Extremophiles (1997) 1:125–134

ORIGINAL PAPER

In-Geol Choi · Sang Suk Kim · Jae-Ryeon Ryu
Ye Sun Han · Won-Gi Bang · Sung-Hou Kim
Yeon Gyu Yu

Random sequence analysis of genomic DNA of a hyperthermophile: *Aquifex pyrophilus*

Received: March 1, 1997 / Accepted: May 9, 1997

Abstract *Aquifex pyrophilus* is one of the hyperthermophilic bacteria that can grow at temperatures up to 95°C. To obtain information about its genomic structure, random sequencing was performed on plasmid libraries containing 0.5–2 kb genomic DNA fragments of *A. pyrophilus*. Comparison of the obtained sequence tags with known proteins revealed that 123 tags showed strong similarity to previously identified proteins in the PIR or Genebank databases. These included three proteases, two amino acid racemases, and three enzymes utilizing oxygen as substrate. Although the GC ratio of the genome is about 40%, the codon usage of *A. pyrophilus* showed biased occurrence of G and C at the third position of codons, especially those for amino acids such as asparagine, aspartic acid, cysteine, glutamine, glutamic acid, histidine, lysine, and tyrosine. A higher ratio of positively charged amino acids in *A. pyrophilus* proteins as compared with proteins from mesophiles suggested that *Aquifex* proteins might contain increased ion-pair interaction that could help to maintain heat stability.

Key words *Aquifex pyrophilus* · Hyperthermophile · Genome sequence · Protease · Racemase · Oxidase

Communicated by G. Antranikian

I.-G. Choi · S.S. Kim · J.-R. Ryu · Y.S. Han · Y.G. Yu (✉)
Structural Biology Center, Korea Institute of Science and Technology, PO Box 131, Cheongryang, Seoul, Korea
Tel. +82-2-958-5936; Fax +82-2-958-5939
e-mail: ygy@kistmail.kist.re.kr

W.-G. Bang
Department of Agricultural Chemistry, Korea University, Seoul, Korea

S.-H. Kim
Department of Chemistry and Lawrence Berkeley National Laboratory, University of California, Berkeley, California, USA

Introduction

Organisms which can grow at near or above the boiling point of water have been identified from various geothermal locations and named hyperthermophiles. Most hyperthermophiles belong to the Archaea except two genera of Bacteria, *Aquifex* and *Thermotoga*. *Aquifex pyrophilus*, which was isolated from a marine thermal deposit near a site of volcanic activity (Huber et al. 1992), has the highest known optimum growth temperature (85°C) among Bacteria. It is a Gram-negative bacterium and a strict chemolithoautotroph that uses CO_2 as a carbon source and H_2 as an electron donor. One of the unique properties of this organism is its ability to utilize oxygen as an electron acceptor. It can also use nitrate as an electron acceptor under anaerobic growth conditions. Comparison of the 16S rRNA sequence from *A. pyrophilus* to those from other bacteria located this organism at the deepest branch-off in the bacterial domain (Burggraf et al. 1992). Hyperthermophily, chemolithoautotrophy, and basal location in the phylogenetic tree suggested that this organism might possess characteristics of an ancestor of Bacteria.

The genome size of *A. pyrophilus* has been determined to be about 1.6×10^6 base pairs and its physical map has been constructed using rarely cutting restriction enzymes (Shao et al. 1994). A few genes from *A. pyrophilus* had been cloned and characterized. Among them are tRNA synthetase (Brown and Doolittle 1995), elongation factor G (Bocchetta et al. 1995), flagellin (Behammer et al. 1995), and the *recA* gene (Wetmur et al. 1994). Amino acid sequences of proteins from *A. pyrophilus* were about 38%–60% identical to those from mesophilic bacteria such as *Bacillus, Escherichia coli*, or *Salmonella*. The extent of these identities was higher than the value between hyperthermophilic Archaea and mesophilic Bacteria.

Sequence analysis of randomly selected cDNA or fragments of genomic DNA has been proved to be quite an effective method for obtaining a large amount of genomic information. In the case of *Thermatoga maritima*, among 175 randomly selected cDNA clones derived from polyA-

containing mRNA, 52 unique clones showed significant amino acid sequence similarity with known proteins (Kim et al. 1993). Sequencing of a large number of randomly prepared short genomic DNA fragments and assembly of the sequences can produce a contiguous sequence of the whole genome. This approach was successfully applied to sequencing the small genomes from organisms such as *Haemophilus influenzae* (Fleischmann et al. 1995) or *Methanococcus jannaschii* (Bult et al. 1996). In these organisms, about 1700 predicted protein-coding regions were identified from the whole genome of 1.8-Mb size. Genes from these organisms were very compactly organized. As a result, the coding region covered more than 85% of the total genome and the distances between two adjacent genes were small. Since *A. pyrophilus* has a small genome, the noncoding regions will be very small compared to the coding regions and the probability that the randomly sequenced genomic DNA is a coding region of a gene would be high.

Here, we report the identification of more than one hundred genes of *A. pyrophilus* using random sequencing of its genomic DNA library. The efficiency of the random sequence tag method, the presence of interesting genes in terms of biotechnology, the codon usage, and the relative amino acid frequency are also discussed.

Materials and methods

Materials. *A. pyrophilus* cells (stock number 6858) were obtained from Deutsch Sammlung von Mikroorganismen (DSM, Braunschweig Germany). Plasmid pBluescript II KS(+) (Strategene, La Jolla, CA, USA) and pUC19 were used for the construction of a plasmid library. The host strain of *E. coli* for plasmid DNA amplification was DH5α [*supE44 ΔlacU169(φ80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1*]. Restriction enzymes and DNA modification enzymes used for this study were of molecular biology grade.

Cell culture and DNA isolation. A modified SME media (Stetter et al. 1983) was used for the growth of *A. pyrophilus*. The cells were inoculated into 20 ml of liquid medium in a 120-ml incubation bottle which was filled to 3 bar with a gas mixture composed of hydrogen, carbon dioxide, and oxygen at a ratio of 79.5:20:0.5 (%). Cells were grown for 12 h at 85°C with moderate shaking (120 rpm) as described previously (Huber et al. 1992). Cells were harvested by centrifugation at 10000 × *g* for 10 min. The cell pellet was washed with 10 mM Tris-HCl (pH 8.0) and stored at −80°C until further usage. For isolation of genomic DNA, a genomic DNA purification kit from Qiagen (Hilden, Germany) was used. The cell pellet of *A. pyrophilus* (100 mg wet weight) was resuspended in 1 ml of 50 mM Tris-HCl (pH 8.0), 50 mM ethylenediaminetetraacetic acid (EDTA), 0.5% Tween-20, and 0.5% Triton X-100, which also contained ribonuclease and lysozyme at a concentration of 0.2 mg/ml and 2 mg/ml, respectively, and was incubated at 37°C for 1 h. After the incubation, proteinase K was added to the mixture to a concentration of 1 mg/ml, and this was incubated at 50°C overnight. After the treatments, the genomic DNA was purified using a silica base column as described in the instruction manual. Plasmid DNA for sequencing analysis was prepared from *E. coli* using a plasmid DNA mini prep kit (Qiagen).

Library construction. Two sets of plasmid library, using pBluescript II KS(+) and pUC19, were prepared. Plasmid pBluescript II KS(+) and pUC19 DNA were digested with *Hin*dIII and *Eco*RI, respectively, and the linear plasmid DNAs were recovered from an 0.8% agarose gel after electrophoresis. Genomic DNA of *A. pyrophilus* was partially digested with *Hin*dIII or *Tsp*509 I until the average size of DNA fragments was 0.5–1.5 kb. The genomic DNA fragments from *Hin*dIII digestion were ligated with pBluescript II KS(+) linearized with *Hin*dIII. The fragments from *Tsp*509 I digestion were ligated with *Eco*RI-restricted pUC19. The ligated plasmid was transformed into *E. coli* DH5α cells. The transformed *E. coli* cells were plated on an LB agar plate (Sambrook et al. 1989) containing 100 μg/ml of ampicillin along with 800 μg of 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) and 800 μg of isopropylthio-β-D-galactoside (IPTG) on the surface of an agar plate. After an overnight incubation, white colonies were selected and their plasmid DNAs were purified. Standard methods for DNA manipulation and analysis were used as described previously (Sambrook et al. 1989).

DNA sequencing. Nucleotide sequences were determined by the dideoxy chain termination method as described in the instruction manual of the ABI sequencing kit (Applied Biosystems, Foster City, CA, USA). Universal primers (T3 or T7 primer) specific for pBluescript II KS(+) and pUC19 plasmid DNAs were used for the sequencing reaction. The sequencing reaction products were separated on an acrylamide gel and the sequences were read using an ABI 373 automatic DNA sequencer.

Sequence analysis. The obtained genomic DNA sequences were converted into amino acids sequences in all six possible reading frames and were compared with the known proteins in the database using the sequence comparison program, BLAST (Altschul et al. 1990). The SEQSEE program (Wishart et al. 1994) was used for sequence alignment between functionally related proteins.

Results

Construction of *A. pyrophilus* genomic libraries

Plasmid libraries containing small fragments of genomic DNA from *A. pyrophilus* were prepared using plasmid vectors pBluescript II KS(+) or pUC19 DNA. Colonies which may contain the foreign DNA were selected by α-complementation. Plasmid DNAs from 178 colonies of the

Table 1. Statistics on genomic library sequencing

Number of plasmids from white clones	528
Number of plasmids which have been sequenced	490
Number of plasmids which gave sequence longer than 350 base pairs	390
Number of clones which have high scoring matches[a]	121
Number of clones which matched with known genes	109
Number of clones which matched with hypothetical protein	12
Matched clones as % of sequenced clones	30%
Sequenced bases estimated	150 kb
Sequenced DNA as % of total genome size	9.4%

[a] Sequence tags with matched score higher than 100 were selected.

pBluescript II KS(+)-based library and 350 colonies of the pUC19 library were purified. Among the 528 selected clones, 38 white colonies contained plasmid DNAs without any insert and 490 colonies had plasmid DNAs harboring *Aquifex* DNA of 0.5−2.0 kb (Table 1). The 38 false positives from the α-complementation test probably resulted from the self-ligation of vector DNA which had damaged cloning sites.

Sequence determination and analysis

The nucleotide sequence of *A. pyrophilus* genomic DNA in the 490 plasmid DNAs was determined using primers specific for the sequences adjacent to the cloning site of the pBluescript II KS(+) or pUC19 DNA. Among those 490 plasmid DNAs, only 390 were successfully sequenced to produce readable sequences of 350–400 bases with the automatic DNA sequencer. The sequencing failure of the remaining 100 plasmid DNAs was probably due to inadequate purity of the DNA or failure in the reaction procedure. Sequencing reactions were not repeated for these plasmid DNAs. The sequence tags obtained from 390 plasmid DNAs were processed for BLAST analysis. Amino acid sequences were deduced in six possible reading frames from each of the obtained sequence tags. Comparison of the sequence tags with the known protein sequences in the Protein Identification Resource (PIR) or Genebank by the use of the BLAST program identified genes having similarity to each tag. A matched score of over 100 from a BLAST search was used as a criterion for deciding whether the unknown sequences were matched to known proteins. There were 121 genes which satisfied this criterion. Among them, 109 tags represented similarity to functionally defined genes and 12 tags were matched to hypothetical proteins whose functions are not yet defined. The tags are categorized based on their function and are listed in Table 2. In addition, two tags similar to proteases with scores less than 100 are also listed.

Interestingly, the sequence tag AQpUs20 had the same DNA sequence as the previously identified elongation factor G gene from *A. pyrophilus* (Bocchetta et al. 1995). The sequence tag represents nucleotides 1857–2227 of the elongation factor G gene. When the two sequences were compared, there were 7 mismatched bases out of 371 bases. The accuracy of sequencing was measured to be more than 98% for this tag.

Genes from *A. pyrophilus*

The 121 sequence tags are listed with genes having strong similarity in Table 2. The assigned functions for each sequence tag included various cellular functions such as amino acid biosynthesis, energy metabolism, macromolecule metabolism, or biosynthesis of a cell wall component.

Several components of protein transport machinery were identified. Genes with similarity to those for SecA (AQpBbH62), SecF (AQpBH67), and signal recognition particle (AQpBH115) were identified. The presence of the *secA* gene along with other components suggested that this organism possesses a protein transport system similar to that commonly found in Bacteria.

There were several genes whose presence had been anticipated. One of them was isocitrate dehydrogenase (AQpUs43). Activity of isocitrate dehydrogenase had been measured in *Aquifex* (Beh et al. 1993). This enzyme is assumed to make isocitrate from α-ketoglutarate and carbon dioxide as a component of the reductive citric acid cycle by which carbon dioxide fixation is conducted. The kinetic properties of this enzyme could reveal more details of the reductive citric acid cycle. Under anaerobic growth conditions, this organism uses nitrate as an electron acceptor. AQpBH32 represents nitrate reductase, which may be involved in nitrate reduction required for the growth under anaerobic conditions or for nitrogen assimilation.

Three sequence tags similar to proteases, together with their sequence identities, were compared with known proteases (Fig. 1). AQpBbH130 had a strong similarity to ATP-dependent endopeptidase La of *E. coli*, which is involved in degradation of damaged cytosolic proteins (Chin et al. 1988). In addition, two tags (AQpBB007 and AQpUs41) had similar sequences to serine-type proteases although their matched score values were lower than 100. AQpBB007 has a coding region of 18 amino acids due to the short insert size. The amino acid sequence was 61% identical to aqualysin, an extracellular serine protease from *Thermus aquaticus* (Kwon et al. 1988). Also, the identical region of the tag contained a histidine residue that composed an active site. Since serine-type proteases have high similarity around the active site, which contains serine, histidine, and aspartic acid, AQpBB007 probably represents one of the protease genes in *Aquifex*.

Several enzymes which use oxygen as a substrate were identified, and their sequence similarity to known proteins is shown in Fig. 2. Two subunits of the respiration complex were identified. AQpBB012 and AQpU238 were similar to cytochrome *c* oxidase chain I and cytochrome *d* complex chain II, respectively. The cytochrome *c* oxidase used oxygen as the terminal electron acceptor and cytochrome *c* as the electron donor. Besides respiratory proteins, aspartate oxidase (AQpU75) and superoxide dismutase (AQpBH15) were observed. Aspartate oxidase converts aspartate to

Table 2. List of genes obtained from *A. pyrophilus* genomic library having matches to known genes in other organisms

pir no.	Gene description/categories	Score	Length of ORF	Clone
	Biosynthesis of small molecules			
pir|S| S08431	2-isopropylmalate synthase (EC 4.1.3.12) – *Salmonella typhimurium*	237	97	AQpU312
pir|S|S23070	6-deoxyerythronolide B synthase II – *Saccharopolyspora erythraea*	103	40	AQpU178
pir|S|XQBS	amidophosphoribosyltransferase (EC 2.4.2.14) – *B. subtilis*	227	88	AQpBH116
pir|S|A38621	aspartate transaminase (EC 2.6.1.1) – *Bacillus* sp. (strain YM-2)	272	97	AQpBH60
pir|S|JS0274	biotin synthetase – *Bacillus sphaericus*	131	117	AQpBB015
pir|S|F39845	carbamoyl-phosphate synthase (EC 6.3.5.5) – *B. subtilis*	331	85	AQpU74
pir|S|SYECKR	chorismate synthase (EC 4.6.1.4) – *E. coli*	119	74	AQpU173
pir|S|16802	cyclase *hisF* – *Azospirillum brasilense*	246	126	AQpU163
pir|S|A31133	diaminopimelate decarboxylase (EC 4.1.1.20) – *Pseudomonas aeruginosa*	101	96	AQpU356
pir|S|DEYCG6	glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) – *Synechococcus* sp. (PCC 7942)	193	86	AQpU195
pir|S|A49473	glutamate racemase (EC 5.1.1.3) – *Lactobacillus fermenti*	223	108	AQpU181
pir|S|S46975	glutamate synthase (ferredoxin) (EC 1.4.7.1) – *Synechocystis* sp.	151	75	AQpU167
pir|S|XNECGM	glutamine–fructose-6-phosphate transaminase (EC 2.6.1.16) – *E. coli*	211	107	AQpu217
pir|S|SYECGU	GMP synthase (glutamine-hydrolyzing) (EC 6.3.5.2) – *E. coli*	221	57	AQpU346
pir|S|B42280	GMP synthetase – *Bacillus subtilis*	257	81	AQpU301
pir|S|DEPSHA	homoserine dehydrogenase (EC 1.1.1.3) – *Pseudomonas aeruginosa*	185	84	AQpU252
pir|S|B42594	*hyuA* – *Pseudomonas* sp.	136	82	AQpU197
pir|S|C42594	*hyuB* – *Pseudomonas* sp.	149	55	AQpU240
pir|S|DEBSMP	IMP dehydrogenase (EC 1.1.1.205) – *Bacillus subtilis*	126	37	AQpBH81
pir|S|A40615	*o*-succinylbenzoate synthase – *E. coli*	329	68	AQpU330
pir|S|DCUTOB	ornithine decarboxylase (EC 4.1.1.17) – *Trypanosoma brucei*	117	49	AQpU151
pir|S|A28088	oxaloacetate decarboxylase (EC 4.1.1.3) – *Klebsiella pneumoniae*	248	106	AQpU159
pir|S|S49373	oxygen independent coproporphyrinogen III oxidase –*E. coli*	164	72	AQpBH164
pir|S|BVECXA	PdxA protein – *E. coli*	135	81	AQpU307
pir|S|1S21418	phospho-2-dehydro-3-deoxyheptonate aldolase (EC 4.1.2.15) – *B. subtilis*	109	39	AQpU260
pir|S|DEBSPE	phosphoribosylaminoimidazole carboxylase (EC 4.1.1.21) – *B. subtilis*	222	108	AQpU310
pir|S|SYBS1G	phosphoribosylformylglycinamidine synthase (EC 6.3.5.3) – *B. subtilis*	165	114	AQpU174
pir|S|S45545	RibA protein – *Bacillus subtilis*	105	33	AQpBH59
pir|S|PN0100	RibG protein – *Bacillus subtilis*	263	78	AQpUs10
pir|S|A42479	*S*-adenosyl-L-methionine uroporphyrinogen III methyltransferase – *Bacillus megaterium*	216	75	AQpU155
pir|S|B47045	uroporphyrinogen decarboxylase (EC 4.1.1.37) – *Bacillus subtilis*	108	58	AQpU234
	Cell structure			
pir|S|I42365	flagellar hook-basal body protein FlgG – *Bacillus subtilis*	109	37	AQpBB013
pir|S|D36914	integral membrane protein ModB – *Rhodobacter capsulatus*	198	88	AQpU90
pir|S|LPECRA	lipoprotein RlpA precursor – *E. coli*	134	76	AQpUs1
pir|S|ZPECP2	penicillin-binding protein 2 – *E. coli*	132	73	AQpU144
pir|S|JC1275	phospho-*N*-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) – *Bacillus subtilis*	116	100	AQpBbH77
pir|S|A47681	short-chain alcohol dehydrogenase homolog EnvM – *E. coli*	208	112	AQpU168
pir|S|JS0701	SmpB protein – *E. coli*	179	63	AQpBH80
pir|S|A44917	UDP-*N*-acetylglucosamine enolpyruvy transferase MurZ – *E. coli*	128	98	AQpU239
	Cellular processes			
pir|S|A33830	cation efflux system membrane protein CzcA – *Alcaligenes eutrophus*	166	94	AQpU308
pir|S|CEECFY	cell division protein FtsY – *E. coli*	303	136	AQpBH115
pir|S|JC1345	CmcT protein – *Streptomyces lactamdurans* (fragment)	111	81	AQpBH24
pir|S|A39194	dipeptide transport protein DppA – *E. coli*	124	85	AQpBH111
pir|S|DEPSGD	GDP mannose 6-dehydrogenase (EC 1.1.1.132) – *Pseudomonas aeruginosa*	173	86	AQpBH1
pir|S|A25504	H^+-transporting ATP synthase (EC 3.6.1.34) – *E. coli*	216	63	AQpBB011
pir|S|C43255	*hoxZ* – *Alcaligenes eutrophus*	142	57	AQpBH31
pir|S|BVECLA	LepA protein – *E. coli*	336	96	AQpU213
pir|S|JN0255	lipoic acid metabolism protein LipA – *E. coli*	358	109	AQpBH68
pir|S|BVECHJ	molybdenum transport protein ChlJ – *E. coli*	140	88	AQpBH13
pir|S|JC1345	multidrug resistant protein EmrB – *E. coli*	147	104	AQpBH25
pir|S|S47149	SecA protein – *Staphylococcus carnosus*	103	77	AQpBbH62
pir|S|JQ0697	secretion protein SecF – *E. coli*	237	125	AQpBH67
pir|S|B44465	sodium ion pump oxaloacetate decarboxylase – *Salmonella typhimurium*	376	118	AQpU170
pir|S|S09250	uptake hydrogenase (EC 1.18.99.-) – *Rhodocychus gelatinosus*	194	86	AQpU292
	Intermediary metabolism			
pir|S|A35742	aqualysin (EC 3.4.21.-) I precursor – *Thermus aquaticus*	62	18	AQpBB007

Table 2. *Continued*

pir no.	Gene description/categories	Score	Length of	Clone
pir|S|S31885	ATP synthase (c) – *Synechococcus* sp.	168	66	AQpUs45
pir|S|S27055	basic proteinase precursor – *Dichelobacter nodosus*	73	84	AQpUs41
pir|S|S20534	cytochrome-*c* oxidase (EC 1.9.3.1) chain I – *Rhodobacter sphaeroides*	316	118	AQpBB012
pir|S|B28940	cytochrome *d* complex (EC 1.10.3.-) chain II – *E. coli*	207	89	AQpU238
pir|S|SUECLA	endopeptidase La (EC 3.4.21.53) – *E. coli*	187	95	AQpBbH130
pir|S|S19606	HydG protein – *Salmonella typhimurium*	174	57	AQpBH42
pir|S|S25692	HypF protein – *Rhodobacter capsulatus*	230	143	AQpBH110
pir|S|B49341	isocitrate dehydrogenase ($NADP^+$B14+) (EC 1.1.1.42) I – *Vibrio* sp. (strain ABE-1)	322	114	AQpUs43
pir|S|OXECLD	L-aspartate oxidase (EC 1.4.3.16) – *E. coli*	236	66	AQpU75
pir|S|S36863	D-lactate dehydrogenase (EC 1.1.1.27) – *Thermotoga maritima*	110	84	AQpU176
pir|S|JS0493	methionyl aminopeptidase (EC 3.4.11.18) – *E. coli*	165	110	AQpBH5
pir|S|A39588	NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) 50K chain – *Paracoccus denitrificans*	162	94	AQpUs6
pir|S|A47751	NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 – *Paracoccus denitrificans*	177	63	AQpU322
pir|S|S22370	NADH dehydrogenase 23K chain homolog – *Rhodobacter capsulatus*	127	89	AQpU200
pir|S|S37065	NADH dehydrogenase I chain NuoH – *E. coli*	170	90	AQpU251
pir|S|S49971	*nifR3*-like gene product – *Azospirillum brasilense*	128	119	AQpU192
pir|S|S29756	NifU protein – *Azotobacter vinelandii*	286	103	AQpU171
pir|S|S36605	nitrate reductase – *Synechococcus* sp.	171	112	AQpBH32
pir|S|S14395	probable cytochrome *a* assembly factor – *Bacillus subtilis*	213	92	AQpU258
pir|S|JT0393	regulatory protein GlnR – *Bacillus subtilis*	105	66	AQpU93
pir|S|NIPSRP	regulatory protein XylR – *Pseudomonas putida* plasmid TOL	126	87	AQpU65
pir|S|A38461	superoxide dismutase (EC 1.15.1.1) (Mn-Fe) – *Methylomonas* sp.	142	101	AQpBH15
pir|S|DWECTD	threonine dehydratase (EC 4.2.1.16) – *E. coli*	228	80	AQpU028
pir|S|A41229	transcription initiation factor sigma 54 – *Bacillus subtilis*	142	92	AQpU202
pir|S|RNEBST	transcription initiation factor sigma 70 – *Salmonella typhimurium*	142	93	AQpUs53
pir|S|JQ0338	transcription initiation factor sigma NtrA – *Pseudomonas putida*	121	77	AQpU78
pir|S|C55216	transcription initiation factor sigmaD – *Bacillus subtilis*	139	67	AQpU199
pir|S|S46946	Zam protein - *Cyanobacterium synechocystis*	126	79	AQpUt5
	Macromolecule metabolism			
pir|S|A33168	26K antigen – *Helicobacter pylori*	139	28	AQpUts7
pir|S|SYECAT	alanine–rRNA ligase (EC 6.1.1.7) – *E. coli*	186	80	AQpU351
pir|S|JC1118	alkyl sulfatase (EC 3.1.6.-) – *Pseudomonas* sp.	113	62	AQpU207
pir|S|S16753	aminoacyl-tRNA hydrolase (EC 3.1.1.29) – *E. coli*	111	63	AQpBH30
pir|S|A44514	AnfA protein – *Azotobacter vinelandii*	348	124	AQpU161
pir|S|A32412	deoxyribonuclease (pyrimidine dimer)(EC 3.1.25.1) – *E. coli*	191	81	AQpU216
pir|S|A40363	DNA ligase (NAD^+) (EC 6.5.1.2) – *Thermus aquaticus*	205	66	AQpU339
pir|S|A49794	DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) *Neisseria gonorrhoeae* (strain MUG116)	242	90	AQpU334
pir|S|IQBSOC	DnaA protein homolog – *Bacillus subtilis*	227	88	AQpU344
pir|S|IQECDB	DnaB protein – *E. coli*	155	61	AQpBH135
pir|S|S38928	elongation factor G – *Aquifex pyrophilus* (fragment)	339	109	AQpUs20
pir|S|SYECET	glutamate–tRNA ligase (EC 6.1.1.17) – *E. coli*	157	39	AQpU311
pir|S|S40178	isoleucine–tRNA ligase (EC 6.1.1.5) – *Staphylococcus aureus*	162	83	AQpU325
pir|S|SYTWMT	methionine–tRNA ligase (EC 6.1.1.10) – *Thermus aquaticus*	325	106	AQpBH73
pir|S|YFBSB	phenylalanine–tRNA ligase (EC 6.1.1.20) – *Bacillus subtilis*	103	33	AQpBH72
pir|S|S35953	prepilin peptidase PilD – *Pseudomonas putida*	202	104	AQpBH17
pir|S|S29885	ribosomal protein S8 – *Micrococcus luteus*	137	55	AQpU262
pir|S|FCECR1	translation releasing factor RF-1 – *E. coli*	255	111	AQpU157
pir|S|S02166	type I site-specific deoxyribonuclease (EC 3.1.21.3) *EcoR* 124/3 HsdM protein – *E. coli* plasmid R 124/3	103	58	AQpUts11
pir|S|A42385	Uvr-402 protein – *Streptococcus pneumoniae* plasmid pSB470	215	117	AQpU81
pir|S|A41251	valine–tRNA ligase (EC 6.1.1.9) – *Neurospora crassa*	228	110	AQpU160
	Other			
pir|S|C41902	arsenate reductase (EC 1.-.-.-) – *Staphylococcus xylosus* plasmid pSX267	114	104	AQpUts6
pir|S|A54058	farnesyltranstransferase (EC 2.5.1.29) – *Sulfolobus acidocaldarius*	129	76	AQpU245
pir|S|JT0761	heat shock protein HslU – *E. coli*	267	97	AQpUs39
pir|S|A43653	HflX protein – *E. coli*	245	122	AQpU208
pir|S|S21957	P-glycoprotein Pgp1 – *Arabidopsis thaliana*	124	97	AQpU341
pir|S|S45881	probable purine nucleotide-binding protein YBR025c – yeast (*Saccharomyces cerevisiae*)	211	79	AQpU253
pir|S|KIBSRS	ribose-phosphate pyrophosphokinase (EC 2.7.6.1) – *Bacillus subtilis*	137	45	AQpBH108
	Hypothetical protein			
pir|S|Q4ECGG	hypothetical 10K protein (*glpD-glgP* intergenic region) – *E. coli*	466	93	AQpUts15

Table 2. *Continued*

pir no.	Gene description/categories	Score	Length of	Clone
pir|S|B33171	hypothetical 17K protein (*lpxA* 5′ region) – *E. coli*	182	59	AQpU299
pir|S|JQ0872	hypothetical 56.2K protein (*ilvG-rrnC* intergenic region) – *E. coli*	225	81	AQpU255
pir|S|JQ1221	hypothetical 60K protein – *Pseudomonas putida*	123	65	AQpBH33
pir|S|S43189	hypothetical protein – *Pseudomonas aeruginosa*	122	41	AQpUts3
pir|S|S48428	hypothetical protein – yeast (*Saccharomyces cerevisiae*)	157	77	AQpU228
pir|S|S27544	hypothetical protein 1 – *Thermoanaerobacterium thermosulfurigenes*	186	91	AQpBH44
pir|S|S26447	hypothetical protein 10 – *Methanobacterium thermoformicicum* plasmid pFV1	137	91	AQpU88
pir|S|S43916	hypothetical protein 3 – *Bacillus stearothermophilus*	120	66	AQpU165
pir|S|S18078	hypothetical protein 3 – *Bacillus subtilis*	102	57	AQpU211
pir|S|PN0543	hypothetical protein 313 – *Micromonospora olivasteropora* (fragment)	151	111	AQpUs51
pir|S|C41872	hypothetical protein (*groEL* 3′-region) – *Clostridium acetobutylicum*	112	47	AQpU337

ORF, open reading frame.

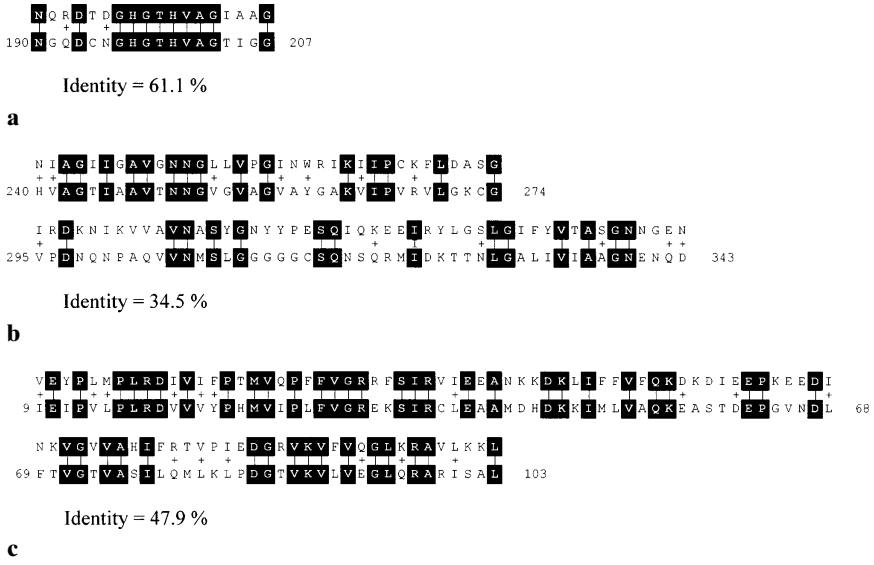

Fig. 1a–c. Sequence alignment of proteases. The amino acid sequences of the *upper* strands are from *Aquifex* proteins and the *lower* strands are from the homologous proteins. **a** AQpBB007 and aqualysin from *Thermus aquaticus* (PIR# A35742). **b** AQpUs41 and basic protease from *Dichelobacter nodosus* (PIR# S27055). **c** AQpBbH130 and endopeptidase La of *Escherichia coli* (PIR# SUECLA). The identity values for the pairs of sequences are listed *below* each pair

iminoaspartate using oxygen as an electron acceptor. Superoxide dismutase, together with catalase whose activity was previously measured in *A. pyrophilus* (Huber et al. 1992), was probably responsible for the removal of oxygen radical. Since most hyperthermophiles have been grown anaerobically, it has been assumed that they would not have superoxide dismutase or amino acid oxidases. Oxygen-utilizing enzymes identified from *Aquifex* in this study might represent one of the extremely heat-stable enzymes in that class.

Two amino acid racemases which convert an L-form to a D-form amino acid were identified. One of them is glutamate racemase (AQpU181). Since glutamic acid was detected in *Aquifex* cell wall peptidoglycan (Huber et al. 1992), this enzyme is assumed to be responsible for the synthesis of D-glutamate, which is used as a building block of peptidoglycan. Alanine racemase (AQpU219) was also observed in this study, although it was not listed in Table 2 due to the matched score being lower than 100. The comparison of the two tags with other bacterial racemases is shown in Fig. 3. The amino acids racemases of *Aquifex* might be among the most heat-stable amino acid racemases.

Codon usage

About 30 sequence tags that had a matched score greater than 200 with known proteins were selected. A total of 2608 codons were obtained from the most probable open reading frames, and codon usage for each amino acid is listed in Table 3, along with the codon bias of *Thermatoga* (Kim et al. 1993) and *E. coli* (obtained from the IUBIO archive at Indiana University by way of Internet: ftp.iubio.bio.edu). *Aquifex* showed codon usage similar to that of *Thermotoga*. Amino acids such as Ala, Gly, Ser, Pro, Thr, and Val used almost evenly distributed codons. For arginine, AGA and AGG codons were dominantly used in hyperthermophilic Bacteria. In *E. coli*, however, these codons only represent 7% of the arginine codons. Another

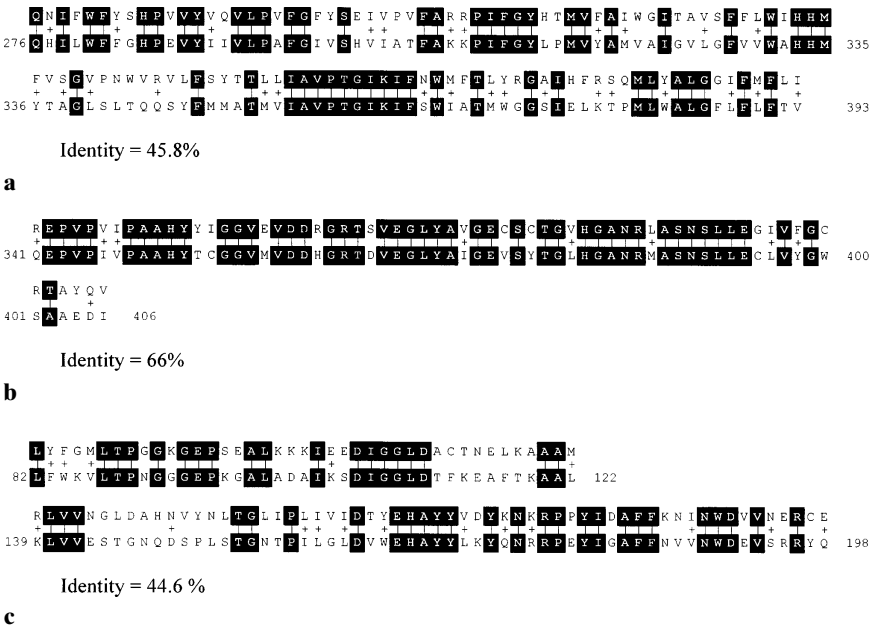

Fig. 2a–c. Sequence alignment of oxidases and superoxide dismutase. The *upper* strands represent *Aquifex* sequences. **a** AQpBB012 and cytochrome *c* oxidase subunit I from *Rhodobacter sphaeroides* (PIR# S20534). **b** AQpU75 and aspartate oxidase from *E. coli* (PIR# OXECLD). **c** AQpBH15 and superoxide dismutase from *Methylomonas* (PIR# A38461)

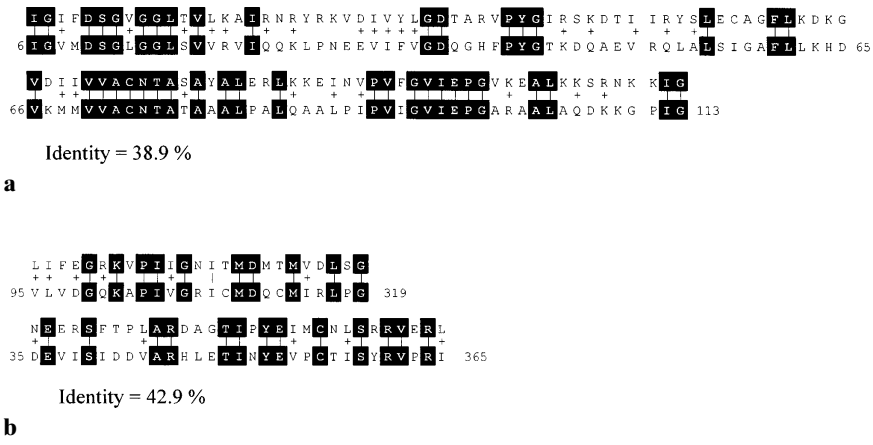

Fig. 3a,b. Sequence alignment of racemases. The *upper* strands represent *Aquifex* sequences. **a** AQpU181 and glutamate racemase from *Lactobacillus fermenti* (PIR# A49473). **b** AQpU219 and alanine racemase from *Bacillus stearothermophilus* (PIR# A24102)

noticeable trend in *Aquifex* was a higher G + C content of 56% at the third position of codons, compared with the average of 40% G + C content in the genome. In the case of amino acids using only two codons, such as Asn, Asp, Cys, Gln, Glu, His, Lys, Phe, and Tyr, the average frequency of G or C at the third position of codons was even higher, at 65%.

High frequency of positively charged amino acids

The frequencies of charged and hydrophobic amino acids in *Aquifex* were compared with those from mesophilic bacteria such as *E. coli* or *B. subtilis*. Among the sequence tags, 32 tags with a high matched score and open reading frames longer than 50 amino acids were selected. The amino acids from the sequence tags and the corresponding region of mesophilic bacterial genes were classified as charged amino acids (arginine, lysine, glutamic acid, and aspartic acid), positively charged amino acids (arginine and lysine), and hydrophobic amino acids (valine, leucine, isoleucine, phenylalanine, and tryptophan). The number of amino acids in each group from the *Aquifex* sequence was compared with the value from the corresponding sequence from mesophilic bacteria, and is listed in Table 4. One of the outstanding trends was the increased occurrence of positively charged amino acids in *Aquifex* proteins. The proportion of positively charged amino acids in the examined region in six *Aquifex* proteins was more than double that in those from mesophilic bacteria. Only 12.5% of the tags had a lower amount of positively charged amino acids. In contrast, the ratio of hydrophobic amino acids was almost unchanged. These results indicated that at least some portions of *A. pyrophilus* proteins have more positively charged amino acids than their mesophilic counterpart. The increased positively charged amino acid content may result in the increase of ion-pair interactions which in turn contribute to thermostability of proteins.

Table 3. Codon usage in *Aquifex pyrophilus*

Residue	Codon	*Aquifex pyrophilus*		*Thermotoga maritima* %[a,b]	*E. coli* %[a,c]	Residue	Codon	*Aquifex pyrophilus*		*Thermotoga maritima* %[a,b]	*E. coli* %[a,c]
		No. of codons	%[a]					No. of codons	%[a]		
Ala	GCA	48	24	28	22		CTG	49	17	21	55
	GCC	43	22	22	25		CTT	82	29	27	10
	GCG	66	34	19	34		TTA	23	8	4	11
	GCT	40	20	31	19		TTG	12	4	10	11
Arg	AGA	45	28	56	4	Lys	AAA	53	33	49	76
	AGG	84	52	22	3		AAG	110	67	51	24
	CGA	7	4	5	5	Met	ATG	53	100	100	100
	CGC	6	4	6	37	Phe	TTC	67	54	82	49
	CGG	11	7	5	8		TTT	58	46	18	51
	CGT	8	5	8	42	Pro	CCA	12	11	16	20
Asn	AAC	70	79	79	61		CCC	57	50	24	10
	AAT	19	21	21	39		CCG	23	20	29	55
Asp	GAC	71	66	58	41		CCT	21	19	31	16
	GAT	36	34	42	59	Ser	AGC	18	13	16	27
Cys	TGC	25	69	33	57		AGT	14	10	15	13
	TGT	11	31	67	43		TCA	22	16	13	12
Gln	CAA	21	38	15	31		TCC	42	30	23	17
	CAG	35	63	85	69		TCG	16	11	11	13
Glu	GAA	83	40	65	70		TCT	28	20	21	19
	GAG	126	60	35	30	Thr	ACA	16	12	43	13
Gly	GGA	86	44	54	9		ACC	59	43	25	43
	GGC	31	16	10	40		ACG	49	36	22	23
	GGG	38	19	11	13		ACT	14	10	10	21
	GGT	41	21	25	38	Trp	TGG	31	100	100	100
His	CAC	29	78	74	48	Tyr	TAC	56	70	79	47
	CAT	8	22	26	52		TAT	24	30	21	53
Ile	ATA	125	69	34	7	Val	GTA	43	20	13	17
	ATC	15	8	48	46		GTC	26	12	23	20
	ATT	40	22	18	47		GTG	32	15	27	34
Leu	CTA	20	7	4	3		GTT	109	52	37	29
	CTC	101	35	34	10	Total codons sequenced: 2608					

For *A. pyrophilus*, the overall %GC at the third position of codons was 55.6%.
[a] Percentage of amino acid residue coded for by respective codon.
[b] Kim et al. 1993.
[c] From IUBIO archive at Indiana University.

Discussion

The random sequence tag method was used to obtain genetic information from *A. pyrophilus*. This small-genome organism which lives at extreme temperatures has a limited number of genes and served as a target organism for the examination of gene structure by random sequence analysis. About 10% of the total genomic DNA sequence of *A. pyrophilus* was determined and possible genes were identified after sequence comparison. The number of genes identified in this study was 109. If the total number of genes in *A. pyrophilus* is about 1600 and the fraction of functionally definable genes is about 60%, as in *H. influenzae* (Fleischmann et al. 1995), the number of identified genes in this study corresponds to 11% of all functionally definable genes. Since we sequenced about 10% of the total genome, this is well matched with the expected values. Besides the tags with a matched score greater than 100, listed in Table 2, some of the sequence tags with a score lower than 100 possibly represent a real gene. For example, AQpBB007 has a sequence similar to aqualysin, a serine protease from *Thermus aquaticus*, with a score of 62. When the entire open reading frame containing the sequence tag was cloned and sequenced, it represented one of the serine-protease genes (data not shown).

Aquifex has unique properties, which cannot be observed in other hyperthermophiles, such as oxygen utilization and chemolithotrophic characteristics. Some of the enzymes involved in those processes, aspartate oxidase, superoxide dismutase, glutamate racemase, and components of respiratory complexes, were detected in this study. Interestingly, the presence of subunits of cytochrome *c* oxidase and of the cytochrome *d* complex indicated that *Aquifex* has an oxygen respiratory chain and that cytochrome *c* oxidase

Table 4. Comparison of amino acid composition of proteins from *A. pyrophilus* and mesophilic bacteria

Proteins[a]	Charged amino acids[b]	Positively charged amino acids[c]	Hydrophobic amino acids[d]	Clone name
phosphoribosylformylglycinamidine synthase	1.3	3.3	1.1	AQpU174
homoserine dehydrogenase	1.4	2.5	1.1	AQpU252
glutamate racemase	1.7	2.2	1.0	AQpU181
aspartate transaminase	1.4	2.0	1.5	AQpBH60
L-aspartate oxidase	1.2	2.0	1.1	AQpU75
LepA protein (GTP-binding membrane protein	1.1	2.0	1.1	AQpU213
hyuA(D-5-substituted hydantoin conversion)	1.3	1.8	1.0	AQpU197
valine–tRNA ligase	1.3	1.8	1.4	AQpU160
deoxyribonuclease	1.3	1.6	1.1	AQpU216
LipA protein (lipoic acid metabolism)	1.2	1.4	1.2	AQpBH68
Uvr-402 protein	1.2	1.4	1.0	AQpU81
GMP-synthase	1.1	1.4	0.9	AQpU346
cytochrome *c* oxidase chain I	0.9	1.3	1.1	AQpBB012
DNA topoisomerase	1.1	1.3	1.0	AQpU334
oxaloacetate decarboxylase	1.2	1.3	1.0	AQpU159
phosphoribosylaminoimidazole carboxylase	1.2	1.3	1.0	AQpU310
translation releasing factor RF-1	1.0	1.3	1.4	AQpU157
S-adenosyl-l-methionine uroporphyringogen III methyltransferase	1.2	1.2	1.0	AQpU155
cytochrome *d* complex chain II	0.8	1.2	1.1	AQpU238
nifU(nitrogen fixation related protein)	0.9	1.1	1.2	AQpU171
2-isopropylmalate synthase	1.2	1.1	0.9	AQpU312
cyclase *hisF*	1.1	1.1	0.9	AQpU163
methionine–tRNA ligase	1.2	1.1	1.2	AQpBH73
amido phosphoribosy transferase	1.0	1.0	1.1	AQpBH116
DNA ligase	0.9	1.0	1.1	AQpU339
dnaB (replicative DNA helicase)	0.8	1.0	1.1	AQpBH135
isocitrate dehydrogenase	1.0	1.0	1.0	AQpUs43
o-succinyl benzoate synthase	1.0	1.0	1.0	AQpU330
isoleucine–tRNA ligase	1.1	0.9	1.3	AQpU325
glutamine–fructose-6-phosphate transaminase	1.0	0.9	1.1	AQpU217
carbamoyl-phosphate synthase	1.0	0.9	1.0	AQpU312
threonine dehydratase	0.8	0.7	1.0	AQpU028
Average of all proteins	1.1	1.4	1.1	

[a] Proteins with more than 50 amino acids were selected for the comparison, and corresponding regions were compared.
[b–d] Number of each amino acid in protein from *A. pyrophilus* divided by number of corresponding amino acid in protein from mesophile.
[b] Charged amino acids: Lys, Arg, Asp, Glu.
[c] Positively charged amino acids: Lys, Arg.
[d] Hydrophobic amino acids: Val, Leu, Ile, Phe, Trp.

is used for the last step of respiration. Sequence analysis and biochemical characterization of these enzymes may provide detailed information about the evolution of molecular oxygen utilization.

The high ratio of positively charged amino acids as compared with that of mesophilic proteins may explain the stability of *Aquifex* proteins. Three-dimensional structures of several hyperthermophile proteins have been determined and compared with the structures of mesophilic proteins. One of the remarkable differences is a large number of ion-pairs or network of ion-pairs at the surface of hyperthermophilic proteins. In case of glutamate dehydrogenase, the *Pyrococcus* protein has 70% more ion-pair interactions within a subunit as well as 100% more intersubunit ion-pair interactions than the *Clostridium symbiosum* protein (Yip et al. 1995). Comparison of the glyceraldehyde 3-phosphate dehydrogenase structure from *Thermotoga* and *Bacillus* also indicated an increased number of ion-pair interactions in hyperthermophilic proteins as a major distinguishing feature (Korndorfer et al. 1995). For increased ion-pair interaction, these proteins had more charged amino acids, especially positive charged amino acids. If ion-pair content or a network of ion-pairs is the main contributor of thermostability to hyperthermophilic proteins, *Aquifex* proteins might be expected to have more charged amino acids. As shown in Table 4, the majority of the examined sequences have higher contents of positively charged amino acids than do proteins from mesophilic bacteria. In contrast, there was no significant difference in the distribution of hydrophobic amino acids. More extensive sequence analysis and structure determination will reveal the role of ion-pair interaction in thermostability of proteins from hyperthermophilic bacteria.

Acknowledgments We thank Dr. K.-S. Kim of the Structural Biology Center of the Korea Institute of Science and Technology for helpful advice on the application of the SEQSEE and BLAST programs. This work is supported by a grant from the Ministry of Science and Technology, Korea. The work of S.-H. Kim has been supported by the Office of Health and Environmental Research, U.S. Department of Energy (DE-AC03-765F00098).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Beh M, Strauss G, Huber R, Stetter KO, Fuchs G (1993) Enzymes of the reductive citric acid cycle in the autotrophic eubacterium *Aquifex pyrophilus* and in the archaebacterium *Thermoproteus neutrophilius*. Arch Microbiol 160:306–311

Behammer W, Shao Z, Mages W, Rachel R, Stetter KO, Schmitt R (1995) Flagella structure and hyperthermophily: analysis of a single flagellin gene and its product in *Aquifex pyrophilus*. J Bacteriol 177:6630–6637

Bocchetta M, Ceccarelli E, Sanangelantoni AM, Tiboni O, Cammarano P (1995) Arrangement and nucleotide sequence of the gene (*fus*) encoding elongation factor G (EF-G) from hyperthermophilic bacterium *Aquifex pyrophilus*: phylogenic depth of hyperthermophilic bacteria inferred from analysis of EF-G *fus* sequences. J Mol Evol 41:803–812

Brown JR, Doolittle WF (1995) Root of the universal tree of life based on ancient aminoacyl t-RNA synthetase gene duplications. Proc Natl Acad Sci USA 92:2441–2445

Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb J-F, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NSM, Weidman JF, Fuhrman J, Nguyen D, Utterback TR, Kelley JM, Peterson JD, Sadow PW, Hanna MC, Cotton MD, Robert KM, Hurst MA, Kaine BP, Borodovsky M, Klenk HP, Fraser CM, Smith HO, Woese CR, Venter JC (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science 273:1058-1073

Burggraf S, Olsen GJ, Stetter KO, Woese CR (1992) A phylogenetic analysis of *Aquifex pyrophilus*. Syst Appl Microbiol 15:352–356

Chin DT, Goff SA, Webster T, Smith T, Goldberg AL (1988) Sequence of the *lon* gene in *Escherichia coli*: a heat shock gene which encodes the ATP-dependent protease La. J Biol Chem 263:11718–11728

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, McKenney K, Sutton G, FitzHugh W, Fields C, Gocayne JD, Scott J, Shirley R, Liu L, Glodek A, Kelley JM, Weidman JF, Phillips CA, Spriggs T, Hedblom E, Cotton MD, Utterback TR, Hanna MC, Nguyen DT, Saudek DM, Brandon RC, Fine LD, Fritchman JL, Fuhrman J, Geoghagen NSM, Gnehm CL, McDonald L, Small KV, Fraser CM, Smith HO, Venter CJ (1995) Whole genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496–512

Huber R, Wilharm T, Huber D, Trincone A, Burggraf S, Konig H, Rachel R, Rockinger I, Fricke H, Stetter KO (1992) *Aquifex pyrophilus* gen. nov. sp. nov., represents a novel group of marine hyperthermophilic hydrogen-oxidizing bacteria. Syst Appl Microbiol 15:340–351

Kim CW, Markiewicz P, Lee JJ, Schierle CF, Miller JH (1993) Studies of the hyperthermophile *Thermotoga maritima* by random sequencing of cDNA and genomic libraries: identification and sequencing of the *trpEF(D)* operon. J Mol Biol 231:960–981

Korndorfer I, Steipe B, Huber R, Tomschy A, Jeaniche R (1995) The crystal structure of holo-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima* at 2.5 Å resolution. J Mol Biol 246:511–521

Kwon ST, Terada I, Matsuzawa H, Ohta T (1988) Nucleotide sequence of the gene for aqualysin 1 (a thermophilic alkaline serine protease) of *Thermus aquaticus* YT-1 and characteristics of the deduced primary structure of the enzyme. Eur J Biochem 173:491–497

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning, a laboratory manual. Cold Spring Harbor University, Cold Spring Harbor, NY

Shao Z, Mages W, Schmitt R (1994) A physical map of the hyperthermophilic bacterium *Aquifex pyrophilus* chromosome. J Bacteriol 176:6776–6780

Stetter KO, Konig H, Stackebrandt E (1983) *Pyrodictium* gen. nov., a new genera of submarine disc-shaped sulfur reducing archaebacteria growing optimally at 105°C. Syst Appl Microbiol 4:535–551

Wetmur JG, Wong DM, Ortiz B, Tong J, Reichert F, Gelfand DH (1994) Cloning, sequencing and expression of RecA proteins from three distantly related thermophilic eubacteria. J Biol Chem 269:25929–25935

Wishart DS, Boyko RF, Willard L, Richard FM, Sykes BD (1994) SEQSEE: a comprehensive program suite for protein sequence analysis. Comp Appl Biosci 10:121–132

Yip KSP, Stillman TJ, Britton KL, Artymiuk PJ, Barker PJ, Sedelnikova SE, Engel PC, Pasquo A, Scandurra R, Rice DW (1995) The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. Structure 3:1147–1158