

Complete genome sequence of *Pyrobaculum oguniense*

David L. Bernick¹, Kevin Karplus¹, Lauren M. Lui¹, Joanna K. C. Coker¹, Julie N. Murphy¹, Patricia P. Chan¹, Aaron E. Cozen¹, Todd M. Lowe¹

¹Biomolecular Engineering, University of California., Santa Cruz, California, USA

Corresponding author: Todd M. Lowe (lowe@soe.ucsc.edu)

Keywords: *Pyrobaculum oguniense*, *Pyrobaculum arsenaticum*, Crenarchaea, inversion

Pyrobaculum oguniense TE7 is an aerobic hyperthermophilic crenarchaeon isolated from a hot spring in Japan. Here we describe its main chromosome of 2,436,033 bp, with three large-scale inversions and an extra-chromosomal element of 16,887 bp. We have annotated 2,800 protein-coding genes and 145 RNA genes in this genome, including nine H/ACA-like small RNA, 83 predicted C/D box small RNA, and 47 transfer RNA genes. Comparative analyses with the closest known relative, the anaerobe *Pyrobaculum arsenaticum* from Italy, reveals unexpectedly high synteny and nucleotide identity between these two geographically distant species. Deep sequencing of a mixture of genomic DNA from multiple cells has illuminated some of the genome dynamics potentially shared with other species in this genus.

Introduction

Pyrobaculum oguniense TE7^T (=DSMZ 13380=JCM10595) was originally isolated from the Tsuetate hot spring in Oguni-cho, Kumamoto Prefecture, Japan [1], and subsequently found to grow heterotrophically at an optimal temperature near 94°C, pH 7.0 (at 25°C), and in the presence or absence of oxygen. Under anaerobic conditions, it can utilize sulfur-containing compounds (sulfur, thiosulfate, L-cystine and oxidized glutathione) but not nitrate or nitrite as terminal electron acceptors.

Initial 16S ribosomal DNA sequence analysis [1] placed *Pyrobaculum oguniense* TE7^T in the *Pyrobaculum* clade and closest to *P. aerophilum* and *Thermoproteus neutrophilus* (recently renamed to *Pyrobaculum neutrophilum* [2]). DNA hybridization studies were conducted with *P. aerophilum* IM2, *P. islandicum* GEO3, *P. organotrophum* H10 and *T. neutrophilus* (*P. neutrophilum*) V24Sta, showing little genomic similarity to those species. *P. arsenaticum* PZ6^T [3], *P. sp.* 1860 [4] and *P. calidifontis* VA1 [5] were not available at that time.

The genus *Pyrobaculum* is known for its range of respiratory capabilities [6]. Three of the currently known members of the genus can respire oxygen; *P. aerophilum* is a facultative micro-aerobe, while *P. calidifontis* and *P. oguniense* can utilize atmospheric oxygen. *P. aerophilum* [7], *P. calidifontis*, and four other metabolically unique *Pyrobaculum* species have been fully sequenced; together with *P.*

oguniense, we sought to further broaden the understanding of this important hyperthermophilic group. Pairwise whole-genome alignments of previously sequenced *Pyrobaculum* species reveal many structural rearrangements. With the availability of high-throughput sequencing, we were able to further explore rearrangements that occur between species, and our use of a not-quite-clonal population allowed exploration of rearrangements within a single species.

Classification and features

Figure 1 and Table 1 summarize the phylogenetic position and characteristics of *Pyrobaculum oguniense* TE7 relative to other members of the *Pyrobaculum* genus, respectively.

Genome sequencing information

Genome project history

Table 2 presents the project information and its association with MGS version 2.0 compliance [23].

Growth conditions and DNA isolation

The initial culture was obtained in 2003 from the Leibniz Institute-German Collection of Microorganisms and Cell Cultures (DSMZ), and grown anaerobically in stoppered, 150ml glass culture bottles at 90°C. This culture was stored at 4°C for an extended period (six years) before being sampled for this study.

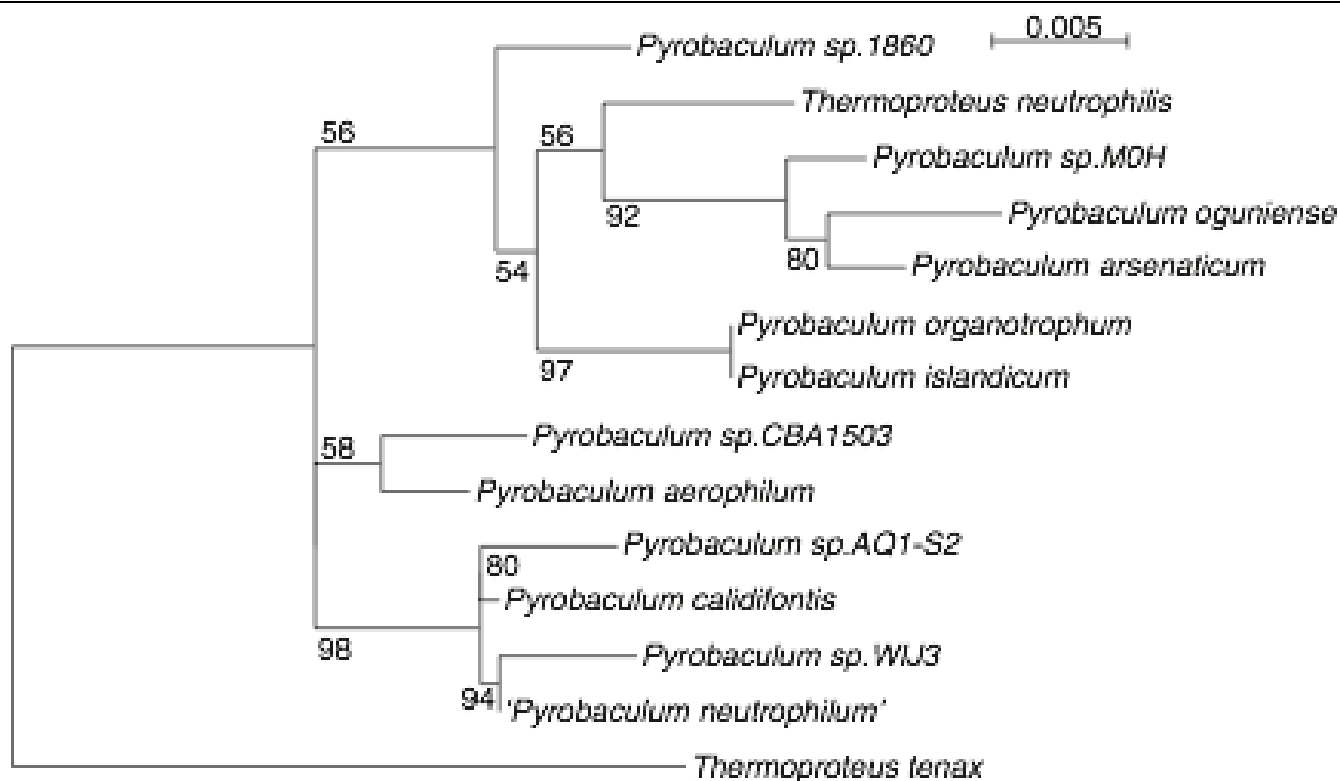


Figure 1. Phylogenetic tree of the known *Pyrobaculum* species based on 16S ribosomal RNA sequence. Accession numbers and associated culture collection identifiers (when available) for 16S ribosomal RNA genes are: *Pyrobaculum aerophilum* (NC_003364.1, DSM 7523); *P. calidifontis* (NC_009073.1, DSM 21063); *P. islandicum* (NC_008701.1, DSM 4184); *P. arsenaticum* (NC_009376.1, DSM 13514); *P. oguniense* (CP003316, DSM 13380); *Pyrobaculum neutrophilum* V24Sta / *Thermoproteus neutrophilus* (NC_010525.1); P.sp. 1860 (CP003098.1); *P. organotrophum* (AB304846.1, DSM 4185); *P. sp.CBA1503* (HM594679.1); *P. sp. MOH* (AB302407.1); *P. sp. AQ1.S2* (DQ778007.1); *P. sp. WIJ3* (AJ277125.1). Sequences were aligned using MAFFT v.6 [8], followed by manual curation [9] to remove 16S ribosomal introns and all terminal gap columns caused by missing sequence. The maximum likelihood tree was constructed using Tree-Puzzle v. 5.2 [10] using exact parameter estimates, 10,000 quartets and 1000 puzzling steps. *Thermoproteus tenax* Kra1 (NC_016070.1, DSM 2078) was included as an outgroup. Numbered branches show bootstrap percentages and branch lengths depict nucleotide mutation rate (see scale bar upper right).

A set of ten-fold dilutions of an actively growing culture ($\sim 10^8$ cells/ml) was carried out and growth was monitored over a five-day period. All cultures were grown at 90°C without shaking in 200ml modified DSM 390 medium, using 1g tryptone, 1g yeast extract, pH 7, supplemented with 10mM $\text{Na}_2\text{S}_2\text{O}_3$ in 1L flasks under a head-space of nitrogen. At day four of growth, a new 400ml aerobic culture was inoculated with 20ml from the penultimate member of the dilution series (10^{-8}) and shaken at 100 rpm, supplemented with 10mM $\text{Na}_2\text{S}_2\text{O}_3$, and subsequently was used for sequencing. We note that at day five, turbid growth was seen in the final member of the dilution series (10^{-9} initial dilution). This implies that

the initial 10^{-8} inoculum used for sequencing likely included more than 10 cells.

Cell pellets were obtained from the 400ml aerobic culture, frozen at -80°C and suspended in 15ml SNET II lysis buffer (20mM Tris-Cl pH 8, 5mM EDTA, 400mM NaCl, 1% SDS) supplemented with 0.5mg/ml Proteinase K and incubated at 55°C for four hours. DNA was extracted from this digest using an equal volume of Tris-buffered (pH 8) PCI (Phenol:Chloroform:Isoamyl-OH (25:24:1)). Following pha se-separation (3220g, 10 min. at 4°C), the resulting aqueous phase was treated with RNase A (25µg/ml) for 30 minutes at 37°C. This reaction was PCI-extracted a second time, followed by CHCl_3 extraction of the resulting aqueous phase and a final phase separation as before.

Table 1. Classification and general features of *Pyrobaculum oguniense* according to the MIGS recommendations [11].

MIGS ID	Property	Term	Evidence code
		Domain <i>Archaea</i>	TAS [12]
		Phylum <i>Crenarchaeota</i>	TAS [13]
		Class <i>Thermoprotei</i>	TAS [14,15]
	Current classification	Order <i>Thermoproteales</i>	TAS [16-19]
		Family <i>Thermoproteaceae</i>	TAS [16-18]
		Genus <i>Pyrobaculum</i>	TAS [20,21]
		Species <i>Pyrobaculum oguniense</i>	TAS [1]
		Type strain <i>TE7</i>	
	Cell shape	rods 0.6-1µm × 2-10µm	TAS [1]
	Sporulation	no	
	Temperature range	70–97°C	
	Optimum temperature	90–94°C	
	Carbon source	heterotroph 1g/L yeast extract or 0.5g/L yeast extract with 0.5g/L tryptone)	TAS [1]
	Energy source	(see carbon source)	TAS [1]
	Terminal electron acceptor	O ₂ , sulfur compounds, no growth on NO ₃ or NO ₂	TAS [1]
MIGS-6	Habitat	hot-spring	TAS [1]
MIGS-6.3	Salinity	0–1.5% (w/v); 0% optimal	TAS [1]
MIGS-22	Oxygen	facultative aerobe	TAS [1]
MIGS-15	Biotic relationship	free-living	NAS
MIGS-14	Pathogenicity	none	NAS
MIGS-4	Geographic location	Tsuetate hot spring, Oguni-cho, Kumamoto prefecture, Japan	TAS [1]
MIGS-5	Sample collection time	June 1997	NAS
MIGS-4.1	Latitude	33.186	NAS
MIGS-4.2	Longitude	131.031	NAS
MIGS-4.3	Depth	hot-spring sediment / fluid	NAS
MIGS-4.4	Altitude	300m	NAS

Evidence codes - TAS: Traceable Author Statement; NAS: Non-traceable Author Statement. These evidence codes are from the Gene Ontology project [22].

Table 2. Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Finished
MIGS-28	Libraries used	Roche 454 Titanium library, SOLiD 2×25 Mate-pair (1k-3.5k insert)
MIGS-29	Sequencing platforms	454 GS FLX Titanium, ABI SOLiD
MIGS-31.2	Fold coverage	59× 454, 500× SOLiD
MIGS-30	Assemblers	Newbler 2.0.01.14, Custom
MIGS-32	Gene calling method	Prodigal, tRNAScan-SE
	Genome Database release	Genbank
	Genbank ID	379005763
		379002962
	Genbank Date of Release	2012-02-12
	GOLD ID	Gc02118
	Project relevance	Biotechnology

DNA was precipitated in an equal volume of isopropyl alcohol at -20°C overnight, followed by centrifugation (3,220 g, 15 min. at 4°C). The resulting pellet was washed in 70% EtOH, pelleted (3220g, 30 min. at 4°C) and aspirated to remove the supernatant. The final DNA pellet was suspended in 1ml TE (50mM Tris-Cl pH 8, 1 mM EDTA) overnight at room temperature, yielding a final DNA concentration of 0.77 µg/µl.

Genome sequencing and assembly

Sequencing was performed by the UCSC genome sequencing center using both Roche/454 GS/FLX Titanium pyrosequencing and the ABI SOLiD system (mate-pair). Pyrosequencing reads were assembled with 59X coverage exceeding Q40 over 99.95% (2,449,310 bases) of the genome, producing 20 contigs at an N50 of 467,815 bp. This assembly included 24 Sanger reads generated by primer-walking across four of the five encoded CRISPR repeat regions. The resulting maximal base-error rate (<Q40) is 25 in 50,000.

Contigs were assembled to a single scaffold using the mate-pair library generated for use on the ABI SOLiD sequencer. The library was produced with an insert size range of 1000–3,500 bp, and final sequencing yielded 30,631,205 read pairs of 25 bp read length. Those read-pairs were mapped to the 20 pyrosequencing-derived contigs to produce a *From::* To table of uniquely mapping read-pairs; read-pair counts were accumulated for each of the 20×20 contig-pair assignments in each of the three possible relative contig orientations (same, converging or diverging). The scaffold closed easily with these data and yielded a single main chromosome with three major inversions and an extra-chromosomal element.

Genome annotation

Gene prediction and annotation was prepared using the IMG/ER service of the Joint Genome Institute [24], where protein coding genes were identified using Prodigal [25]. RNase P RNA [26], SRP RNA and ribosomal RNA(5S, 16S, 23S) were identified by homology to the currently described *Pyrobaculum* members using the UCSC Archaeal Genome Browser (archaea.ucsc.edu) [27]. Annotation of transfer RNA (tRNA) genes was established using tRNAscan-SE [28], supplemented with manual curation of non-canonical introns. C/D box sRNA genes were identified computationally using Snoscan [29]

with extensions supported by transcriptional sequencing [30]. H/ACA-like sRNA genes were identified using transcriptionally-supported homology modeling of experimentally validated sRNA transcripts [31]. CRISPR repeats were identified using CRT [32] or CRISPR-finder [33], with strandedness established by transcriptional sequencing.

Genome properties

The properties and overall statistics of the genome are summarized in Table 3, Table 4, Table 5, Table 6, and Table 7. The single main chromosome (55.08% GC content) has a total size of 2,436,033 bp. Ultra-deep mate-pair sequencing has revealed three regions of the genome that are present in an inverted orientation within a minority of the population (Table 7). The genome also includes an extra-chromosomal element of 16, 887 bp (50.58% GC), that encodes 35 predicted protein-coding genes. Of those genes, seven have an annotated function and the remaining 28 genes are annotated as hypothetical proteins. Of the seven annotated genes, three are coded with viral functions [35].

The majority of the *P. oguniense* genome is structurally syntenic to the genome of *P. arsenaticum*, and genes found in both species show an average of approximately 97% nucleotide identity. The *P. oguniense* genome is approximately 15% larger than *P. arsenaticum*, with the former encoding 536 more (2835 - 2299) open reading frames (ORFs) predicted to be genes. Vast stretches of sequence space are syntenic between the two species (Figure 2, regions in blue), broken by relatively few regions that appear to arise from either gene loss in *P. arsenaticum* or genomic expansion in *P. oguniense*, possibly a result of the numerous paREP elements present in these genomes (Figure 2). These repetitive regions are difficult to assemble, and some are putative transposons (PaREP2b, for example).

We can identify specific genes and gene clusters that are present in *P. oguniense* but are missing in *P. arsenaticum*. Notably, the cobalamin synthetic cluster and two thiamine synthetic genes (ThiW and ThiC) are absent in *P. arsenaticum*. The terminal cytochrome cluster associated with aerobic respiration [36] is also absent in *P. arsenaticum* as expected from an obligate anaerobe. Among the 16 largest deletions in *P. arsenaticum* (relative to *P. oguniense*), four are associated with paREP2 genes, six with paREP1/8, and one with paREP6 (Table 5).

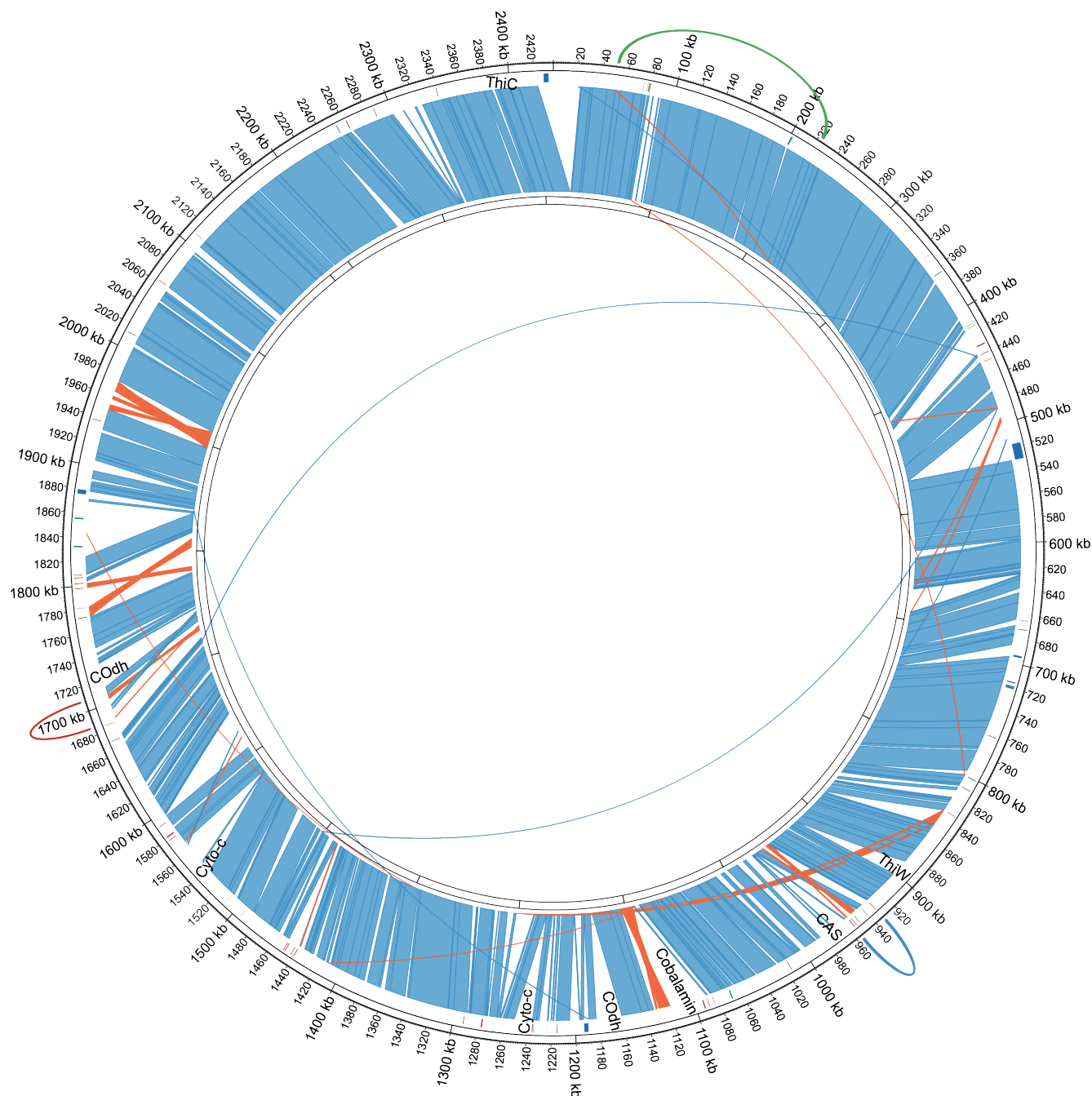


Figure 2. Genomic alignment of *P. oguniense* with *P. arsenaticum*. Outer ring: *P. oguniense* (+ strand); Inner ring: *P. arsenaticum* (- strand). Inter-species alignment blocks shown in light blue and gold (inverted orientation). Intra-species *P. oguniense* genomic inversions shown as arcs of different colors along outer ring: C8 inversion (red); Glutamate Dehydrogenase (GluDH) inversion (green); RAMP/paREP inversion (blue). Positions of paREP elements shown as ticks inside outer ring: paREP1 (red); paREP2b (blue); paREP7 (green). Positions of selected genes which are present in *P. oguniense* and missing in *P. arsenaticum* are shown in text inside outer ring: thiamine biosynthesis genes (ThiW and ThiC); CRISPR Cassette (CAS); cobalamin cluster; CO dehydrogenase (COdh); and the aerobic cytochrome clusters (Cyto-c). Aligned regions smaller than 500 nucleotides have been removed for clarity.

Table 3. Nucleotide content and gene count levels of the main chromosome^a

Attribute	Value	% of total
Genome size (bp)	2,436,033	100
DNA Coding region (bp)	2,164,251	88.84
DNA G+C content (bp)	1,341,816	55.08
Total genes	2,980	100
RNA genes	145	4.74
rRNA operons	1	
Protein-coding genes	2,800	93.96
Genes in paralog clusters	1,214	40.74
Genes assigned to COGs	1,797	60.30
Genes assigned PFAM domains	1,719	57.68
Genes with signal peptides	794	26.64
Genes with transmembrane helices	646	21.68
CRISPR arrays	5	% of total

^aThe ECE (16,887 bp) contains 35 genes, has a 50.58% G+C content, and is excluded from this table. Total gene count includes 35 pseudogenes.

Table 4. Number of genes associated with the 25 general COG functional categories

Code	Value	%age ^a	Description
J	163	8.53	Translation
A	5	0.26	RNA processing and modification
K	112	5.86	Transcription
L	100	5.23	Replication, recombination and repair
B	4	0.21	Chromatin structure and dynamics
D	22	1.15	Cell cycle control, mitosis and meiosis
Y	NA		Nuclear structure
V	15	0.78	Defense mechanisms
T	45	2.35	Signal transduction mechanisms
M	47	2.46	Cell wall/membrane biogenesis
N	4	0.21	Cell motility
Z	1	0.05	Cytoskeleton
W	NA		Extracellular structures
U	22	1.15	Intracellular trafficking and secretion
O	87	4.55	Post-translational modification, protein turnover, chaperones
C	182	9.52	Energy production and conversion
G	82	4.29	Carbohydrate transport and metabolism
E	159	8.32	Amino acid transport and metabolism
F	58	3.04	Nucleotide transport and metabolism
H	115	6.02	Coenzyme transport and metabolism
I	60	3.14	Lipid transport and metabolism
P	83	4.34	Inorganic ion transport and metabolism
Q	26	1.36	Secondary metabolites biosynthesis, transport and catabolism
R	323	16.90	General function prediction only
S	196	10.26	Function unknown
-	1,144		Not in COGs

^aThe total is based on the 1,911 COG assignments made across 1,701 protein-coding genes with at least one COG assignment. The Not in COGs category is made up of 1,099 hypothetical protein coding genes and 145 RNA genes. The 35 genes in the ECE are excluded from this analysis.

Table 5. Sixteen largest regions present in *Pyrobaculum oguniense* and absent in *P. arsenaticum*.

Region coordinates (kb)	PaRep type	Gene cluster
2,420 - 0,020	paREP2	
420 - 440	six with paREP1 or paREP1/8	
485 - 530	paREP2	
682 - 695	paREP2	
887 - 900		ThiW
955 - 985	paREP1/8	CRISPR cassette
1,090 - 1,120	paREP1	Cobalamin biosynthesis cassette
1,160 - 1,180		CO dehydrogenase
1,235 - 1,250	paREP1/8	
1,440 - 1,460	paREP1/8	
1,540 - 1,565		aerobic terminal cytochromes
1,672 - 1,690	paREP6	
1,715 - 1,735		CO dehydrogenase
1,780 - 1,795	paREP1	
1,825 - 1,870	paREP2	
2,300 - 2,385		ThiC

Table 6. Summary of genome: one chromosome and one extra-chromosomal element

Label	Size (bp)	Topology	INSDC identifier
Chromosome (Chr)	2,436,033	circular	NC_016885.1
Extra-chromosomal Element (ECE)	16,887	circular	NC_016886.1

Table 7. Genomic inversions present within the sampled population

Inversion name	Coordinates		Length	Frequency
	Start	End		
GluDH	50,930	223,540	172,611	0.17
RAMP/paREP	932,090	955,719	23,630	0.18
C8	1,686,376	1,708,299	21,924	0.35

^aMinority inversion frequency established as described previously [34].

Conclusion

Genomic sequencing and assembly of *Pyrobaculum oguniense* has yielded a complete genome and an extra-chromosomal element. The main chromosome is largely syntenic to *Pyrobaculum arsenaticum* and contains a number of gene clusters that are absent in that species. This is of particular interest considering that these species were isolated on opposite sides of the Eurasian continent; *P. oguniense* was isolated in Japan, while *P. arsenaticum* was isolated in an arsenic-rich anaerobic pool in Italy.

The synteny that has been retained between the genomes of *P. oguniense* and *P. arsenaticum* allows a close examination of gene gain or loss events in the genetic history of these two species. *P. arsenaticum* is missing the gene clusters that support cobalamin and thiamine synthesis, and it is missing the aerobic cytochrome cluster. Given that *P. oguniense* and the next closest member in the clade, *P. aerophilum*, have both retained these capabilities; the most parsimonious explanation is gene loss in *P. arsenaticum*. Because these genes are located at disparate positions in the *P. oguniense* genome, it would further appear that these losses are the result of multiple events in the evolutionary history of *P. arsenaticum*.

Within this genome, 145 non-coding RNA genes are described. These include a single operon encoding 16S and 23S ribosomal RNA, the associated 5S rRNA, the 7S signal recognition particle (SRP), and the RNase P RNA. There are 47 annotated

tRNA genes, plus a single tRNA pseudogene. Also included are 83 predicted C/D box sRNA genes and nine additional H/ACA-like sRNA, each of which has been transcriptionally validated [31]. The non-coding RNA content of the *P. oguniense* genome has become the most extensively annotated among crenarchaeal genomes to date.

The use of a not-quite-clonal cell population for DNA isolation, coupled with ultra-deep sequencing has provided a view of three major inversions that are each present in over 17% of the sample population. The boundaries of one of these inversions are defined by an inverted repeat encoding a duplication of glutamate dehydrogenase (GluDH). Notably, this duplication appears to be present in each of the currently sequenced *Pyrobaculum* members, suggesting that those genomes may also host similar inversions. A second inversion has at its termini another inverted duplication, encoding a gene associated with one of the paREP members and a CRISPR-associated gene. It remains unclear if these common structural variants impart a physiological advantage, and if so, how the variation provides utility to its host. Based on our expanded genome diversity observations, we suggest that avoiding the use of a strictly clonal population for sequencing purposes can provide a significant benefit to understanding both the biology of the host and a clearer understanding of the genome dynamics of the species.

Acknowledgements

Sequencing was provided by the UCSC Genome Sequencing Center. We would like to thank Nathan Boyd, Eveline Hesson and Nader Pourmand for their expertise and advice in this work. This work was supported by National Science Foundation Grant DBI-0641061

(T.L. and D.B.) and the Graduate Research and Education in Adaptive Bio-Technology (GREAT) Training Program sponsored by the University of California Biotechnology Research and Education Program (D.B.).

References

1. Sako Y, Nunoura T, Uchida A. *Pyrobaculum oguniense* sp. nov., a novel facultatively aerobic and hyperthermophilic archaeon growing at up to 97 degrees C. *Int J Syst Evol Microbiol* 2001; **51**:303-309. [PubMed](#)
2. Chan PP, Cozen AE, Lowe TM. Reclassification of *Thermoproteus neutrophilus* Stetter and Zillig 1989 as *Pyrobaculum neutrophilum* comb. nov. based on phylogenetic analysis. *Int J Syst Evol Microbiol* 2012. [PubMed](#) <http://dx.doi.org/10.1099/ij.s.0.043091-0>
3. Huber R, Sacher M, Vollmann A, Huber H, Rose D. Respiration of arsenate and selenate by hyperthermophilic archaea. *Syst Appl Microbiol* 2000; **23**:305-314. [PubMed](#) [http://dx.doi.org/10.1016/S0723-2020\(00\)80058-2](http://dx.doi.org/10.1016/S0723-2020(00)80058-2)
4. Mardanov AV, Gumerov VM, Slobodkina GB, Beletsky AV, Bonch-Osmolovskaya EA, Ravin NV, Skryabin KG. Complete genome sequence of strain 1860, a crenarchaeon of the genus pyrobaculum able to grow with various electron

- acceptors. *J Bacteriol* 2012; **194**:727-728. [PubMed](http://dx.doi.org/10.1128/JB.06465-11) <http://dx.doi.org/10.1128/JB.06465-11>
5. Amo T, Paje ML, Inagaki A, Ezaki S, Atomi H, Imanaka T. *Pyrobaculum calidifontis* sp. nov., a novel hyperthermophilic archaeon that grows in atmospheric air. *Archaea* 2002; **1**:113-121. [PubMed](http://dx.doi.org/10.1155/2002/616075) <http://dx.doi.org/10.1155/2002/616075>
 6. Cozen AE, Weirauch MT, Pollard KS, Bernick DL, Stuart JM, Lowe TM. Transcriptional map of respiratory versatility in the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *J Bacteriol* 2009; **191**:782-794. [PubMed](http://dx.doi.org/10.1128/JB.00965-08) <http://dx.doi.org/10.1128/JB.00965-08>
 7. Fitz-Gibbon ST, Ladner H, Kim UJ, Stetter KO, Simon MI, Miller JH. Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *Proc Natl Acad Sci USA* 2002; **99**:984-989. [PubMed](http://dx.doi.org/10.1073/pnas.241636498) <http://dx.doi.org/10.1073/pnas.241636498>
 8. Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 2008; **9**:286-298. [PubMed](http://dx.doi.org/10.1093/bib/bbn013) <http://dx.doi.org/10.1093/bib/bbn013>
 9. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009; **25**:1189-1191. [PubMed](http://dx.doi.org/10.1093/bioinformatics/btp033) <http://dx.doi.org/10.1093/bioinformatics/btp033>
 10. Strimmer K, von Haeseler A. Quartet Puzzling: A Quartet Maximum-Likelihood Method for Reconstructing Tree Topologies. *Mol Biol Evol* 1996; **13**:964-969. [PubMed](http://dx.doi.org/10.1093/oxfordjournals.molbev.a025664) <http://dx.doi.org/10.1093/oxfordjournals.molbev.a025664>
 11. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, et al. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. [PubMed](http://dx.doi.org/10.1038/nbt1360) <http://dx.doi.org/10.1038/nbt1360>
 12. 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](http://dx.doi.org/10.1073/pnas.87.12.4576) <http://dx.doi.org/10.1073/pnas.87.12.4576>
 13. Garrity GM, Holt JG. Phylum AI. *Crenarchaeota* phy. nov. In: Garrity GM, Boone DR, Castenholz RW (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 1, Springer, New York, 2001, p. 169-210.
 14. List Editor. Validation List no. 85. Validation of publication of new names and new combinations previously effectively published outside the IJSEM. *Int J Syst Evol Microbiol* 2002; **52**:685-690. [PubMed](http://dx.doi.org/10.1099/ijs.0.02358-0) <http://dx.doi.org/10.1099/ijs.0.02358-0>
 15. Reysenbach AL. Class I. *Thermoprotei* class. nov. In: Garrity GM, Boone DR, Castenholz RW (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 1, Springer, New York, 2001, p. 169.
 16. Validation of the publication of new names and new combinations previously effectively published outside the IJSB. List No. 8. *Int J Syst Bacteriol* 1982; **32**:266-268. <http://dx.doi.org/10.1099/00207713-32-2-266>
 17. Zillig W, Stetter KO, Schäfer W, Janekovic D, Wunderl S, Holz J, Palm P. *Thermoproteales*: a novel type of extremely thermoacidophilic anaerobic archaeobacteria isolated from Icelandic solfataras. [Orig A]. *Zentralbl Bakteriol* 1981; **C2**:205-227.
 18. Burggraf S, Huber H, Stetter KO. Reclassification of the crenarchael orders and families in accordance with 16S rRNA sequence data. *Int J Syst Bacteriol* 1997; **47**:657-660. [PubMed](http://dx.doi.org/10.1099/00207713-47-3-657) <http://dx.doi.org/10.1099/00207713-47-3-657>
 19. Judicial Commission of the International Committee on Systematics of Prokaryotes. The nomenclatural types of the orders *Acholeplasmatales*, *Halanaerobiales*, *Halobacteriales*, *Methanobacteriales*, *Methanococcales*, *Methanomicrobiales*, *Planctomycetales*, *Prochlorales*, *Sulfobolales*, *Thermococcales*, *Thermoproteales* and *Verrucomicrobiales* are the genera *Acholeplasma*, *Halanaerobium*, *Halobacterium*, *Methanobacterium*, *Methanococcus*, *Methanomicrobium*, *Planctomyces*, *Prochloron*, *Sulfobolus*, *Thermococcus*, *Thermoproteus* and *Verrucomicrobium*, respectively. Opinion 79. *Int J Syst Evol Microbiol* 2005; **55**:517-518. [PubMed](http://dx.doi.org/10.1099/ijs.0.63548-0) <http://dx.doi.org/10.1099/ijs.0.63548-0>
 20. List Editor. Validation of the publication of new names and new combinations previously effectively published outside the IJSB. List No. 25. *Int J Syst Bacteriol* 1988; **38**:220-222. <http://dx.doi.org/10.1099/00207713-38-2-220>
 21. Huber R, Kristjansson JK, Stetter KO. *Pyrobaculum* gen. nov., a new genus of neutrophilic, rod-shaped archaeobacteria from continental solfataras growing optimally at 100 C. *Arch Microbiol* 1987; **149**:95-101. <http://dx.doi.org/10.1007/BF00425072>

22. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29. [PubMed](#) <http://dx.doi.org/10.1038/75556>
23. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, Ashburner M. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. [PubMed](#) <http://dx.doi.org/10.1038/nbt1360>
24. DOE. Joint Genome Institute. <http://img.jgi.doe.gov>
25. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**:119. [PubMed](#) <http://dx.doi.org/10.1186/1471-2105-11-119>
26. Lai LB, Chan PP, Cozen AE, Bernick DL, Brown JW, Gopalan V, Lowe TM. Discovery of a minimal form of RNase P in *Pyrobaculum*. *Proc Natl Acad Sci USA* 2010; **107**:22493-22498. [PubMed](#) <http://dx.doi.org/10.1073/pnas.1013969107>
27. Chan PP, Holmes AD, Smith AM, Tran D, Lowe TM. The UCSC Archaeal Genome Browser: 2012 update. *Nucleic Acids Res* 2012; **40**(Database issue):D646-D652. [PubMed](#) <http://dx.doi.org/10.1093/nar/gkr990>
28. 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](#)
29. Lowe TM, Eddy SR. A computational screen for methylation guide snoRNAs in yeast. *Science* 1999; **283**:1168-1171. [PubMed](#) <http://dx.doi.org/10.1126/science.283.5405.1168>
30. Bernick DL, Dennis PP, Lui LM, Lowe TM. Diversity of antisense and other non-coding RNAs in Archaea revealed by comparative small RNA sequencing in four *Pyrobaculum* species. *Frontiers in Microbiology* 2012;3. <http://dx.doi.org/10.3389/fmicb.2012.00231>
31. Bernick DL, Dennis PP, Hochsmann M, Lowe TM. Discovery of *Pyrobaculum* small RNA families with atypical pseudouridine guide RNA features. *RNA* 2012; **18**:402-411. [PubMed](#) <http://dx.doi.org/10.1261/rna.031385.111>
32. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 2007; **8**:209. [PubMed](#) <http://dx.doi.org/10.1186/1471-2105-8-209>
33. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 2007;35(Web Server issue):W52-57
34. Bernick DL. Sequential discovery - from small RNA to genomes, an investigation of the hyperthermophilic genus *Pyrobaculum*. Santa Cruz, California USA: University of California, Santa Cruz; 2010. 120 p.
35. Krupovic M, Bamford DH. Archaeal proviruses TKV4 and MVV extend the PRD1-adenovirus lineage to the phylum *Euryarchaeota*. *Virology* 2008; **375**:292-300. [PubMed](#) <http://dx.doi.org/10.1016/j.virol.2008.01.043>
36. Nunoura T, Sako Y, Wakagi T, Uchida A. Regulation of the aerobic respiratory chain in the facultatively aerobic and hyperthermophilic archaeon *Pyrobaculum oguniense*. *Microbiology* 2003; **149**:673-688. [PubMed](#) <http://dx.doi.org/10.1099/mic.0.26000-0>
37. Sako Y, Nunoura T, Uchida A. *Pyrobaculum oguniense* sp. nov., a novel facultatively aerobic and hyperthermophilic archaeon growing at up to 97 degrees C. *Int J Syst Evol Microbiol* 2001; **51**:303-309. [PubMed](#)