

12. Garrity GM, Lilburn TG, Cole JR, Harrison SH, Euzéby J, Tindall BJ. Taxonomic Outline of the *Bacteria* and *Archaea*. Release 7.7 March 6, 2007. Michigan State University Board of Trustees. DOI: 10.1601/TOBA7.7. http://www.taxonomicoutline.org/index.php/toba

13. Grant WD, Kamekura M, McGenity TJ, Ventosa A. Class III. *Halobacteria* class. nov. *In*: DR Boone, RW Castenholz, GM Garrity (eds): *Bergey's Manual of Systematic Bacteriology*, 2nd edition, Vol. 1, The *Archaea* and the deeply branching and phototrophic *Bacteria*, Springer-Verlag, New York, 2001.p 294.

14. Grant WD, Larsen H. Group III. Extremely halophilic archaeobacteria. Order *Halobacteriales* ord. nov. *In*: JT Staley, MP Bryant, N. Pfennig, JG Holt (eds), Bergey's Manual of Systematic Bacteriology, 1st edition, Vol. 3, The Williams & Wilkins Co., Baltimore, 1989, pp. 2216-2218.

15. Gibbons NE. Family V. *Halobacteriaceae* fam. nov. *In*: RE Buchanan, NE Gibbons (eds): *Bergey's Manual of Determinative Bacteriology*, 8th edition, The Williams & Wilkins Co, Baltimore, 1974, p. 269.

16. Anonymous. Biological Agents: Technical rules for biological agents www.baua.de TRBA 466.

17. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al*. The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nat Genet* 2000; **25**:25-29. PubMed doi:10.1038/75556

18. Norton C. Rediscovering the ecology of halobacteria. *ASM News* 1992; **58**:363-367.

19. List of media used at DSMZ for cell growth: http://www.dsmz.de/microorganisms/media_list.php

20. Wu M, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova N, Kunin V, Goodwin L, Wu M, Tindall BJ, *et al*.. A phylogeny-driven genomic encyclopedia of Bacteria and Archaea. *Nature* 2009 (In press)

21. Besemer J, Lomsadze A, Borodovsky M. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* 2001; **29**:2607-2618. PubMed PubMed doi:10.1093/nar/29.12.2607

22. Markowitz VM, Mavromatis K, Ivanova NN, Chen IMA, Chu K, Kyrpides NC. *Expert Review of Functional Annotations for Microbial Genomes. Bioinformatics* 2009 **25**:2271-2278.

23 Pati A, Ivanova N, Mikhailova, N, Ovchinikova G, Hooper SD, Lykidis A, Kyrpides NC. GenePRIMP: A Gene Prediction Improvement Pipeline for microbial genomes. (Submitted) 2009.

24. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997; **25**:955-964. PubMed doi:10.1093/nar/25.5.955

25. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 2007; **35**:3100-3108. PubMed doi:10.1093/nar/gkm160

26. Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* 2005; **33**:D121-D124. PubMed doi:10.1093/nar/gki081

27. Markowitz VM, Szeto E, Palaniappan K, Grechkin Y, Chu K, Chen IMA, Dubchak I, Anderson I, Lykidis A, Mavromatis K, *et al*. The Integrated Microbial Genomes (IMG) system in 2007: data content and analysis tool extensions. *Nucleic Acids Res* 2008; **36**:D528-D533. PubMed doi:10.1093/nar/gkm846

28. Karp PD, Paley SM, Romero P. The Pathway Tools Software. *Bioinformatics* 2002; **18**:S225-S232. PubMed

29. Caspi R, Karp P, Foerster H, Fulcher CA, Kaipa P, Krummenacker M, Latendresse M, Paley SM, Rhee SY, Shearer A, *et al*. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of pathway/Genome Databases. *Nucleic*

Acids Res 2008; **36**:D623-D631. PubMed PubMed doi:10.1093/nar/gkm900

30. Karp PD, Ouzounis CA, Moore-Kochlacs C, Goldovsky L, Kaipa P, Ahren D, Tsoka S, Darzentas N, Kunin V, Lopez-Bigas N. Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res* 2005; **33**:6083-6089. PubMed doi:10.1093/nar/gki892