JOURNAL OF BACTERIOLOGY, Apr. 2008, p. 2957–2965 Vol. 190, No. 8
0021-9193/08/$08.00+0 doi:10.1128/JB.01949-07

Genome Sequence of *Thermofilum pendens* Reveals an Exceptional Loss of Biosynthetic Pathways without Genome Reduction▽†

Iain Anderson,[1]* Jason Rodriguez,[2,3] Dwi Susanti,[2,5] Iris Porat,[6]‡ Claudia Reich,[7] Luke E. Ulrich,[8] James G. Elkins,[9] Kostas Mavromatis,[1] Athanasios Lykidis,[1] Edwin Kim,[1] Linda S. Thompson,[1,10] Matt Nolan,[1] Miriam Land,[9] Alex Copeland,[1] Alla Lapidus,[1] Susan Lucas,[1] Chris Detter,[1,10] Igor B. Zhulin,[8] Gary J. Olsen,[7] William Whitman,[6] Biswarup Mukhopadhyay,[2,3,4,5] James Bristow,[1] and Nikos Kyrpides[1]

Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598[1]; Virginia Bioinformatics Institute[2] and Departments of Biochemistry[3] and Biological Sciences[4] and Genetics, Bioinformatics and Computational Biology Graduate Program,[5] Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061; Department of Microbiology, University of Georgia, Athens, Georgia 30602[6]; Department of Microbiology, University of Illinois, Urbana, Illinois 61801[7]; Joint Institute for Computational Sciences, University of Tennessee-Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831[8]; Bioscience Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831[9]; and Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545[10]

Received 14 December 2007/Accepted 1 February 2008

We report the complete genome of *Thermofilum pendens*, a deeply branching, hyperthermophilic member of the order *Thermoproteales* in the archaeal kingdom *Crenarchaeota*. *T. pendens* is a sulfur-dependent, anaerobic heterotroph isolated from a solfatara in Iceland. It is an extracellular commensal, requiring an extract of *Thermoproteus tenax* for growth, and the genome sequence reveals that biosynthetic pathways for purines, most amino acids, and most cofactors are absent. In fact, *T. pendens* has fewer biosynthetic enzymes than obligate intracellular parasites, although it does not display other features that are common among obligate parasites and thus does not appear to be in the process of becoming a parasite. It appears that *T. pendens* has adapted to life in an environment rich in nutrients. *T. pendens* was known previously to utilize peptides as an energy source, but the genome revealed a substantial ability to grow on carbohydrates. *T. pendens* is the first crenarchaeote and only the second archaeon found to have a transporter of the phosphotransferase system. In addition to fermentation, *T. pendens* may obtain energy from sulfur reduction with hydrogen and formate as electron donors. It may also be capable of sulfur-independent growth on formate with formate hydrogen lyase. Additional novel features are the presence of a monomethylamine:corrinoid methyltransferase, the first time that this enzyme has been found outside the *Methanosarcinales*, and the presence of a presenilin-related protein. The predicted highly expressed proteins do not include proteins encoded by housekeeping genes and instead include ABC transporters for carbohydrates and peptides and clustered regularly interspaced short palindromic repeat-associated proteins.

Crenarchaeota is one of the two major divisions of the *Archaea*, and it is the least well represented taxon in terms of genome sequences. Only seven crenarchaeal genomes have been sequenced and published so far, and three of these are genomes of members of the genus *Sulfolobus*. For the order *Thermoproteales*, the complete sequence of only one organism, *Pyrobaculum aerophilum*, has been determined and published so far, although several more species of *Pyrobaculum*, *Caldivirga maquilingensis*, and *Thermoproteus tenax* have been or are currently being sequenced (29). *Thermofilum pendens* represents a deep branch in the order *Thermoproteales*, and this organism grows only in rich medium with a fraction of the polar lipids of *T. tenax* (64), a property that has not been seen before in archaea. Therefore, it was an attractive sequencing target. We report here the genome sequence of *T. pendens* and analysis of the type strain, *T. pendens* Hrk5.

T. pendens is an anaerobic, sulfur-dependent hyperthermophile isolated from a solfatara in Iceland. It forms long thin filaments and may have an unusual mode of reproduction in which spherical bulges form at one end of the cell. It requires complex media and a lipid extract from the related organism *T. tenax* for growth (64). The unknown lipid may be a cellular component or may make sulfur more available to the cells. Complex media, such as tryptone or yeast extract, are required for growth, and CO_2 and H_2S are produced, similar to characteristics of other anaerobic members of the *Crenarchaeota* and the euryarchaeal family *Thermococcaceae*. The genome shows that this organism appears to have lost most biosynthetic pathways, yet does not have a reduced genome size compared to other *Crenarchaeota*.

MATERIALS AND METHODS

Frozen *T. pendens* Hrk5 cells were obtained from Karl Stetter. Cells were resuspended in 0.25 M sucrose in Tris-EDTA buffer. Sodium dodecyl sulfate was

* Corresponding author. Mailing address: Joint Genome Institute, 2800 Mitchell Dr., Walnut Creek, CA 94598. Phone: (925) 296-5784. Fax: (925) 296-5850. E-mail: IJAnderson@lbl.gov.

† Supplemental material for this article may be found at http://jb.asm.org/.

‡ Present address: Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA 16802.

▽ Published ahead of print on 8 February 2008.

TABLE 1. General statistics for *T. pendens*

Parameter	Value
Chromosome size (bp)	1,781,889
Chromosome G+C content (bp) (%)	1,027,538 (57.6)
Plasmid size (bp)	31,504
Plasmid G+C content (bp) (%)	17,813 (56.5)
Total genome size (bp)	1,813,393
Total genome G+C content (bp) (%)	1,045,351 (57.6)
Total no. of genes	1,923
No. (%) of RNA genes	40 (2.1)
No. (%) of protein-encoding genes	1,883 (97.9)
No. (%) of genes with function prediction	1,170 (60.8)
No. (%) of genes in ortholog clusters	1,541 (80.1)
No. (%) of genes in paralog clusters	805 (41.9)
No. (%) of genes assigned to COGs	1,264 (65.7)
No. (%) of genes assigned to Pfam domains	1,209 (62.9)
No. (%) of genes with signal peptides	134 (7.0)
No. (%) of genes with transmembrane helices	437 (22.7)
No. (%) of fusion genes	79 (4.11)

added to a concentration of 1%, and cells were lysed by three cycles of freezing and thawing. Proteinase K was added to a concentration of 50 μg/ml, and the lysate was incubated at 60°C for 30 min. Undigested proteins were precipitated by addition of 0.5 M NaCl and were removed by centrifugation. Nucleic acids in the supernatant were precipitated by addition of an equal volume of cold isopropanol and collected by centrifugation. After digestion with RNase A, the DNA was purified by successive extractions with phenol and phenol-chloroform and recovered by ethanol precipitation. DNA was resuspended in Tris-EDTA buffer and sent to the Joint Genome Institute.

The genome of *T. pendens* was sequenced at the Joint Genome Institute using a combination of 3-, 6-, and 40-kb (fosmid) DNA libraries. All general aspects of library construction and sequencing performed at the Joint Genome Institute are described at http://www.jgi.doe.gov/. Draft assemblies were based on 21,478 total reads. All three libraries provided 11× coverage of the genome. The Phred/Phrap/Consed software package (www.phrap.com) was used for sequence assembly and quality assessment (7, 8, 10). After the shotgun stage, reads were assembled with parallel phrap (High Performance Software, LLC). Possible misassemblies were corrected with Dupfinisher (16) or transposon bombing of bridging clones (Epicenter Biotechnologies, Madison, WI). Gaps between contigs were closed by editing in Consed, custom primer walking, or PCR amplification (Roche Applied Science, Indianapolis, IN). A total of 465 additional reactions were necessary to close gaps and to increase the quality of the finished sequence. Genes were identified using a combination of Critica (2) and Glimmer (6), followed by a round of manual curation.

Analysis of the *T. pendens* genome was carried out with the Integrated Microbial Genomes (IMG) system (33). Protein families unique to *T. pendens* or missing from *T. pendens* but present in other *Crenarchaeota* were identified with the phylogenetic profiler in the IMG system. Analysis of signal transduction was carried out using the MiST database (60), which uses SMART (28) and Pfam (9) domain assignments. A cumulative GC skew plot was generated with a 35-kb sliding window using GraphDNA (56).

Predicted highly expressed (PHX) genes were determined with the EMBOSS (45) programs cusp and cai. The training set of PHX genes was compiled from the data of Karlin et al. (21). The genes with a codon adaptation index in the top 5% were considered to be PHX genes.

Nucleotide sequence accession numbers. The sequences of *T. pendens*, consisting of the sequences of one chromosome and one plasmid, can be accessed using GenBank accession numbers CP000505 and CP000506. The Genomes On Line Database accession number for *T. pendens* is Gc00473.

RESULTS

General features. The genome of *T. pendens* Hrk5 consists of a circular 1.78-Mbp chromosome and a 31,504-bp plasmid (Table 1). The G+C content is 58%, which is higher than the G+C contents of other *Crenarchaeota*. A total of 1,923 genes were identified, and 1,883 of these genes encode proteins. The percentage of the genome devoted to encoding genes is 91%, which is slightly higher than the values for other sequenced *Crenarchaeota*. About 59% of the protein-encoding genes begin with an AUG codon, 32% begin with a GUG codon, and 10% begin with UUG. About 66% of the protein-encoding genes have been assigned to clusters of orthologous groups (COGs) (55), and about 63% have Pfam (9) domains, similar to the values for other archaeal genomes. There is one copy of each rRNA. *T. pendens* has the highest percentage of fusion genes among members of the *Crenarchaeota*. Several proteins that are present in *T. pendens* have not been found in crenarchaeotes or in archaea previously (Table 2), and genes encoding several proteins found in all other crenarchaeotes are missing from the *T. pendens* genome (Table 3).

The plasmid is predicted to encode 52 proteins, only 2 of which have similarity to proteins in the GenBank nonredundant protein database. The Tpen_1849 protein is similar to a *T. pendens* chromosomal protein with an unknown function (Tpen_0735), and the Tpen_1875 protein is a predicted helicase. In addition, the Tpen_1891 protein is predicted to be a site-specific recombinase (COG4974). The function of the plasmid and whether it is beneficial to the host are currently unknown.

Cumulative GC skew analysis of the *T. pendens* genome was used to identify the potential origin(s) of replication (12). A global minimum was located at position 488884, which is near a 478-bp intergenic region between positions 487890 and 488368. The intergenic spacer contains several repetitive sequences similar to conserved crenarchaeal origin recognition boxes (46).

No repetitive elements were found when the ISfinder database (53) was searched with the *T. pendens* coding sequences. However, in a 100,000-bp section of the genome there are 12 stretches of clustered regularly interspaced short palindromic repeat (CRISPR) elements that are interspersed with protein-encoding genes. CRISPR elements and their associated genes constitute a defense against viruses (3). Among the 97 PHX genes (see Table S1 in the supplemental material) are 6 genes encoding CRISPR-associated proteins (Tpen_1263, Tpen_1287, Tpen_1288, Tpen_1316, Tpen_1342, and Tpen_1356). Interestingly, a group of seven consecutive genes (Tpen_1287 to Tpen_1293), including two genes encoding CRISPR-associated proteins, is PHX. While other thermophilic archaea have similar numbers of genes encoding CRISPR-associated proteins in their genomes, most do not have genes encoding CRISPR-associated proteins that are PHX genes, although *Staphylothermus marinus* and *P. aerophilum* do (four and two genes, respectively). However, *T. pendens* has the highest number of CRISPR-associated genes that are PHX genes and the highest percentage. Thus, protection against viral infection appears to be a major priority for *T. pendens*. *Crenarchaeota* from hot spring environments are known to host a wide variety of viruses with distinctive morphologies (for a review, see reference 42).

Central metabolism. *T. pendens* contains complete glycolysis and gluconeogenesis pathways. Glyceraldehyde-3-phosphate: ferredoxin oxidoreductase, found in some archaeal hyperthermophiles as part of an alternative step in glycolysis (37), is not present in *T. pendens*. Phosphoenolpyruvate synthase, used in the last step in glycolysis in *Thermococcus kodakaraensis* (19), is present in *T. pendens* (Tpen_0588). Phosphoenolpyruvate

TABLE 2. Unique genes in *T. pendens* with COG hits

CDS	COG no.	Function
COGs not found in any other sequenced *Archaea*		
Tpen_1241	0698	Ribose-5-phosphate isomerase rpiB
Tpen_1297	3525	Glycosyl hydrolase family 20
Tpen_1097	3444	PTS IIB subunit
Tpen_1100	3715	PTS IIC subunit
Tpen_1100	3716	PTS IID subunit
Tpen_1090	4821	Phosphosugar binding protein, SIS domain
COGs not found in any other sequenced *Crenarchaeota*		
Tpen_1155	1554	Glycoside hydrolase family 65
Tpen_1624	3836	2-Dehydro-3-deoxyglucarate aldolase
Tpen_0948	0207	Thymidylate synthase
Tpen_0017	3613	Nucleoside 2-deoxyribosyltransferase
Tpen_1467	5598	Trimethylamine:corrinoid methyltransferase
Tpen_1211		Monomethylamine:corrinoid methyltransferase
Tpen_1092	1080	Phosphoenolpyruvate-protein kinase (enzyme I of PTS)
Tpen_1091	1925	PTS HPr protein
Tpen_1098	2893	PTS IIA component
Tpen_1491	1268	Biotin transporter BioY
Tpen_0929	3601	Riboflavin transporter
Tpen_0191	2116	Formate transporter
Tpen_1479	2060	Potassium-transporting ATPase, A chain
Tpen_1480	2216	Potassium-transporting ATPase, B chain
Tpen_1481	2156	Potassium-transporting ATPase, c chain
Tpen_0197	0474	Cation transport ATPase (P-type
Tpen_1427		ATPase)
Tpen_1010	1327	Predicted transcriptional regulator, Zn ribbon and ATP cone
Tpen_1048	1510	Predicted transcriptional regulator
Tpen_0270	4190	Predicted transcriptional regulator
Tpen_0889	2150	Predicted regulator of amino acid metabolism, ACT domain
Tpen_0253	2229	Predicted GTPase
Tpen_1457	1773	Rubredoxin
Tpen_1536	1811	Uncharacterized membrane protein, DUF554
Tpen_0198	2047	Uncharacterized protein, ATP-grasp superfamily
Tpen_0838	2164	Uncharacterized conserved protein,
Tpen_1835		DUF369
Tpen_1118	2908	Uncharacterized conserved protein
Tpen_0381	3863	Uncharacterized relative of cell wall-associated hydrolases
Tpen_1090	4821	Uncharacterized protein with phosphosugar binding domain

TABLE 3. Genes present in all *Crenarchaeota* except *T. pendens* with COG hits

COG	Function
0214	Pyridoxine biosynthesis enzyme (YaaD)
0311	Glutamine amidotransferase involved in pyridoxine synthesis (YaaE)
0413	Ketopantoate hydroxymethyltransferase
0452	Phosphopantothenoylcysteine synthetase/decarboxylase
0108	3,4-Dihydroxy-2-butanone-4-phosphate synthase
1985	Pyrimidine reductase, riboflavin biosynthesis
0054	Riboflavin synthase beta chain
0163	3-Polyprenyl-4-hydroxybenzoate decarboxylase
1635	Flavoprotein involved in thiazole biosynthesis
0112	Glycine/serine hydroxymethyltransferase
0189	Glutathione synthase/ribosomal protein S6P modification enzyme/L-2-aminoadipate *N*-acetyltransferase
0105	Nucleoside diphosphate kinase
2046	Sulfate adenylyltransferase
1650	Uncharacterized protein conserved in archaea
1701	Uncharacterized protein conserved in archaea

synthase could be involved in glycolysis and/or in gluconeogenesis. Starch synthesis and utilization pathways are also present.

Pentoses are synthesized through the ribulose monophosphate pathway that is common in archaea (for a review, see reference 22). *T. pendens* genes encode two ribose-5-phosphate isomerases, one RpiA-type enzyme (Tpen_0327) and one RpiB-type enzyme (Tpen_1241). This is the first time that an RpiB enzyme has been found in archaea. The *rpiB* gene is adjacent to a uridine phosphorylase gene (Tpen_1240), suggesting that this protein has a function in nucleoside utilization. Under conditions in which excess ribonucleosides are present, RpiB may be involved in conversion of ribose phosphate to hexoses through the ribulose monophosphate pathway, a reversal of the pathway from its predicted function in archaea.

ATP can be generated from pyruvate through the consecutive action of pyruvate:ferredoxin oxidoreductase and ADP-forming acetyl-coenzyme A (acetyl-CoA) synthase, similar to the activity in *Thermococcales* (32). The *T. pendens* pyruvate:ferredoxin oxidoreductase (Tpen_0571 to Tpen_0574) is similar to the *Thermotoga maritima* enzyme that has been characterized (23). In *Thermococcales* and *Crenarchaeota*, ADP-forming acetyl-CoA synthase has two subunits, the alpha and beta subunits. *T. pendens* contains one alpha subunit (Tpen_0336), one beta subunit (Tpen_0109), and one protein with alpha and beta subunits fused together (Tpen_0602). Two AMP-forming acyl-CoA synthases are also present (Tpen_0893 and Tpen_1611). *T. pendens* has four other enzymes similar to pyruvate:ferredoxin oxidoreductase (Tpen_0540 to Tpen_0543, Tpen_0781 and Tpen_0782, Tpen_0856 and Tpen_0857, and Tpen_1455 and Tpen_1456), and these enzymes are likely to be involved in amino acid degradation pathways in which the amino acid is first converted to the 2-ketoacid, then to the acyl-CoA, and finally to an acid, with ATP generated by acyl-CoA synthases (32). Four aldehyde:ferredoxin oxidoreductases are also present (Tpen_0094, Tpen_0176, Tpen_1413, and Tpen_1817), and these enzymes could be involved in peptide fermentation (1). The 2-oxoacid oxidoreductases produce aldehydes, which are converted to acids. Reduced ferredoxin is produced, but there is no ATP production in this pathway.

T. pendens appears to assimilate glycerol. There is a glycerol kinase gene (Tpen_1128) adjacent to the gene encoding subunit A of glycerol-3-phosphate dehydrogenase (Tpen_1127).

Next to these genes are three genes with similarity to genes encoding subunits B, C, and D of succinate dehydrogenases (Tpen_1124 to Tpen_1126). A gene encoding subunit A of succinate dehydrogenase has not been found in the genome. It appears that the three succinate dehydrogenase-related subunits along with the glycerol-3-phosphate dehydrogenase subunit A may form a novel glycerol-3-phosphate dehydrogenase that may transfer electrons to a quinone or another acceptor.

Unlike most of the sequenced *Crenarchaeota*, *T. pendens* has a ribulose-1,5-bisphosphate carboxylase (Tpen_1227). It also has the recently discovered enzymes involved in conversion of the ribose phosphate group of AMP to ribulose 1,5-bisphosphate (50): AMP phosphorylase (Tpen_0093) and ribose-1,5-bisphosphate isomerase (Tpen_0384). Under conditions in which acetate is incorporated into the gluconeogenesis pathway, the AMP-forming acetyl-CoA synthetase and phosphoenolpyruvate synthase could produce substantial amounts of AMP. The *T. pendens* AMP-forming acetyl-CoA synthetase (Tpen_0893) has a very high level of similarity to the *P. aerophilum* enzyme that has been characterized (4). A large amount of AMP may also be generated by ribose-phosphate pyrophosphokinase, which is required for pyrimidine synthesis, and phosphoribosyltransferases.

Biosynthesis. *T. pendens* is known to require an extract of *T. tenax* for growth. While the specific compound from *T. tenax* required could not be identified, the genome analysis revealed a vast reduction in the ability of *T. pendens* to synthesize basic metabolites. *T. pendens* appears to be dependent on its environment for purines, most cofactors, and most amino acids. A list of 125 COGs involved in synthesis of nucleobases, amino acids, and cofactors was compiled (see Table S2 in the supplemental material). COGs encoding archaeal biosynthetic enzymes were included where they are known. The presence of these COGs in all complete bacterial and archaeal genomes was determined using the function profile feature in the IMG system. *T. pendens* possesses only 11 of these COGs. The COGs found in *T. pendens* include genes encoding five enzymes involved in pyrimidine synthesis, aspartate carbamoyltransferase (COG0540), dihydroorotase (COG0044), dihydroorotate dehydrogenase (COG0167), orotate phosphoribosyltransferase (COG0461), and orotidine-5′-phosphate decarboxylase (COG0284). Also present are genes for threonine synthase (COG0498), 2-polyprenylphenol hydroxylase (COG0543), protoporphyrinogen oxidase (COG1232), 4-hydroxybenzoate polyprenyltransferase (COG0382; two copies), 1,4-dihydroxy-2-naphthoate octaprenyltransferase (COG1575; two copies), and a methylase involved in ubiquinone/menaquinone biosynthesis (COG2226; six copies). The only organisms with fewer members of this COG set were obligate parasites or commensals. In fact, some obligate parasites, such as *Rickettsia* species, have greater biosynthetic capabilities than *T. pendens*.

While it is possible that *T. pendens* has different pathways for metabolite synthesis or has many enzymes that were replaced through nonorthologous gene displacement, this is unlikely to account for the lack of biosynthetic enzymes because other *Crenarchaeota* have recognized pathways for basic metabolites. For example, all *Crenarchaeota* except *T. pendens* have homologs of the pyridoxine biosynthesis genes *pdx1* and *pdx2* (*yaaD* and *yaaE* in *Bacillus subtilis*) and the bifunctional CoA biosynthetic enzyme phosphopantothenoylcysteine synthetase/decarboxylase. Table 3 shows the COGs missing from *T. pendens* that are found in all other sequenced *Crenarchaeota*. Nine of these COGs are involved in pyridoxine, CoA, riboflavin, ubiquinone, and thiamine biosynthesis. In addition, most *Crenarchaeota* have homologs of several heme biosynthetic enzymes, but *T. pendens* lacks these enzymes. They are not shown in Table 3 because they are also not present in the *Staphylothermus marinus* genome. Also, COG1731, encoding archaeal riboflavin synthase, is not shown in Table 3 because it is not present in either *T. pendens* or *Cenarchaeum symbiosum*; however, *C. symbiosum* has the bacterium-type riboflavin synthase (COG0307), but *T. pendens* lacks both the bacterial and archaeal enzymes.

In accordance with the predicted lack of biosynthetic capacity, *T. pendens* is the only crenarchaeote that has a BioY family biotin transporter and a riboflavin transporter (Table 2). In addition, *T. pendens* has an expanded number of ABC transporters related to the ABC transporters involved in cobalt uptake. While most *Crenarchaeota* have zero to two representatives of this family, *T. pendens* has seven. One of these transporters has an additional membrane protein related to *B. subtilis* YkoE, and such transporters are predicted to transport the thiamine precursor hydroxymethylpyrimidine (47).

T. pendens has genes for limited amino acid synthesis. It has a putative cysteine synthase (Tpen_1605) related to a previously characterized *Aeropyrum pernix* enzyme (41), but no serine acetyltransferase. Probably, like *A. pernix*, *T. pendens* uses *O*-phosphoserine rather than *O*-acetylserine as the intermediate in cysteine synthesis. Cysteine synthesis may have been preserved in *T. pendens* so that cysteine can help protect the cell against oxidative stress, a phenomenon that is thought to occur in some parasitic protists (for a review, see reference 38).

Glutamine can be synthesized from glutamate on its tRNA (Tpen_0360 and Tpen_0361) and also by a cytosolic glutamine synthase (Tpen_1089). Cytosolic glutamine synthesis has probably been preserved because of its role as a nitrogen donor. *T. pendens* has six proteins with glutamine amidotransferase domains, including CTP synthase (Tpen_1163) and glucosamine-6-phosphate synthetase (Tpen_0085 and Tpen_1094). Asparagine can be synthesized by a tRNA synthetase-related, archaeal asparagine synthetase (Tpen_1140) (48).

T. pendens has a cobalamin-independent methionine synthase (Tpen_1819) but no homoserine biosynthesis genes; thus, it can probably not make methionine de novo, but it can recycle homoserine resulting from *S*-adenosylmethionine-dependent methylation reactions. Interestingly, *T. pendens* has genes related to genes encoding monomethylamine and trimethylamine methyltransferases in *Methanosarcinales* (Tpen_1211 and Tpen_1467). The *T. pendens* monomethylamine methyltransferase is related to the *Methanosarcina* enzymes, and this is the first time that this protein family has been found outside the *Methanosarcinales*. Both putative methyltransferase genes are adjacent to corrinoid protein genes (Tpen_1212 and Tpen_1468), supporting their function as methyltransferases. Where the *Methanosarcina* proteins have pyrrolysine residues, both proteins of *T. pendens* have leucine. The methyl groups transferred from methylamines could be used to recycle methionine after methylation reactions.

T. pendens can synthesize pyrimidines but not purines de

novo. Carbamoyl phosphate for pyrimidine synthesis is generated by carbamate kinase (Tpen_0172), not by carbamoyl phosphate synthase, similar to the activity in *Pyrococcus furiosus* (61). There are a variety of phosphorylases and phosphoribosyltransferases that could be used for salvage of bases. In addition, *T. pendens* has an open reading frame (Tpen_1649) with 66% similarity to *A. pernix* APE0012, which encodes a broad-range nucleoside kinase as well as a phosphofructokinase (17); thus, nucleosides may also be salvaged. No genes encoding transporters belonging to known families of nucleobase or nucleoside transporters could be identified in the genome.

T. pendens appears to be able to synthesize phospholipids de novo. It may have a modified mevalonate pathway like that predicted for *Methanocaldococcus jannaschii* (13) as it has a homolog of the MJ0044 protein, which was shown to be an isopentyl phosphate kinase (Tpen_0607). It has the enzymes for synthesis of *sn*-glycerol 1-phosphate (Tpen_1231) and geranylgeranyl diphosphate (Tpen_0606) and for attaching the geranylgeranyl groups to glycerol 1-phosphate (Tpen0633, Tpen_0636, and Tpen_1449). Like many archaea, it has only one identifiable CDP-alcohol phosphatidyltransferase (Tpen_0218), and this enzyme is most closely related to archaetidylinositol synthases. *myo*-Inositol-1-phosphate synthase is present (Tpen_1660). It is not known whether *T. pendens* makes additional phospholipids.

Carbohydrate metabolism and transport. *T. pendens* requires a complex growth medium, such as yeast extract, tryptone, or gelatin, and sucrose stimulates growth (64). It was concluded that *T. pendens* grows mainly by peptide fermentation. While *T. pendens* does have enzymes for amino acid degradation, the genome analysis revealed that sugars and sugar polymers may also be important growth substrates for this organism.

One source of evidence that carbohydrates are important growth substrates is the set of transporters encoded in the genome. *T. pendens* genes encode eight ABC transporters belonging to family 1 (Transport Classification Database [http://www.tcdb.org]), which are involved in sugar uptake (Tpen_1055 to Tpen_1057, Tpen_1149 to Tpen_1152, Tpen_1174 to Tpen_1177, Tpen_1255 to Tpen_1257, Tpen_1451 to Tpen_1453, Tpen_1547 to Tpen_1550, Tpen_1588 to Tpen_1590, and Tpen_1617 to Tpen_1619). Within the archaea, only *Haloarcula marismortui* possesses as many family 1 ABC transporters. *T. pendens* also has one ABC transporter belonging to family 2, which is likely to be involved in sugar uptake (Tpen_1208 to Tpen_1210). The only other family 2 ABC transporter in the archaea is a transporter in *Sulfolobus acidocaldarius*. One of the four family 5 ABC transporters in *T. pendens* (Tpen_1676 to Tpen_1680) is similar to a *P. furiosus* cellobiose transporter (24) and a *T. maritima* transporter for mannobiose (TM1223) (39). Two members of the glycoside-pentoside-hexuronide:cation symporter family are also present (Tpen_1599 and Tpen_1831).

T. pendens is the only sequenced crenarchaeote that has the phosphotransferase system (PTS) for carbohydrate uptake (Table 2). The only other sequenced archaeon that has a PTS transporter is *H. marismortui*. *Haloquadratum walsbyi* has enzyme I and HPr proteins of the PTS, but it does not have identifiable PTS transporters. An enzyme I phylogenetic tree shows that the *T. pendens* and halophile proteins are not closely related, suggesting that they were independently acquired through separate lateral transfer events (not shown). The enzyme I gene in *T. pendens* is adjacent to the genes encoding a predicted *N*-acetylglucosamine-6-phosphate deacetylase, suggesting that *N*-acetylglucosamine may be the substrate for this transporter.

T. pendens has a set of 15 glycosyl hydrolases, about the same number as *Sulfolobus* species and greater than the numbers in other *Crenarchaeota*. There are several genes that are involved in starch utilization. One cluster of genes encoding two glycosyl hydrolases and an ABC transporter (Tpen_1451 to Tpen_1454 and Tpen_1458) is similar to a cluster in *Thermococcus* sp. strain B1001 involved in extracellular formation of cyclomaltodextrins, transport of cyclomaltodextrins into the cell, and intracellular degradation of the cyclomaltodextrins (18). In addition, there is an alpha-glucosidase (Tpen_1511) similar to the characterized NAD^+-dependent *T. maritima* enzyme (44).

Cellulose may also be utilized by *T. pendens*. This species has a secreted family 12 glycosyl hydrolase (Tpen_1681; for glycosyl hydrolase classification, see CAZy [http://www.cazy.org]) with weak similarity to cellulases, as well as an ABC transporter with a high level of similarity to a characterized cellobiose transporter in *P. furiosus* (Tpen_1676 to Tpen_1680) (24). Cellobiose and larger oligosaccharides may be broken down by an intracellular beta-glucosidase (Tpen_1494).

Sucrose stimulates the growth of *T. pendens* but does not serve as a sole energy source (64). The enzymes involved in sucrose metabolism could not be identified from the genome sequence. No beta-fructofuranosidase (invertase) or sucrose phosphorylase could be identified, and there is no homolog of PF0132, which encodes the invertase purified from *P. furiosus* (27).

Genes encoding three glycosidases (Tpen_1511, Tpen_1269, and Tpen_1458) and three ABC transporter-associated sugar-binding proteins (Tpen_1055, Tpen_1208, and Tpen_1257) are among the PHX genes of *T. pendens*, providing further evidence of the importance of carbohydrate metabolism. Genes encoding subunits of two peptide ABC transporters are also PHX genes (Tpen_1635 and Tpen_1636, Tpen_1638, Tpen_1245, and Tpen_1247 to Tpen_1249). This reflects the fact that *T. pendens* needs to obtain many amino acids from external sources and utilizes peptides for energy.

Electron transport. *T. pendens* requires sulfur for growth and produces H_2S, and genes encoding some of the potential catalysts for this metabolism can be identified in the genome sequence. *T. pendens* does not have a hydrogenase related to sulfhydrogenase and hydrogenase II of *P. furiosus*, which reduce sulfur as well as protons (31). It also does not possess a sulfide dehydrogenase (30). There is a homolog (Tpen_0143; 48% identity and 66% similarity) of the recently identified CoA-dependent NADPH:sulfur oxidoreductase in *P. furiosus* (52). However, *T. pendens* does not have the *mbx* protein complex that is predicted to transfer electrons from ferredoxin to NADPH. Also, *T. pendens* does not have a homolog of bacterial ferredoxin-$NADP^+$ reductases (COG1018), so the pathway for recycling ferredoxin is unknown. *T. pendens* has a large set of adjacent genes (Tpen_1070 to Tpen_1088) with similarity to genes encoding NADH dehydrogenases and membrane-bound hydrogenases. This cluster may encode one or more multisubunit enzymes that oxidize ferredoxin and trans-

fer the electrons to NADP, a quinone, or another electron carrier.

T. pendens has an operon (Tpen_1121 to Tpen_1123) similar to the *psrABC* genes encoding the polysulfide reductase in *Wolinella succinogenes* (25). The protein similarity is weak, but the three proteins encoded by the *T. pendens* operon belong to the same protein families as the polysulfide reductase subunits. The A subunit is predicted by ProSite to have a twin-arginine signal peptide, so the enzyme probably reduces its substrate extracellularly.

A substantial amount of formate may be produced by fermentative organisms in the environment in which *T. pendens* lives, and *T. pendens* appears to have two pathways for utilizing formate. Like *P. aerophilum* and *Hyperthermus butylicus*, *T. pendens* has a three-subunit, membrane-bound, molybdopterin-dependent formate dehydrogenase. The alpha subunit has a predicted twin-arginine signal peptide, so the topology of the enzyme is likely to be similar to the solved structure of *Escherichia coli* formate dehydrogenase N with formate oxidation occurring outside the cell (for a review, see reference 20). This enzyme likely channels electrons from formate to a quinone or other carrier and then to sulfur as the final electron acceptor.

T. pendens is the only crenarchaeote that has a formate transporter (Tpen_0191). The transporter gene is adjacent to a putative operon (Tpen_0190 to Tpen_0178) with a high level of similarity to the genes encoding *E. coli* hydrogenase 4. In *E. coli*, hydrogenase 4 forms part of the formate hydrogen lyase complex which oxidizes formate and produces hydrogen under conditions in which no electron acceptors other than protons are present. The *T. pendens* operon contains the gene encoding a formate dehydrogenase alpha subunit, providing strong evidence that this operon encodes formate hydrogen lyase. The formate dehydrogenase protein does not have a signal peptide, suggesting that formate oxidation occurs in the cytoplasm, as observed for the *E. coli* complex (for a review, see reference 51). Formate hydrogen lyase contributes to the generation of a proton gradient in two ways: by using protons from inside the cell to make H_2, which then diffuses out of the cell, and under some conditions by pumping protons out of the cell (14). This enzyme complex is expressed in *E. coli* only when no electron acceptors are present, suggesting that *T. pendens* may use this pathway when sulfur is scarce.

T. pendens may also use hydrogen as an electron donor as it contains genes (Tpen_0591 to Tpen_0594) similar to the genes encoding four subunits of a membrane-bound uptake hydrogenase in *Acidianus ambivalens* (26). The *A. ambivalens* hydrogenase is predicted to use a quinone to transfer electrons from hydrogen to sulfur. This type of pathway is common among archaeal autotrophs, and in *T. pendens* it may supplement the energy derived from peptides and sugars.

Signal transduction. Archaea have significantly fewer signal transduction systems than bacteria. On average, 2.63% of archaeal proteomes and 5.4% of bacterial proteomes consist of signal transduction proteins (60). Moreover, it has been shown previously that in archaeal signal transduction a substantially reduced repertoire of sensory (input) and regulatory (output) domains is utilized (59). The median level of archaeal one-component systems in each genome is roughly 50 times greater than the median level of two-component systems, and the majority of these systems regulate gene expression at the transcriptional level (60). Two-component systems have been found only in *Euryarchaeota* and appear to have been laterally transferred from bacteria. In general, crenarchaeal species have fewer signal transduction systems (only 0.7% of the proteome) than euryarchaeotes.

The *T. pendens* genome contains genes encoding 45 one-component systems (regulatory proteins that contain one or more sensory domains [59]) that comprise 2.4% of the proteome. This percentage is more than three times greater than the average percentage for *Crenarchaeota*. Thirty-one (69%) of these systems are encoded by operons containing predominantly enzymatic genes and are predicted to regulate their transcription in response to environmental and intracellular signals. Interestingly, *T. pendens* possesses three families of transcription regulators that have not been found previously in *Crenarchaeota* (Table 2). *T. pendens* encodes more PadR domains than any other crenarchaeal species, possibly indicating that there is a high level of phenolic acid metabolism. Unlike all other crenarchaeotes, *T. pendens* does not have a member of the *fur* family, which is responsible for regulating metal ion uptake, although it does possess an iron-dependent repressor gene (Tpen_0973), which is located beside an iron transporter gene.

Presenilin. A gene encoding a protein belonging to the presenilin family is present in the *T. pendens* genome (Tpen_0870). In eukaryotes presenilin is an integral membrane protease and part of the gamma-secretase complex (63). Mutations in presenilin cause it to cut amyloid precursor peptide in a different place and generate amyloid precursor peptide forms that are more likely to aggregate and form plaques. A family of proteins weakly related to presenilins, known as presenilin homologs, has been identified in eukaryotes and archaea (11, 43). The *T. pendens* protein and a related protein in *P. aerophilum* are not closely related to these presenilin homologs; they represent a new subfamily of presenilins. These crenarchaeal proteins are about 150 amino acids shorter than the mammalian presenilins and lack hydrophilic regions at the N terminus and in an internal loop (Fig. 1). They contain seven to nine predicted transmembrane helices and the conserved YD, LGXGD, and PALP motifs. The gamma-secretase complex includes three other proteins, but none of these proteins are present in *T. pendens* or *P. aerophilum*. Characterization of this new subfamily of presenilins may shed light on the structure and function of the eukaryotic proteins.

DISCUSSION

The genome sequence of *T. pendens* shows that this organism lost enough biosynthetic pathways that it is not a free-living organism and is a commensal that is dependent on another archaeon. Besides the lack of biosynthesis, several nutrient transporters that are not found in any other crenarchaeote are present in *T. pendens*. A lack of biosynthetic ability and an increase in nutrient transport ability are features commonly found in obligate parasites (49). However, *T. pendens* lacks other features of obligate intracellular parasites, such as a reduced genome size (40, 62), a loss of signal transduction and DNA repair proteins (36), an increased percentage of A · T base pairs (35, 62), and a decreased number of fusion proteins (K. Mavromatis and N. C. Kyrpides, unpublished results). It

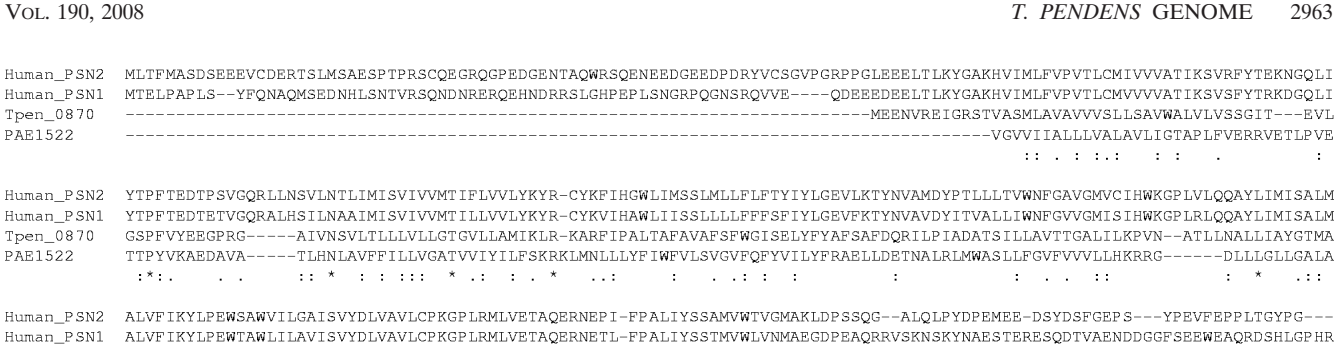

FIG. 1. Alignment of presenilins from a human and *Thermoproteales*. The alignment was constructed with Clustal W (57).

has been hypothesized that one reason that genome size reduction occurs in intracellular bacteria is that there is no possibility for lateral gene transfer from other bacteria (54); however, *T. pendens* does have the opportunity to be exposed to DNA of other bacteria and archaea, and this may help explain the finding that its genome is a normal size. Also, since *T. pendens* is an extracellular symbiont rather than an intracellular symbiont, it may require a larger genome to deal with environmental perturbations. Some extracellular symbionts (e.g., *E. coli*, *Haemophilus influenzae*, and *Pasteurella multocida*) encode biosynthetic pathways for most nutrients (34), while others lack some pathways. For example, *Helicobacter pylori* cannot synthesize purines and some amino acids (58); however, *T. pendens* is unique among extracellular symbionts in the extent of its pathway loss.

Most of the crenarchaeal genomes sequenced so far are genomes of heterotrophs, but they do not show the same extent of loss of biosynthetic pathways as *T. pendens*. For example, *S. acidocaldarius* encodes biosynthetic pathways for all amino acids and nucleotides (5), while *C. symbiosum* has biosynthetic pathways for all amino acids except proline (15). *H. butylicus* has lost most amino acid biosynthetic pathways, but cofactor synthesis pathways appear to be retained. *T. pendens* is unique in its loss of pathways for cofactor biosynthesis, as shown in Table 3.

T. pendens does not appear to be parasitic, as it is not known to cause harm to another organism. However, it is limited to growth in nutrient-rich environments, to the point of depending on a specific organism (*T. tenax*) for an essential nutrient. This type of dependence may be one reason why many microbes cannot be cultivated.

PHX genes in archaea are generally found to be housekeeping genes (21); however, this is not the case in *T. pendens*. Surprisingly, the *T. pendens* PHX genes include many CRISPR-associated genes and genes encoding ABC transporters for carbohydrates and peptides. These findings suggest that *T. pendens* is constantly under attack from viruses in its environment. The large number of CRISPR elements also supports this conclusion. The presence of peptide ABC transporter genes that are PHX genes suggests that *T. pendens* places a higher priority on nutrient acquisition than on maximization of cell growth and division, which is in agreement with its lack of biosynthetic pathways for most amino acids and cofactors.

ACKNOWLEDGMENTS

This work was performed under the auspices of the U.S. Department of Energy's Office of Science, Biological and Environmental Research Program and was supported by the University of California Lawrence Berkeley National Laboratory under contract DE-AC02-05CH11231, by the Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344, and by Los Alamos National Laboratory under contract DE-AC02-06NA25396. J.R., D.S., and B.M. were supported by NASA Astrobiology: Exobiology and Evolutionary Biology grant NNG05GP24G to B.M. I.P. and W.W. were supported by Department of Energy contract DE-FG02-97ER20269. L.E.U. and I.B.Z. were supported by grant GM72285 from the National Institutes of Health. J.G.E. was supported by the DOE Genomes to Life program. L.S.T. and C.D. were supported by the Department of Energy under contract W-7405-ENG-36. M.L. was supported by the Department of Energy under contract DE-AC05-000R22725.

REFERENCES

1. **Adams, M. W. W., J. F. Holden, A. L. Menon, G. J. Schut, A. M. Grunden, C. Hou, A. M. Hutchins, F. E. Jenney, Jr., C. Kim, K. Ma, G. Pan, R. Roy, R. Sapra, S. V. Story, and M. F. J. M. Verhagen.** 2001. Key role for sulfur in peptide metabolism and in regulation of three hydrogenases in the hyperthermophilic archaeon *Pyrococcus furiosus*. J. Bacteriol. **183:**716–724.
2. **Badger, J. H., and G. J. Olsen.** 1999. CRITICA: coding region identification tool invoking comparative analysis. Mol. Biol. Evol. **16:**512–524.
3. **Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, and P. Horvath.** 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315:**1709–1712.
4. **Bräsen, C., C. Urbanke, and P. Schönheit.** 2005. A novel octameric AMP-forming acetyl-CoA synthetase from the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. FEBS Lett. **579:**477–482.
5. **Chen, L., K. Brügger, M. Skovgaard, P. Redder, Q. She, E. Torarinsson, B. Greve, M. Awayez, A. Zibat, H.-P. Klenk, and R. A. Garrett.** 2005. The genome of *Sulfolobus acidocaldarius*, a model organism of the *Crenarchaeota*. J. Bacteriol. **187:**4992–4999.
6. **Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27:**4636–4641.
7. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:**186–194.
8. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8:**175–185.
9. **Finn, R. D., J. Tate, J. Mistry, P. C. Coggill, S. J. Sammut, H.-R. Hotz, G. Ceric, K. Forslund, S. R. Eddy, E. L. L. Sonnhammer, and A. Bateman.** 2008. The Pfam protein families database. Nucleic Acids Res. **36:**D281–D288.
10. **Gordon, D., C. Abajian, and P. Green.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8:**195–202.
11. **Grigorenko, A. P., Y. K. Moliaka, G. I. Korovaitseva, and E. I. Rogaev.** 2002.

Novel class of polytopic proteins with domains associated with putative protease activity. Biochemistry (Moscow) **67:**826–835.
12. **Grigoriev, A.** 1998. Analyzing genomes with cumulative skew diagrams. Nucleic Acids Res. **26:**2286–2290.
13. **Grochowski, L. L., H. Xu, and R. H. White.** 2006. *Methanocaldococcus jannaschii* uses a modified mevalonate pathway for biosynthesis of isopentenyl diphosphate. J. Bacteriol. **188:**3192–3198.
14. **Hakobyan, M., H. Sargsyan, and K. Bagramyan.** 2005. Proton translocation coupled to formate oxidation in anaerobically grown fermenting *Escherichia coli*. Biophys. Chem. **115:**55–61.
15. **Hallam, S. J., K. T. Konstantinidis, N. Putnam, C. Schleper, Y. Watanabe, J. Sugahara, C. Preston, J. de la Torre, P. M. Richardson, and E. F. DeLong.** 2006. Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. Proc. Natl. Acad. Sci. USA **103:**18296–18301.
16. **Han, C. S., and P. Chain.** 2006. Finishing repeat regions automatically with Dupfinisher, p. 141–146. *In* H. R. Arabnia and H. Valafar (ed.), Proceedings of the 2006 International Conference on Bioinformatics and Computational Biology. CSREA Press, Las Vegas, NV.
17. **Hansen, T., L. Arnfors, R. Ladenstein, and P. Schönheit.** 2007. The phosphofructokinase-B (MJ0406) from *Methanocaldococcus jannaschii* represents a nucleoside kinase with a broad substrate specificity. Extremophiles **11:**105–114.
18. **Hashimoto, Y., T. Yamamoto, S. Fujiwara, M. Takagi, and T. Imanaka.** 2001. Extracellular synthesis, specific recognition, and intracellular degradation of cyclomaltodextrins by the hyperthermophilic archaeon *Thermococcus* sp. strain B1001. J. Bacteriol. **183:**5050–5057.
19. **Imanaka, H., A. Yamatsu, T. Fukui, H. Atomi, and T. Imanaka.** 2006. Phosphoenolpyruvate synthase plays an essential role for glycolysis in the modified Embden-Meyerhof pathway in *Thermococcus kodakaraensis*. Mol. Microbiol. **61:**898–909.
20. **Jormakka, M., B. Byrne, and S. Iwata.** 2003. Formate dehydrogenase—a versatile enzyme in changing environments. Curr. Opin. Struct. Biol. **13:**418–423.
21. **Karlin, S., J. Mrazek, J. Ma, and L. Brocchieri.** 2005. Predicted highly expressed genes in archaeal genomes. Proc. Natl. Acad. Sci. USA **102:**7303–7308.
22. **Kato, N., H. Yurimoto, and R. K. Thauer.** 2006. The physiological role of the ribulose monophosphate pathway in bacteria and archaea. Biosci. Biotechnol. Biochem. **70:**10–21.
23. **Kletzin, A., and M. W. W. Adams.** 1996. Molecular phylogenetic characterization of pyruvate and 2-ketoisovalerate ferredoxin oxidoreductases from *Pyrococcus furiosus* and pyruvate ferredoxin oxidoreductase from *Thermotoga maritima*. J. Bacteriol. **178:**248–257.
24. **Koning, S. M., M. G. L. Elferink, W. N. Konings, and A. J. M. Driessen.** 2001. Cellobiose uptake in the hyperthermophilic archaeon *Pyrococcus furiosus* is mediated by an inducible, high-affinity ABC transporter. J. Bacteriol. **183:**4979–4984.
25. **Krafft, T., M. Bokranz, O. Klimmek, I. Schröder, F. Fahrenholz, E. Kojro, and A. Kröger.** 1992. Cloning and nucleotide sequence of the *psrA* gene of *Wolinella succinogenes* polysulphide reductase. Eur. J. Biochem. **206:**503–510.
26. **Laska, S., F. Lottspeich, and A. Kletzin.** 2003. Membrane-bound hydrogenase and sulfur reductase of the hyperthermophilic and acidophilic archaeon *Acidianus ambivalens*. Microbiology **149:**2357–2371.
27. **Lee, H.-S., K. R. Shockley, G. J. Schut, S. B. Conners, C. I. Montero, M. R. Johnson, C.-J. Chou, S. L. Bridger, N. Wigner, S. D. Brehm, F. E. Jenney, Jr., D. A. Comfort, R. M. Kelly, and M. W. W. Adams.** 2006. Transcriptional and biochemical analysis of starch metabolism in the hyperthermophilic archaeon *Pyrococcus furiosus*. J. Bacteriol. **188:**2115–2125.
28. **Letunic, I., R. R. Copley, B. Pils, S. Pinkert, J. Schultz, and P. Bork.** 2006. SMART 5: domains in the context of genomes and networks. Nucleic Acids Res. **34:**D257–D260.
29. **Liolios, K., N. Tavernarakis, P. Hugenholtz, and N. C. Kyrpides.** 2006. The Genomes On Line Database (GOLD) v. 2: a monitor of genome projects worldwide. Nucleic Acids Res. **34:**D332–D334.
30. **Ma, K., and M. W. W. Adams.** 1994. Sulfide dehydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus*: a new multifunctional enzyme involved in the reduction of elemental sulfur. J. Bacteriol. **176:**6509–6517.
31. **Ma, K., R. Weiss, and M. W. W. Adams.** 2000. Characterization of hydrogenase II from the hyperthermophilic archaeon *Pyrococcus furiosus* and assessment of its role in sulfur reduction. J. Bacteriol. **182:**1864–1871.
32. **Mai, X., and M. W. W. Adams.** 1996. Purification and characterization of two reversible and ADP-dependent acetyl coenzyme A synthetases from the hyperthermophilic archaeon *Pyrococcus furiosus*. J. Bacteriol. **178:**5897–5903.
33. **Markowitz, V. M., E. Szeto, K. Palaniappan, Y. Grechkin, K. Chu, I.-M. A. Chen, I. Dubchak, I. Anderson, A. Lykidis, K. Mavromatis, N. N. Ivanova, and N. C. Kyrpides.** 2006. The integrated microbial genomes (IMG) system in 2007: data content and analysis tool extensions. Nucleic Acids Res. **34:**D344–D348.
34. **May, B. J., Q. Zhang, L. L. Li, M. L. Paustian, T. S. Whittam, and V. Kapur.** 2001. Complete genomic sequence of *Pasteurella multocida* Pm70. Proc. Natl. Acad. Sci. USA **98:**3460–3465.
35. **Moran, N. A.** 1996. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. USA **93:**2873–2878.
36. **Moran, N. A., and J. J. Wernegreen.** 2000. Lifestyle evolution in symbiotic bacteria: insights from genomics. Trends Ecol. Evol. **15:**321–326.
37. **Mukund, S., and M. W. W. Adams.** 1995. Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon *Pyrococcus furiosus*. J. Biol. Chem. **270:**8389–8392.
38. **Müller, S., E. Liebau, R. D. Walter, and R. L. Krauth-Siegel.** 2003. Thiol-based redox metabolism of protozoan parasites. Trends Parasitol. **19:**320–328.
39. **Nanavati, D. M., K. Thirangoon, and K. M. Noll.** 2006. Several archaeal homologs of putative oligopeptide-binding proteins encoded by *Thermotoga maritima* bind sugars. Appl. Environ. Microbiol. **72:**1336–1345.
40. **Ochman, H., and N. A. Moran.** 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science **292:**1096–1099.
41. **Oda, Y., K. Mino, K. Ishikawa, and M. Ataka.** 2005. Three-dimensional structure of a new enzyme, O-phosphoserine sulfhydrylase, involved in L-cysteine synthesis by a hyperthermophilic archaeon, *Aeropyrum pernix* K1, at 2.0 Å resolution. J. Mol. Biol. **351:**334–344.
42. **Ortmann, A. C., B. Wiedenheft, T. Douglas, and M. Young.** 2006. Hot crenarchaeal viruses reveal deep evolutionary connections. Nat. Rev. Microbiol. **4:**520–528.
43. **Ponting, C. P., M. Hutton, A. Nyborg, M. Baker, K. Jansen, and T. E. Golde.** 2002. Identification of a novel family of presenilin homologues. Hum. Mol. Genet. **11:**1037–1044.
44. **Raasch, C., W. Streit, J. Schanzer, M. Bibel, U. Gosslar, and W. Liebl.** 2000. *Thermotoga maritima* AglA, an extremely thermostable NAD^+-, Mn^{2+}-, and thiol-dependent α-glucosidase. Extremophiles **4:**189–200.
45. **Rice, P., I. Longden, and A. Bleasby.** 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. **16:**276–277.
46. **Robinson, N. P., I. Dionne, M. Lundgren, V. L. Marsh, R. Bernander, and S. D. Bell.** 2004. Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. Cell **116:**25–38.
47. **Rodionov, D. A., A. G. Vitreschak, A. A. Mironov, and M. S. Gelfand.** 2002. Comparative genomics of thiamin biosynthesis in prokaryotes. New genes and regulatory mechanisms. J. Biol. Chem. **277:**48949–48959.
48. **Roy, H., H. D. Becker, J. Reinbolt, and D. Kern.** 2003. When contemporary aminoacyl-tRNA synthetases invent their cognate amino acid metabolism. Proc. Natl. Acad. Sci. USA **100:**9837–9842.
49. **Sakharkar, K. R., P. K. Dhar, and V. T. Chow.** 2004. Genome reduction in prokaryotic obligatory intracellular parasites of humans: a comparative analysis. Int. J. Syst. Evol. Microbiol. **54:**1937–1941.
50. **Sato, T., H. Atomi, and T. Imanaka.** 2007. Archaeal type III rubiscos function in a pathway for AMP metabolism. Science **315:**1003–1006.
51. **Sawers, R. G.** 2005. Formate and its role in hydrogen production in *Escherichia coli.* Biochem. Soc. Trans. **33:**42–46.
52. **Schut, G. J., S. L. Bridger, and M. W. W. Adams.** 2007. Insights into the metabolism of elemental sulfur by the hyperthermophilic archaeon *Pyrococcus furiosus*: characterization of a coenzyme A-dependent NAD(P)H sulfur oxidoreductase. J. Bacteriol. **189:**4431–4441.
53. **Siguier, P., J. Perochon, L. Lestrade, J. Mahillon, and M. Chandler.** 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. **34:**D32–D36.
54. **Tamas, I., L. Klasson, B. Canbäck, A. K. Näslund, A.-S. Eriksson, J. J. Wernegreen, J. P. Sandström, N. A. Moran, and S. G. E. Andersson.** 2002. 50 million years of genomic stasis in endosymbiotic bacteria. Science **296:**2376–2379.
55. **Tatusov, R. L., N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin, E. V. Koonin, D. M. Krylov, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, B. S. Rao, S. Smirnov, A. V. Sverdlov, S. Vasudevan, Y. I. Wolf, J. J. Yin, and D. A. Natale.** 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4:**41.
56. **Thomas, J. M., D. Horspool, G. Brown, V. Tcherepanov, and C. Upton.** 2007. GraphDNA: a Java program for graphical display of DNA composition analyses. BMC Bioinformatics **8:**21.
57. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
58. **Tomb, J.-F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzegerald, N. Lee, M. D. Adams, E. K. Hickey, D. E. Berg, J. D. Gocayne, T. R. Utterback, J. D. Peterson, J. M. Kelley, M. D. Cotton, J. M. Weidman, C. Fujii, C. Bowman, L. Watthey, E. Wallin, W. S. Hayes, M. Borodovsky, P. D. Karp, H. O. Smith, C. M. Fraser, and J. C. Venter.** 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature **388:**539–547.

59. **Ulrich, L. E., E. V. Koonin, and I. B. Zhulin.** 2005. One-component systems dominate signal transduction in prokaryotes. Trends Microbiol. **13:**52–56.
60. **Ulrich, L. E., and I. B. Zhulin.** 2007. MiST: a microbial signal transduction database. Nucleic Acids Res. **35:**D386–D390.
61. **Uriarte, M., A. Marina, S. Ramón-Maiques, I. Fita, and V. Rubio.** 1999. The carbamoyl-phosphate synthetase of *Pyrococcus furiosus* is enzymologically and structurally a carbamate kinase. J. Biol. Chem. **274:**16295–16303.
62. **Waters, E., M. J. Hohn, I. Ahel, D. E. Graham, M. D. Adams, M. Barnstead, K. Y. Beeson, L. Bibbs, R. Bolanos, M. Keller, K. Kretz, X. Lin, E. Mathur, J. Ni, M. Podar, T. Richardson, G. G. Sutton, M. Simon, D. Söll, K. O. Stetter, J. M. Short, and M. Noordewier.** 2003. The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. Proc. Natl. Acad. Sci. USA **100:**12984–12988.
63. **Wolfe, M. S.** 2006. The gamma-secretase complex: membrane-embedded proteolytic ensemble. Biochemistry **45:**7931–7939.
64. **Zillig, W., A. Gierl, G. Schreiber, S. Wunderl, D. Janekovic, K. O. Stetter, and H. P. Klenk.** 1983. The archaebacterium *Thermofilum pendens* represents a novel genus of the thermophilic, anaerobic sulfur respiring Thermoproteales. Syst. Appl. Microbiol. **4:**79–87.