Title
Genome sequence of Thermofilum pendens reveals an exceptional loss of biosynthetic pathways without genome reduction

Permalink
https://escholarship.org/uc/item/7qd3t2fr

Author
Fenner, Marsha W

Publication Date
2008-01-01

Peer reviewed
Genome sequence of *Thermofilum pendens* reveals an exceptional loss of biosynthetic pathways without genome reduction


1Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598

2Virginia Bioinformatics Institute and Departments of Biochemistry and Biological Sciences, Genetics, Bioinformatics and Computational Biology Graduate Program, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061

3Dept. of Microbiology, University of Georgia, Athens, GA 30602

4Dept. of Microbiology, University of Illinois, Urbana, IL 61801

5Joint Institute for Computational Sciences, University of Tennessee – Oak Ridge National Laboratory, Oak Ridge, TN 37831

6Bioscience Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831

7Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545

8Current address: Dept. of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA 16802

Running title: *Thermofilum pendens* genome

*Corresponding author: Iain Anderson

Joint Genome Institute
2800 Mitchell Dr.
Walnut Creek, CA 94598
Tel: 925-296-5784
Fax: 925-296-5850
Email: IJAnderson@lbl.gov
Abstract

We report the complete genome of *Thermofilum pendens*, a deep-branching, hyperthermophilic member of the order Thermoproteales within 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. *T. pendens* does not display other features 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 to utilize peptides as an energy source, but the genome reveals 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 gain energy from sulfur reduction with hydrogen and formate as electron donors. It may also be capable of sulfur-independent growth on formate with formate hydrogenlyase. Additional novel features are the presence of a monomethylamine:corrinoid methyltransferase, the first time this enzyme has been found outside of Methanosarcinales, and a presenilin-related protein. Predicted highly expressed proteins do not include housekeeping genes, and instead include ABC transporters for carbohydrates and peptides, and CRISPR-associated proteins.
Introduction

Crenarchaeota is one of the two major divisions of the Archaea, and it is the least well-represented in terms of genome sequences. Only six crenarchaeal complete genomes have been published so far, and three of these are from the genus Sulfolobus. Within the order Thermoproteales, only one organism has been completely sequenced so far, Pyrobaculum aerophilum, although several more species of Pyrobaculum, Caldivirga maquilingensis, and Thermoproteus tenax are currently being sequenced (23). *Thermofilum pendens* represents a deep branch within the order Thermoproteales, and the organism grows only in the presence of a fraction of the polar lipids of *T. tenax* (55), a property that has not been seen before in archaea. Therefore, it was an attractive sequencing target. We report here the genome sequence and analysis of the type strain *Thermofilum 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 (55). 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 CO₂ and H₂S are produced, similar to other anaerobic members of the Crenarchaeota and the euryarchaeal family Thermococcaceae. The genome reveals an organism that appears to have lost many biosynthetic capabilities yet does not have a reduced genome size compared to other Crenarchaeota.
Materials and Methods

Frozen *T. pendens* Hrk5 cells were obtained from Dr. Karl Stetter. Cells were resuspended in 0.25 M sucrose in TE buffer. Sodium dodecyl sulfate was added to 1% concentration, and cells were lysed by three cycles of freezing and thawing. Proteinase K was added to 50 µg/ml, and the lysate was incubated at 60°C for 30 minutes. Undigested proteins were precipitated by addition of NaCl to 0.5 M concentration and removed by centrifugation. Nucleic acids in the supernatant were precipitated by the 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 TE buffer and sent to the Joint Genome Institute.

The genome of *T. pendens* was sequenced at the Joint Genome Institute (JGI) using a combination of 3kb, 6kb and 40kb (fosmid) DNA libraries. All general aspects of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov/. Draft assemblies were based on 21478 total reads. All three libraries provided 11x coverage of the genome. The Phred/Phrap/Consed software package (www.phrap.com) was used for sequence assembly and quality assessment (5-7). After the shotgun stage, reads were assembled with parallel phrap (High Performance Software, LLC). Possible mis-assemblies were corrected with Dupfinisher (11) or transposon bombing of bridging clones (Epicentre Biotechnologies, Madison, WI). Gaps between contigs were closed by editing in Consed, custom primer walk or PCR amplification (Roche Applied Science, Indianapolis, IN). A total of 465 additional reactions were necessary to close gaps and to raise the quality of the finished sequence.
The sequences of *T. pendens*, consisting of one chromosome and one plasmid, can be accessed using the GenBank accession numbers NC_008698 and NC_008696 and the Genomes On Line Database accession number Gc00473. Genes were identified using a combination of Critica (2) and Glimmer (4) followed by a round of manual curation.

Analysis of the *T. pendens* genome was carried out with the Integrated Microbial Genomes (IMG) system (27). Protein families unique to *T. pendens* or missing from *T. pendens* but present in other Crenarchaeota were identified with the phylogenetic profiler in IMG. Analysis of signal transduction was carried out using the MiST database (51). A cumulative GC-skew plot was generated with a 35 kb sliding window using GraphDNA (48).

Predicted highly expressed (PHX) genes were determined with the EMBOSS (38) programs cusp and cai. The training set of PHX genes was compiled from Karlin et al. (16). The genes with a codon adaptation index (CAI) in the top 5% were taken to be PHX genes.

Results
General features

The genome of *T. pendens* Hrk5 consists of a circular chromosome of 1.78 Mbp and a plasmid of 31,504 bp (Table 1). The G+C percentage is 58%, higher than that of other Crenarchaeota. 1,923 genes were identified, of which 1,883 encode proteins. The percentage of the genome devoted to encoding genes is 91%, slightly higher than for other sequenced Crenarchaeota. About 59% of protein-coding genes begin with an AUG codon, 32% with a GUG, and 10% with UUG. About 66% of protein-coding genes have
COG domains, and about 63% have Pfam domains, similar to other archaeal genomes. There is one copy of each ribosomal RNA. T. pendens has the highest percentage of fusion genes among the Crenarchaeota. Several proteins in T. pendens have not been found in Crenarchaeotes or in archaea before (Table 2) and several proteins found in all other crenarchaeota are missing from the T. pendens genome (Table 3).

The plasmid is predicted to encode 52 proteins, of which only two have similarity to proteins in the GenBank nonredundant protein database. Tpen_1849 is similar to a T. pendens chromosomal protein of unknown function (Tpen_0735), and Tpen_1875 is a predicted helicase. In addition, Tpen_1891 is predicted to be a site-specific recombinase (COG4974). The function of the plasmid and whether or not it is beneficial to the host is currently unknown.

Cumulative GC-skew analysis of the T. pendens genome was used to identify the potential origin(s) of replication (9). A global minimum was located at position 488,884 which is near a 478 bp intergenic region between positions 487,890 and 488,368. The intergenic spacer contains several repetitive sequences similar to conserved crenarchaeal origin recognition boxes (39).

No repetitive elements were found when the ISfinder database (46) was searched with the T. pendens coding sequences. However, within a 100,000 bp section of the genome there are twelve stretches of clustered regularly interspaced short palindromic repeat (CRISPR) elements interspersed with protein-coding genes. Among the 97 predicted highly expressed (PHX) genes (supplementary table 1) there are 6 CRISPR-associated proteins (Tpen_1263, Tpen_1287, Tpen_1288, Tpen_1316, Tpen_1342, Tpen_1356). Interestingly, a group of seven consecutive genes (Tpen_1287-Tpen_1293) including two
CRISPR-associated proteins is PHX. While other thermophilic archaea have similar numbers of CRISPR-associated proteins within their genomes, most do not have CRISPR-associated proteins as PHX genes, although *Staphylothermus marinus* and *P. aerophilum* do (4 and 2 genes respectively). However *T. pendens* has the highest number of CRISPR-associated genes as 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 (reviewed in 35).

Central metabolism

*T. pendens* contains complete glycolysis and gluconeogenesis pathways. Glyceraldehyde 3-phosphate:ferredoxin oxidoreductase, found in some archael hyperthermophiles as an alternative step in glycolysis (30), is not present in *T. pendens*. Phosphoenolpyruvate (PEP) synthase, used as the last step in glycolysis in *Thermococcus kodakaraensis* (14), is present in *T. pendens* (Tpen_0588). PEP 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 common in archaea (reviewed in 17). *T. pendens* encodes two ribose 5-phosphate isomerases, one RpiA-type (Tpen_0327) and one RpiB-type (Tpen_1241). This is the first time an RpiB has been found in archaea. The RpiB is adjacent to uridine phosphorylase (Tpen_1240) suggesting a function in nucleoside utilization. Under conditions in which ribonucleosides are present in excess, 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 archaeb.

ATP can be generated from pyruvate through the consecutive action of pyruvate:ferredoxin oxidoreductase (PFOR) and ADP-forming acetyl-CoA synthase similar to Thermococcales (26). The *T. pendens* PFOR (Tpen_0571-0574) is similar to the characterized *Thermotoga maritima* enzyme (18). In Thermococcales and Crenarchaeota, ADP-forming acetyl-CoA synthase is split into two subunits, alpha and beta. *T. pendens* contains one alpha subunit (Tpen_0336), one beta subunit (Tpen_0109), and one protein with both alpha and beta subunits fused together (Tpen_0602). Two AMP-forming acyl-CoA synthases are also present (Tpen_0893, Tpen_1611). *T. pendens* has four other enzymes similar to pyruvate:ferredoxin oxidoreductase (Tpen_0540-0543, Tpen_0781-0782, Tpen_0856-0857, Tpen_1455-1456), and these 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 (26). Four aldehyde:ferredoxin oxidoreductases are also present (Tpen_0094, Tpen_0176, Tpen_1413, Tpen_1817), and these could be involved in peptide fermentation (1). The 2-oxoacid oxidoreductases produce aldehydes which are then converted to acids. Reduced ferredoxin is produced but there is no ATP production by this pathway.

*T. pendens* appears to assimilate glycerol. There is a glycerol kinase (Tpen_1128) adjacent to subunit A of glycerol 3-phosphate dehydrogenase (Tpen_1127). Next to these are three genes with similarity to subunits B, C, and D of succinate dehydrogenases (Tpen_1124-1126). A gene encoding subunit A of succinate dehydrogenase is not found
within 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 other acceptor.

Unlike most of the sequenced Crenarchaeota, *T. pendens* has ribulose 1,5-bisphosphate carboxylase (Rubisco, Tpen_1227). It also has the recently discovered enzymes involved in conversion of the ribose phosphate group of AMP to ribulose 1,5-bisphosphate (43): 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 AMP. The *T. pendens* AMP-forming acetyl-CoA synthetase (Tpen_0893) has very high similarity to the *P. aerophilum* characterized enzyme (3). A large amount of AMP may also be generated by ribose-phosphate pyrophosphokinase, required for pyrimidine synthesis, and phosphoribosyltransferases.

**Biosynthesis**

*T. pendens* is known to require an extract of *T. tenax* for growth. While the specific compound required from *T. tenax* could not be identified, the genome reveals a vast reduction in its ability 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 (supplementary table 2). 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 IMG. *T.*
*T. pendens* possesses only 11 of these COGs. The only organisms with fewer 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 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 for *T. pendens* have homologs of the pyridoxine biosynthesis genes PDX1 and PDX2 (*yaaD* and *yaaE* in *Bacillus subtilis*) and the bifunctional coenzyme A biosynthetic enzyme phosphopantothenoylcysteine synthetase/decarboxylase. Table 3 lists the COGs missing from *T. pendens* that are found in all other sequenced Crenarchaeota. The top nine of these COGs are involved in pyridoxine, coenzyme A, riboflavin, ubiquinone, and thiamine biosynthesis. In addition, most Crenarchaeota have homologs of several heme biosynthetic enzymes, but *T. pendens* lacks these. They are not found in Table 3 because they are also missing from the *Staphylothermus marinus* genome. Also COG1731, archaeal riboflavin synthase, is not found in Table 3 because it is missing from both *T. pendens* and *Cenarchaeum symbiosum*; however, *C. symbiosum* has the bacterial-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 to have a bioY family biotin transporter and a riboflavin transporter (see Table 2). In addition *T. pendens* has an expansion of ABC transporters related to those involved in cobalt uptake. While most Crenarchaeota have 0-2 representatives from this
family, *T. pendens* has 7. 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 (40).

*T. pendens* has genes for limited amino acid synthesis. There is a putative cysteine synthase (Tpen_1605) related to a characterized *A. pernix* enzyme (34), 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 to protect the cell against oxidative stress, as is thought to occur in some parasitic protists (reviewed in 31).

Glutamine can be synthesized from glutamate on its tRNA (Tpen_0360-0361), and also by a cytosolic glutamine synthase (Tpen_1089). Cytosolic glutamine synthesis has probably been preserved for 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, Tpen_1094). Asparagine can be synthesized by a tRNA synthetase-related, archaeal asparagine synthetase (Tpen_1140; 41).

*T. pendens* has a 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 monomethylamine and trimethylamine methyltransferases from Methanosarcinales (Tpen_1211, Tpen_1467). The *T. pendens* monomethylamine methyltransferase is related to the Methanosarcina enzymes, and this is the first time this protein family has been found outside the Methanosarcinales. Both putative
methyltransferases are adjacent to corrinoid proteins (Tpen_1212, Tpen_1468),
supporting their function as methyltransferases. Where the Methanosarcina proteins have
pyrrolysine residues, *T. pendens* has leucine in both proteins. The methyl groups
transferred from methylamines could be used to recycle methionine after methylation
reactions.

*T. pendens* can synthesize pyrimidines *de novo* but not purines. Carbamoyl phosphate
for pyrimidine synthesis is generated by carbamate kinase (Tpen_0172), not by
carbamoyl phosphate synthase, similar to *P. furiosus* (52). There are a variety of
phosphorylases and phosphoribosyltransferases that could be used for salvage of bases.
In addition *T. pendens* has an ORF (Tpen_1649) with 66% similarity to *A. pernix*
APE0012, which is a broad-range nucleoside kinase as well as a phosphofructokinase
(12), thus nucleosides may also be salvaged. No 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 as predicted for *Methanocaldococcus jannaschii*
(Grochowski et al., 2006) 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,
Tpen_1449). Like many archaea it has only one identifiable CDP-alcohol
phosphatidyltransferase (Tpen_0218), and this is most closely related to
archaetidylinositol synthases. Myo-inositol-1-phosphate synthase is present
(Tpen_1660). It is unknown 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 (55). It was concluded that *T. pendens* grows mainly by peptide fermentation. While *T. pendens* does have enzymes for amino acid degradation, the genome reveals 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* encodes 8 ABC transporters of family 1, which are involved in sugar uptake (Tpen_1055-1057, Tpen_1149-1152, Tpen_1174-1177, Tpen_1255-1257, Tpen_1451-1453, Tpen_1547-1550, Tpen_1588-1590, Tpen_1617-1619). Within the archaea, only *Haloarcula marismortui* possesses as many family 1 ABC transporters. *T. pendens* also has one ABC transporter from family 2, likely to be involved in sugar uptake (Tpen_1208-1210). The only other family 2 ABC transporter in the archaea is in *Sulfolobus acidocaldarius*. One of the four family 5 ABC transporters in *T. pendens* (Tpen_1676-1680) is similar to a *P. furiosus* cellobiose transporter (19) and a *T. maritima* transporter for mannobiose (TM1223; 32). Two members of the glycoside-pentoside-hexuronide (GPH): cation symporter family are also present (Tpen_1599, Tpen_1831).

*T. pendens* is the only sequenced crenarchaeote to have the phosphotransferase system (PTS) for carbohydrate uptake (see Table 2). The only other sequenced archaeon to have a PTS transporter is *H. marismortui*. *Haloquadratum walsbyi* has Enzyme I and HPp proteins from the PTS, but it does not have identifiable PTS transporters. A phylogenetic tree of Enzyme I shows that the *T. pendens* and halophile proteins are not closely related,
suggesting they were independently acquired through separate lateral transfer events (not shown). *T. pendens* Enzyme I is adjacent to 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 other Crenarchaeota. There are several genes involved in starch utilization. One cluster of genes encoding two glycosyl hydrolases and an ABC transporter (Tpen_1451-1454, 1458) is similar to a cluster from *Thermococcus* sp. B1001 involved in extracellular formation of cyclomaltodextrins, transport of cyclomaltodextrins into the cell, and intracellular degradation of the cyclomaltodextrins (13). In addition, there is an alpha-glucosidase (Tpen_1511) similar to the characterized NAD+-dependent *T. maritima* enzyme (37).

Cellulose may also be utilized by *T. pendens*. There is a secreted family 12 glycosyl hydrolase (Tpen_1681) with weak similarity to cellulases as well as an ABC transporter with high similarity to a characterized cellobiose transporter from *P. furiosus* (Tpen_1676-1680; 19). Cellobiose and larger oligosaccharides may be broken down by an intracellular beta-glucosidase (Tpen_1494).

Sucrose stimulates growth of *T. pendens* but does not serve as the sole energy source (55). The enzymes involved in sucrose metabolism can not be identified from the genome sequence. No beta-fructofuranosidase (invertase) or sucrose phosphorylase can be identified, and there is no homolog of PF0132, which encodes the invertase purified from *P. furiosus* (22).
Three glycosidases (Tpen_1511, Tpen_1269, Tpen_1458) and three ABC transporter-associated sugar-binding proteins (Tpen_1055, Tpen_1208, Tpen_1257) are among the PHX genes of *T. pendens*, providing further evidence of the importance of carbohydrate metabolism. Subunits of two peptide ABC transporters are also PHX genes (Tpen_1635-1636, Tpen_1638, Tpen_1245, Tpen_1247-1249). This reflects the need of *T. pendens* to obtain many amino acids from external sources and the utilization of peptides for energy.

Electron transport

*T. pendens* requires sulfur for growth and produces H$_2$S, and 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 (25). It also does not possess a sulfide dehydrogenase (24). There is a homolog (Tpen_0143, 48% identity, 66% similarity) of the recently identified CoA-dependent NADPH:sulfur oxidoreductase from *P. furiosus* (45). 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-Tpen_1088) with similarity to NADH dehydrogenases and membrane bound hydrogenases. This cluster may encode one or more multisubunit enzymes that oxidize ferredoxin and transfer the electrons to NADP, a quinone, or another electron carrier.

*T. pendens* has an operon (Tpen_1121-Tpen_1123) similar to the psrABC genes of the polysulfide reductase from *Wolinella succinogenes* (20). The protein similarity is weak, but the three proteins in 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 environments 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 *E. coli* formate dehydrogenase N with formate oxidation occurring outside the cell (reviewed in 15). 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 to have a formate transporter (Tpen_0191). The transporter is found adjacent to a putative operon (Tpen_0190-Tpen_0178) with high similarity to *E. coli* hydrogenase 4. In *E. coli*, hydrogenase 4 forms part of the formate hydrogenlyase complex which oxidizes formate and produces hydrogen under conditions in which no electron acceptors other than protons are present. The *T. pendens* operon contains a formate dehydrogenase alpha subunit, providing strong evidence that this operon encodes formate hydrogenlyase. The formate dehydrogenase protein does not have a signal peptide, suggesting that formate oxidation occurs in the cytoplasm as for the *E. coli* complex (reviewed in 44). Formate hydrogenlyase contributes to the generation of a proton gradient in two ways: by using protons from inside the cell to make H₂, which then diffuses out of the cell, and under some conditions by pumping protons out of the cell (Hakobyan et al., 2005). 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_0591-0594) similar to the four subunits of a membrane-bound uptake hydrogenase from *Acidianus ambivalens* (21). 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 (51). Moreover, it has been shown previously that archaeal signal transduction utilizes a substantially reduced repertoire of sensory (input) and regulatory (output) domains (50). The median level of archaeal, one-component systems per genome is roughly 50 times greater than that of two-component systems, and the majority of these systems regulate gene expression at the transcriptional level (51). Two-component systems have only been found 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 45 one-component systems – regulatory proteins that contain one or more sensory domains (50) – that comprise 2.4% of its proteome. This percentage is over three times as high as the average for Crenarchaeota. Thirty-one (69%) of these are located in 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 previously been found in Crenarchaeota (Table 2). *T. pendens* contains more PadR domains than any other Crenarchaeal species, possibly indicating 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 metal-ion uptake, although it does possess an iron-dependent repressor (Tpen_0973), which is positioned beside an iron transporter.

**Presenilin**

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 (54). Mutations in presenilin cause it to cut amyloid precursor peptide (APP) in a different place and generate APP 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 (8, 36). The *T. pendens* protein and a related protein from *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, lacking hydrophilic regions at the N-terminus and in an internal loop (Figure 1). They contain 7-9 predicted transmembrane helices and the conserved YD, LGXGD, and PALP motifs. The gamma-secretase complex includes three other proteins, but none of these 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 a loss of biosynthetic pathways to a degree that it is no longer a free-living organism, and has become a commensal, dependent on another archaeon. Along with this lack of biosynthesis, several nutrient transporters that are not found in any other crenarchaeote are present in *T. pendens*. A lack of biosynthetic capability and an increase in nutrient transport capability are features commonly found in obligate parasites (42). However *T. pendens* lacks other features of obligate intracellular parasites, such as genome size reduction (33, 53), loss of signal transduction and DNA repair proteins (29), increased percentage of A-T base pairs (28, 53), and decreased number of fusion proteins (Mavromatis, K. and Kyrpides, N. C., unpublished results). It is hypothesized that one reason genome size reduction in intracellular bacteria occurs because there is no possibility for lateral gene transfer from other bacteria (47), however *T. pendens* does have the opportunity to be exposed to DNA of other bacteria and archaea, and this may help to explain its maintenance of a normal genome size. Also since *T. pendens* is an extracellular rather than intracellular symbiont, it may require a larger genome to deal with environmental perturbations. *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 for an essential nutrient. This type of dependence may be one reason why many microbes are not able to be cultivated.

Predicted highly expressed genes in archaea are generally found to be housekeeping genes (16), however this is not the case in *T. pendens*. Surprisingly *T. pendens* PHX genes contain many CRISPR associated genes and 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 transporters as 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 US Department of Energy's 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. J. R., D. S., and B. M. were supported by a NASA Astrobiology: Exobiology and Evolutionary Biology grant NNG05GP24G to B. M. I. P. and W. W. were supported by DOE contract number DE-FG02-97ER20269. L. E. U. and I. B. Z. were supported by grant number 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


47. Tamas, I., Klasson, L., Canbäck, B., Näslund, A. K., Eriksson, A.-S.,
2002. 50 million years of genomic stasis in endosymbiotic bacteria. Science
296:2376-2379.
GraphDNA: a Java program for graphical display of DNA composition analyses.
BMC Bioinformatics 8:21.
49. Thompson, J. D., Higgins, D. G., and Gibson, T. J. 1994. CLUSTAL W:
improving the sensitivity of progressive multiple sequence alignment through
sequence weighting, position-specific gap penalties and weight matrix choice.
52. Uriarte, M., Marina, A., Ramón-Maiques, S., Fita, I., and Rubio, V. 1999. The
carbamoyl-phosphate synthetase of Pyrococcus furiosus is enzymologically and
53. Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M.,
Beeson, K. Y., Bibbs, L., Bolanos, R., Keller, M., Kretz, K., Lin, X., Mathur, E.,
Ni, J., Podar, M., Richardson, T., Sutton, G. G., Simon, M., Söll, D., Stetter, K.
O., Short, J. M., and Noordewier, M. 2003. The genome of Nanoarchaeum


Table 1. General statistics

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome size (bp)</td>
<td>1781889</td>
</tr>
<tr>
<td>Chromosome G+C (bp)</td>
<td>1027538 (57.6%)</td>
</tr>
<tr>
<td>Plasmid size (bp)</td>
<td>31504</td>
</tr>
<tr>
<td>Plasmid G+C (bp)</td>
<td>17813 (56.5%)</td>
</tr>
<tr>
<td>Total genome size (bp)</td>
<td>1813393</td>
</tr>
<tr>
<td>Total genome G+C (bp)</td>
<td>1045351 (57.6%)</td>
</tr>
<tr>
<td>Total genes</td>
<td>1923</td>
</tr>
<tr>
<td>RNA genes</td>
<td>40 (2.1%)</td>
</tr>
<tr>
<td>Protein-coding genes</td>
<td>1883 (97.9%)</td>
</tr>
<tr>
<td>Genes with function prediction</td>
<td>1170 (60.8%)</td>
</tr>
<tr>
<td>Genes in ortholog clusters</td>
<td>1541 (80.1%)</td>
</tr>
<tr>
<td>Genes in paralog clusters</td>
<td>805 (41.9%)</td>
</tr>
<tr>
<td>Genes assigned to COGs</td>
<td>1264 (65.7%)</td>
</tr>
<tr>
<td>Genes assigned Pfam domains</td>
<td>1209 (62.9%)</td>
</tr>
<tr>
<td>Genes with signal peptides</td>
<td>134 (7.0%)</td>
</tr>
<tr>
<td>Genes with transmembrane helices</td>
<td>437 (22.7%)</td>
</tr>
<tr>
<td>Fusion genes</td>
<td>79 (4.11%)</td>
</tr>
</tbody>
</table>
Table 2. Unique genes in T. pendens with COG hits

<table>
<thead>
<tr>
<th>CDS</th>
<th>COG</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>COGs not found in any other sequenced Archaea</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tpen_1241</td>
<td>0698</td>
<td>Ribose 5-phosphate isomerase rpiB</td>
</tr>
<tr>
<td>Tpen_1297</td>
<td>3525</td>
<td>Glycosyl hydrolase family 20</td>
</tr>
<tr>
<td>Tpen_1097</td>
<td>3444</td>
<td>Phosphotransferase system IIB subunit</td>
</tr>
<tr>
<td>Tpen_1100</td>
<td>3715</td>
<td>Phosphotransferase system IIC subunit</td>
</tr>
<tr>
<td>Tpen_1100</td>
<td>3716</td>
<td>Phosphotransferase system IID subunit</td>
</tr>
<tr>
<td>Tpen_1090</td>
<td>4821</td>
<td>Phosshosugar binding protein, SIS domain</td>
</tr>
<tr>
<td><strong>COGs not found in any other sequenced Crenarchaeota</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tpen_1155</td>
<td>1554</td>
<td>Glycoside hydrolase family 65</td>
</tr>
<tr>
<td>Tpen_1624</td>
<td>3836</td>
<td>2-Dehydro-3-deoxyglucarate aldolase</td>
</tr>
<tr>
<td>Tpen_0948</td>
<td>0207</td>
<td>Thymidylate synthase</td>
</tr>
<tr>
<td>Tpen_0017</td>
<td>3613</td>
<td>Nucleoside 2-deoxyribosyltransferase</td>
</tr>
<tr>
<td>Tpen_1467</td>
<td>5598</td>
<td>Trimethylamine:corrinoid methyltransferase</td>
</tr>
<tr>
<td>Tpen_1211</td>
<td></td>
<td>Monomethylamine:corrinoid methyltransferase</td>
</tr>
<tr>
<td>Tpen_1092</td>
<td>1080</td>
<td>Phosphoenolpyruvate-protein kinase (Enzyme I of PTS)</td>
</tr>
<tr>
<td>Tpen_1091</td>
<td>1925</td>
<td>Phosphotransferase system HPr protein</td>
</tr>
<tr>
<td>Tpen_1098</td>
<td>2893</td>
<td>Phosphotransferase system IIA component</td>
</tr>
<tr>
<td>Tpen_1491</td>
<td>1268</td>
<td>Biotin transporter bioY</td>
</tr>
<tr>
<td>Tpen_0929</td>
<td>3601</td>
<td>Riboflavin transporter</td>
</tr>
<tr>
<td>Tpen_0191</td>
<td>2116</td>
<td>Formate transporter</td>
</tr>
<tr>
<td>Tpen_1479</td>
<td>2060</td>
<td>Potassium-transporting ATPase, A chain</td>
</tr>
<tr>
<td>Tpen_1480</td>
<td>2216</td>
<td>Potassium-transporting ATPase, B chain</td>
</tr>
<tr>
<td>Tpen_1481</td>
<td>2156</td>
<td>Potassium-transporting ATPase, c chain</td>
</tr>
<tr>
<td>Tpen_0197</td>
<td>0474</td>
<td>Cation transport ATPase (P-type ATPase)</td>
</tr>
<tr>
<td>Tpen_1427</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tpen_1010</td>
<td>1327</td>
<td>Predicted transcriptional regulator, Zn ribbon and ATP-cone</td>
</tr>
<tr>
<td>Tpen_1048</td>
<td>1510</td>
<td>Predicted transcriptional regulator</td>
</tr>
<tr>
<td>Tpen_0270</td>
<td>4190</td>
<td>Predicted transcriptional regulator</td>
</tr>
<tr>
<td>Tpen_0889</td>
<td>2150</td>
<td>Predicted regulator of amino acid metabolism, ACT domain</td>
</tr>
<tr>
<td>Tpen_0253</td>
<td>2229</td>
<td>Predicted GTPase</td>
</tr>
<tr>
<td>Tpen_1457</td>
<td>1773</td>
<td>Rubredoxin</td>
</tr>
<tr>
<td>Tpen_1536</td>
<td>1811</td>
<td>Uncharacterized membrane protein, DUF554</td>
</tr>
<tr>
<td>Tpen_0198</td>
<td>2047</td>
<td>Uncharacterized protein, ATP-grasp superfamily</td>
</tr>
<tr>
<td>Tpen_0838</td>
<td>2164</td>
<td>Uncharacterized conserved protein, DUF369</td>
</tr>
<tr>
<td>Tpen_1835</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tpen_1118</td>
<td>2908</td>
<td>Uncharacterized conserved protein</td>
</tr>
<tr>
<td>Tpen_0381</td>
<td>3863</td>
<td>Uncharacterized relative of cell wall-associated hydrolases</td>
</tr>
<tr>
<td>Tpen_1090</td>
<td>4821</td>
<td>Uncharacterized protein with phosshosugar binding domain</td>
</tr>
</tbody>
</table>
Table 3. Genes present in all Crenarchaeota except *T. pendens* with COG hits

<table>
<thead>
<tr>
<th>COG</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>0214</td>
<td>Pyridoxine biosynthesis enzyme (yaaD)</td>
</tr>
<tr>
<td>0311</td>
<td>Glutamine amidotransferase involved in pyridoxine synthesis (yaaE)</td>
</tr>
<tr>
<td>0413</td>
<td>Ketopantoate hydroxymethyltransferase</td>
</tr>
<tr>
<td>0452</td>
<td>Phosphopantothenoylcysteine synthetase/decarboxylase</td>
</tr>
<tr>
<td>0108</td>
<td>3,4-dihydroxy-2-butanone 4-phosphate synthase</td>
</tr>
<tr>
<td>1985</td>
<td>Pyrimidine reductase, riboflavin biosynthesis</td>
</tr>
<tr>
<td>0054</td>
<td>Riboflavin synthase beta chain</td>
</tr>
<tr>
<td>0163</td>
<td>3-polyprenyl-4-hydroxybenzoate decarboxylase</td>
</tr>
<tr>
<td>1635</td>
<td>Flavoprotein involved in thiazole biosynthesis</td>
</tr>
<tr>
<td>0112</td>
<td>Glycine/serine hydroxymethyltransferase</td>
</tr>
<tr>
<td>0189</td>
<td>Glutathione synthase/Ribosomal protein S6P modification enzyme/L-2-aminoacidipate N-acetyltransferase</td>
</tr>
<tr>
<td>0105</td>
<td>Nucleoside diphosphate kinase</td>
</tr>
<tr>
<td>2046</td>
<td>Sulfate adenylyltransferase</td>
</tr>
<tr>
<td>1650</td>
<td>Uncharacterized protein conserved in archaea</td>
</tr>
<tr>
<td>1701</td>
<td>Uncharacterized protein conserved in archaea</td>
</tr>
</tbody>
</table>
Figure 1. Alignment of presenilins from human and Thermoproteales. Alignment was carried out with Clustal W (49) and shaded with GeneDoc (http://www.nrbsc.org/gfx/genedoc/index.html).