Extremophiles (1997) 1:36–51

ORIGINAL PAPER

Sorel Fitz-Gibbon · Anthony J. Choi
Jeffrey H. Miller · Karl O. Stetter · Melvin I. Simon
Ronald Swanson · Ung-Jin Kim

A fosmid-based genomic map and identification of 474 genes of the hyperthermophilic archaeon *Pyrobaculum aerophilum*

Received: October 18, 1996 / Accepted: October 30, 1996

Abstract We have constructed a physical map of the approximately 1.7-Mb genome of the hyperthermophilic archaeon *Pyrobaculum aerophilum*. Derived from a 12× coverage genomic fosmid library with an average insert size of 36 Kb, the map consists of a single circular contig of 96 overlapping fosmid clones with 211 markers ordered along them. One hundred of the sequence markers have strong similarities to known genes. Many overlaps were also checked using restriction fingerprint analysis. This map is an important step in the elucidation of the sequence of the entire genome of *Pyrobaculum aerophilum.* To this end we have determined more than 95% of the genome with 15 000 random sequences. Each sequence has been screened against the public sequence databases to identify similarities to known genes. We report here a list of the 474 putative genes we have identified.

Key words *Pyrobaculum aerophilum* · Genome · Hyperthermophile · Archaea · Eocyte

Communicated by J. Wiegel

S. Fitz-Gibbon · A.J. Choi · J.H. Miller (✉)
Department of Microbiology and Molecular Genetics and The Molecular Biology Institute, University of California at Los Angeles, 1602 Molecular Sciences Building, 405 Hilgard Avenue, Los Angeles, CA 90095-1489, USA
Tel. +1-310-825-8460; Fax +1-310-206-3088
e-mail: jhmiller@ewald.mbi.ucla.edu

K.O. Stetter
Archaeenzentrum, Regensburg University, 93053 Regensburg, Germany

M.I. Simon · R. Swanson[1] · U.-J. Kim
Division of Biology, 147-75, California Institute of Technology, Pasadena, CA 91125, USA

Present address:
[1] Recombinant Biocatalysis Inc., Philadelphia, PA, USA

Introduction

Pyrobaculum aerophilum is a rod-shaped hyperthermophilic archaeon isolated from a boiling marine water hole (Völkl et al. 1993). Due to its distinctive characteristics, *P. aerophilum* is both an interesting and a suitable organism for evolutionary and comparative studies. It has an extremely high optimal growth temperature of 98°C, and yet, in contrast to other hyperthermophiles, it is facultatively aerobic. Furthermore, *P. aerophilum* can be plated to form colonies within four days with up to 100% efficiency (Völkl et al. 1993). Since strictly anaerobic metabolism predominates at the upper temperature limits of life (Robb and Place 1995), cultivation of hyperthermophiles has been difficult and consequently they are only poorly described. Based on 16S rRNA analysis, hyperthermophiles are the living organisms with the greatest similarity to ancient organisms (Stetter 1992; Woese 1987). Thus hyperthermophiles are of great interest and, due to its ease of cultivation, *P. aerophilum* is an ideal candidate for development as a model hyperthermophilic organism. We plan the construction of a genetic system which will allow for mutagenesis and reverse genetic studies, and for the analysis of gene function. Determining the entire genome sequence will simplify the development of a genetic system for *P. aerophilum* and will provide valuable data not only for research on hyperthermophily but also for phylogeny research and the elucidation of the least studied domain of life, Archaea.

Our strategy for sequencing the entire genome of *P. aerophilum* has been primarily a random "shotgun" approach. We have sequenced random fragments from the genome, used computer algorithms to compare the sequences, and then assembled the overlapping sequences into contigs. While the "shotgun" sequencing approach (Fleischmann et al. 1995) is an efficient way to obtain most of the sequence, it becomes impractical to complete the sequence by this method. To close the final gaps in the sequence, a directed approach is desirable. We have developed a genome map in order to facilitate a directed approach to sequence completion, to confirm our sequence

contigs, and to aid in resolving repetitive and otherwise difficult regions.

Results

We have constructed a genomic map consisting of 211 markers on 96 overlapping large clones. The large inserts (~36 Kb) are carried in the fosmid vector pFOS1. This vector uses the *E. coli* F-factor origin of replication and thus is kept at a low copy number. This vector has been shown to maintain large insert DNA fragments with relatively high stability (Kim et al. 1992). The entire genomic fosmid library consists of 768 clones, which corresponds to a 12× genomic coverage based on an estimated genome size of 1.7 Mb. This level of redundancy was sufficient to cover the whole genome without gaps. By hybridizing the library with various probes we were able to place markers on the map and determine the overlaps of the fosmids.

Construction and screening of fosmid library

Genomic DNA was partially digested with *Sau*3AI before ligation to the *Bam*HI cloning site of the fosmid arms. The ligation reaction contained a large molar excess of fosmid arms over genomic DNA in order to minimize the chance of chimera formation. Packaging was done using the Stratagene GigaPack XL kit which preferentially packages large inserts. The average insert size for this library was 36 Kb as determined by *Not*I digestion and pulsed field gel electrophoresis (not shown). The entire library was gridded onto single small (8 cm × 12 cm) nylon membranes using the Biomek 1000 laboratory workstation. The library was screened by a variety of probes that are described in detail later.

Fosmid contig assembly and verification

Three methods were used to identify overlapping fosmids. First, small inserts from the pUC18 clones were amplified by PCR and hybridized to high-density colony membranes of the large insert fosmid library. As part of our large-scale random sequencing project, both ends of these pUC18 inserts had previously been sequenced and compared to the public sequence databases. Of the 146 markers on the map, 112 were putative genes. These include ribosomal genes, DNA and RNA polymerase genes, tRNA genes, various repair genes, *recA* homologs, and numerous metabolic genes (Fig. 1). Analysis of these data revealed several ambiguous regions, possibly due to repetitive sequences in the markers. These regions were clarified by a fingerprinting method consisting of computer assisted analysis of the banding patterns of fosmids digested with *Ban*I and *Msp*I and labeled at the *Ban*I ends (Coulson et al. 1986; Sulston et al. 1988). Twenty-four overlaps were confirmed by this method (Fig. 1), yielding 10 clear contigs. The final 10 gaps were closed using RNA probes transcribed from the T7 and Sp6 promoters flanking the fosmid vector cloning site. These RNA probes were radiochemically labeled and hybridized back to the whole fosmid library. Prior to generating the probe, the fosmid DNA was digested with *Hin*cII, which yields DNA fragments averaging approximately 1000 bp. *Hin*cII was chosen over a less frequent cutter, to keep the length of the labeled transcript short. This was important for minimizing the inclusion of repetitive sequences in the riboprobes. We found larger probes often gave ambiguous results. Using this method we successfully identified fosmids spanning the 10 remaining gaps and confirmed 38 other overlaps (Fig. 1). In three cases, a second round of probing was necessary to cross the gaps, using probes from the ends of the fosmids identified in the first round. An overlapping set of 96 fosmids was identified, giving a redundancy of approximately 2×. From the average size of the 96 clones in the map and the overall redundancy that they represent, we estimate the genome size of this microbe to be 1.7 Mb.

Genome sequence

The map reported here will be used as an aid to determining the entire genome sequence of *P. aerophilum.* The sequencing effort has so far yielded more than 15 000 random single-pass sequences which were determined using fluorescent multiplex automated sequencers (ABI 373, Applied Biotechnology). Base calls were made from the ABI trace data using the Phred software package (Brent Ewing). This software not only assigns bases from the trace data but also assigns a quality value for each base call. This factor becomes extremely important when the sequences are assembled, obviating the need for extensive manual editing of contigs. The Phrap software package (Phil Green) makes use of the quality information provided by the Phred software, while other assembly programs allow regions of low-quality data to disrupt the consensus sequence. The Phrap software then uses a novel method of assembly to create contigs which are a mosaic of the highest quality parts of reads (Table 1).

After having completed the random sequencing phase of the project, we have slightly over 800 contigs and thus approximately 800 gaps in our sequence. Poisson analysis suggests an average gap size of less than 200 bases long. Several methods are being used in parallel in order to move quickly to a finished sequence. Where possible, new sequences are being derived from the adjacent regions of partially sequenced pUC18 clones. In most cases an oligo primer has to be designed and synthesized for this purpose. On a small number of clones, a deletion subcloning method (unpublished) is being used to reach the unsequenced region at the center of the insert while still using the standard priming sites on the vector.

Of the 800 gaps in our sequence there are approximately 200 gaps which are not represented in the pUC18 library. The fosmid map and templates will be used to extend the sequences across these gaps. The fosmid map includes 146 sequence markers which have been assembled into the se-

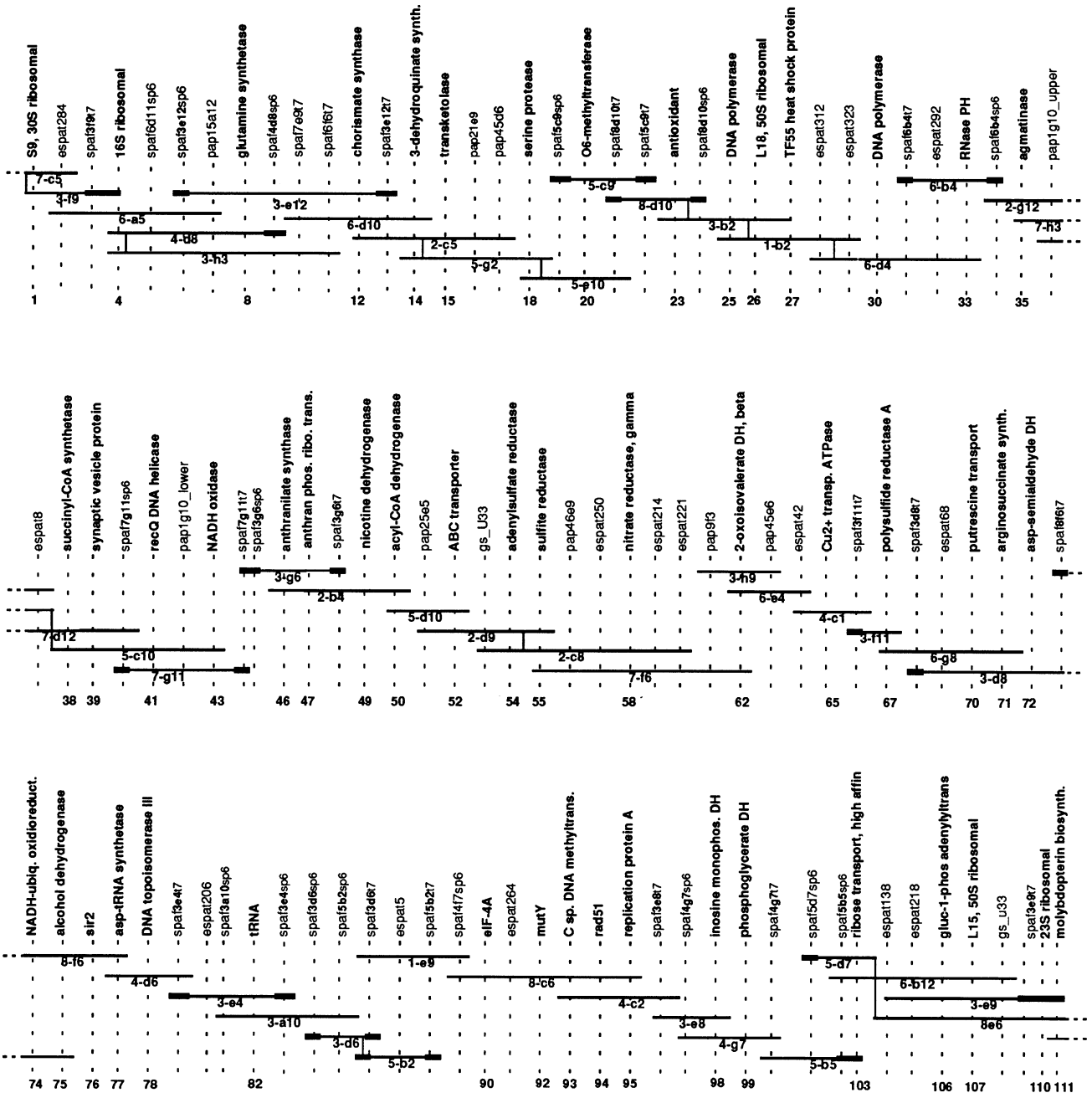

Fig. 1. The *horizontal lines* represent the inserts of fosmid clones covering the entire circular genome of *Pyrobaculum aerophilum*. Fosmid designations are indicated *immediately below the horizontal lines*. *Vertical dashed lines* represent sequence markers placed along the genome via hybridization experiments. Where the sequence tag has similarity to a known gene, the gene name is written in *bold*. All markers were tentatively identified as the listed genes by their similarity to RNA or amino acid sequences in the public databases.

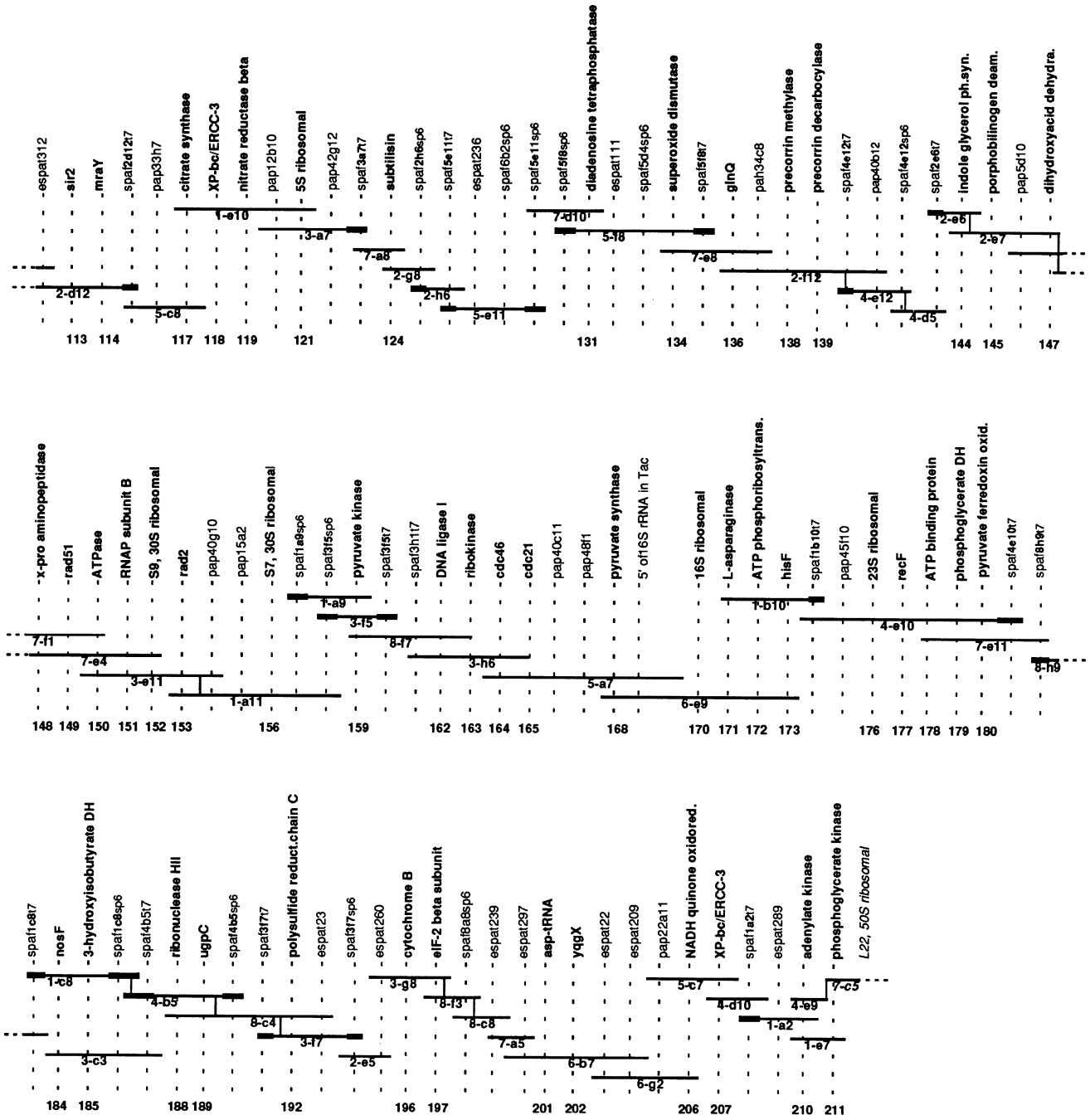

Fig. 1. (*continued*) Details such as blast scores, accession numbers and full names are shown in Table 1a by the corresponding number. Riboprobes are labeled "spaf . . . t7" or "spaf . . . sp6". *Bold regions of the horizontal lines* represent the regions from which the riboprobes were transcribed. *Vertical solid lines* represent the 24 overlaps confirmed by fingerprinting

Table 1a. Similarity data for 100 putative genes which were placed on the fosmid map (Fig. 1). All matches have a spurious match probability of less than 10^{-5} except for eukaryotic IF-4A (90). The numbers (*Map #*) correspond to the markers' position on the fosmid map (Fig. 1)

Map #	Hit identification	Probability (-E)	Accession	Organism
1	S9, 30S ribosomal	30	sp|P39468|	*Sulfolobus acidocaldarius*
4	16S ribosomal	54	gb|M35966|	*Thermoproteus tenax*
8	glutamine synthetase	35	sp|P36205|	*Thermotoga maritima*
12	chorismate synthase	37	pir||S13070	*Synechocystis sp.*
14	3-dehydroquinate synthase	62	sp|P07639|	*Escherichia coli*
15	transketolase	26	gi|1149710	*Clostridium perfringens*
18	serine protease	26	sp|P29139|	*Bacillus polymyxa*
20	O6-methyltransferase	9	sp|P16455|	*Homo sapiens*
23	antioxidant	55	gi|1045502	*Sulfolobus sp.*
25	DNA polymerase	150	gi|807830	*Pyrodictium occultum*
26	L18, 50S ribosomal	52	sp|P14033|	*Methanococcus vannielii*
27	TF55 heat shock protein	79	sp|P28488|	*Sulfolobus shibatae*
30	DNA polymerase	200	gi|807828	*Pyrodictium occultum*
33	RNase PH	15	sp|P37939|	*Mycobacterium leprae*
35	agmatinase	16	sp|P37819|	*Streptomyces clavuligerus*
38	succinyl-CoA synthetase alpha&beta	86	sp|P25126|	*Thermus flavus*
39	synaptic vesicle protein	37	pir||S34961	*Rattus norvegicus*
41	DNA Helicase, *recQ*	22	sp|P30015|	*Escherichia coli*
43	NADH oxidase	20	sp|P37061|	*Enterococcus faecalis*
46	anthranilate synthase	91	sp|Q06128|	*Sulfolobus solfataricus*
47	anthranilate phosphoribosyltransferase	69	gi|149037	*Bacillus pumilus*
49	nicotine dehydrogenase	44	pir||S37570	*Arthrobacter nicotinovorans*
50	acyl-CoA dehydrogenase	33	sp|P46703|	*Mycobacterium leprae*
52	ABC transporter	11	sp|P39326|	*Escherichia coli*
54	adenylyl-sulphate reductase alpha-subunit	44	gi|1183905	*Desulfovibrio vulgaris*
55	sulfite reductase	56	pir|S27479	*Archaeoglobus fulgidus*
58	nitrate reductase gamma subunit	24	sp|P42177|	*Bacillus subtilis*
62	2-oxoisovalerate dehydrogenase, beta	82	sp|P37941|	*Bacillus stearothermophilus*
65	Cu^{2+} transporting ATPase	31	sp|P32113|	*Enterococcus hirae*
67	polysulfide reductase chain *psrA*	29	sp|P31075|	*Wolinella succinogenes*
70	putrescine transport	10	sp|P31136|	*Escherichia coli*
71	argininosuccinate synthase	109	sp|P13256|	*Methanococcus vannielii*
72	aspartate-semialdehyde dehydrogenase	60	sp|P41394|	*Leptospira interrogans*
74	NADH-ubiquinone oxidoreductase	40	sp|P42026|	*Bos taurus*
75	alcohol dehydrogenase	28	sp|P39462|	*Sulfolobus solfataricus*
76	*sir2*	36	gi|845686	*Staphylococcus aureus*
77	asp-tRNA synthetase-prokaryote	62	gi|1146247	*Bacillus subtilis*
78	DNA topoisomerase III	41	gi|1292913	*Homo sapiens*
82	tRNA	12	emb|X05070|	*Thermoproteus tenax*
90	eukaryotic IF-4A	5	gi|1072122	*Homo sapiens*
92	*mutY* = adenine glycosylase	37	sp|P29588|	*Methanobacterium thermo.*
93	cytosine specific DNA methyltransferase	14	gi|1165245	*Neisseria gonorrhoeae*
94	*rad51*	24	pir||S37673	*Homo sapiens*
95	replication protein A	7	sp|P05682|	*Agrobacterium rhizogenes*
98	inosine monophosphate dehydrogenase	7	pir||B48868	*Methanosarcina thermophila*
99	phosphoglycerate dehydrogenase	55	gi|1146196	*Bacillus subtilis*
103	ribose transport, high affinity	23	gi|290599	*Escherichia coli*
106	glucose-1-phosphate adenylyltranferase	21	gi|699151	*Mycobacterium leprae*
107	L15, 50S ribosomal	8	sp|P14032|	*Methanococcus vannielii*
110	23S ribosomal RNA	228	emb|X05480|	*Desulfurococcus mobilis*
111	molybdopterin biosynthesis	44	gi|1002858	*Anabaena sp.*
113	*sir2*	21	gi|845686	*Staphylococcus aureus*
114	phospho-*N*-acetylmuramoyl-pentapeptide-transferase	7	sp|Q03521|	*Bacillus subtilis*
117	citrate synthase	64	sp|P39120|	*Bacillus subtilis*

Table 1a. *Continued*

Map #	Hit identification	Probability (-E)	Accession	Organism
118	DNA excision repair helicase, XP-bc/ERCC-3	14	gi|902048	*Arabidopsis thaliana*
119	nitrate reductase beta	157	sp|P42176|	*Bacillus subtilis*
121	5S ribosomal RNA	40	gb|M16530|	*Sulfolobus sp.*
124	subtilisin	61	sp|P29139|	*Bacillus polymyxa*
131	diadenosine tetraphosphatase (mutT-like)	18	gi|1054947	*Sus scrofa*
134	superoxide dismutase (Mn)	19	sp|Q08713|	*Sulfolobus acidocaldarius*
136	glutamine transport protein *glnQ*	46	sp|P10346|	*Escherichia coli*
138	precorrin methylase	38	sp|Q05630|	*Salmonella typhimurium*
139	precorrin decarbocylase	18	sp|Q05632|	*Salmonella typhimurium*
144	indole-3-glycerol-phosphate synthase	19	sp|Q06121|	*Sulfolobus solfataricus*
145	porphobilinogen deaminase	52	sp|P16616|	*Bacillus subtilis*
147	dihydroxyacid dehydrase	43	sp|P31959|	*Clostridium pasteurianum*
148	X-PRO aminopeptidase	41	gi|1303901	*Bacillus subtilis*
149	*rad51*	91	gi|1054624	*Xenopus laevis*
150	ATPase, alpha, membrane associated	126	sp|P09639|	*Sulfolobus acidocaldarius*
151	RNA polymerase B	94	sp|P11513|	*Sulfolobus acidocaldarius*
152	S9, 30S ribosomal	30	sp|P39468|	*Sulfolobus acidocaldarius*
153	*rad2*	24	sp|P26793|	*Saccharomyces cerevisiae*
156	S7, 30S ribosomal	55	sp|P41206|	*Desulfurococcus mobilis*
159	pyruvate kinase	30	gi|1041097	*Bacillus psychrophilus*
162	DNA ligase	32	sp|Q02093|	*Desulfurolobus ambivalens*
163	DNA ligase	122	sp|Q02093|	*Caenorhabditis elegans*
164	DNA replication licensing factor (MCM2, *cdc46*)	98	gi|852053	*Drosophila melanogaster*
165	*cdc21*	5	sp|P33991|	*Homo sapiens*
168	pyruvate synthase	43	pir||S22396	*Halobacterium halobium*
170	16S ribosomal	47	gb|M38637|	*Thermoplasma acidophilum*
171	L-asparaginase	34	gi|496102	*Lupinus albus*
172	ATP phosphoribosyltransferase	21	gi|1117923	*Yarrowia lipolytica*
173	*hisF* (cyclase) (histidine biosynthesis)	70	sp|Q02133|	*Lactococcus lactis*
176	23S ribosomal RNA	26	gb|M86626|	*Pyrobaculum occultum*
177	*recF*	8	sp|P24900|	*Salmonella thyphimurium*
178	ATP-binding protein	23	gi|1184189	*Escherichia coli*
179	phosphoglycerate dehydrogenase	37	sp|P37666|	*Escherichia coli*
180	pyruvate ferredoxin oxidoreductase	55	gi|1197358	*Pyrococcus furiosus*
184	Cu transporting ATP-binding *nosF*	11	sp|P19844|	*Pseudomonas stutzeri*
185	3-hydroxyisobutyrate DH	47	gi|1001605	*Synechocystis sp.*
188	RNAse HII	9	pir||S53S08	*Saccharomyces cerevisiae*
189	nucleotide-binding protein *ugpC*	80	sp|P10907|	*Escherichia coli*
192	polysulfide reductase chain C, *psrC*	28	sp|P31076|	*Wolinella succinogenes*
196	cytochrome *b*	19	gi|927525	*Sulfolobus acidocaldarius*
197	eIF-2 beta subunit	6	sp|P41375|	*Drosophila melanogaster*
201	asp-tRNA	9	gb|L07321|	*Thermococcus celer*
202	*yqgX*	11	gi|1303871	*Bacillus subtilis*
206	NADH-quinone oxidoreductase	25	sp|P29925|	*Paracoccus denitrificans*
207	DNA excision repair helicase, XP-bc/ERCC-3	16	gi|902048	*Arabidopsis thaliana*
210	adenylate kinase	29	gi|1086550	*Methanococcus jannaschii*
211	phosphoglycerate kinase	69	sp|P20971	*Methanothermus fervidus*

Approximately 15 000 sequences were screened individually against the nonredundant set of public databases using BLAST (Altschul et al. 1990; States et al. 1991) with default settings. The spurious match probabilities (as determined by the BLAST algorithm) are reported as orders of magnitude, e.g., 3.6E-24 becomes 24. All of the matches reported have spurious match probabilities of less than 10^{-5}; most of the matches reported have spurious match probabilities of less than 10^{-10}. Detailed analyses of matches have not been performed. Duplicate copies of genes are not known or reported. The identification and categorization of the matches should be considered tentative. Assignments are based upon Monica Riley's categories (Riley 1993; see also Bult et al. 1996). The cited names of the organisms are as they appear in the data entry of the gene bank and thus do not necessarily reflect the present correct nomenclature of the listed organisms. Upon completion of the sequence, extensive analysis will be performed and reported. We are developing a web site on which our data will be publicly available. E-mail to sorel@ewald.mbi.ucla.edu for web site information or for help with specific data.

Table 1b. Similarity data for 474 *Pyrobaculum aerophilum* random sequence tags

ID #	Hit identification	Probability (-E)	Accession	Organism
Amino acid biosynthesis				
1	glutamate-1-semialdehyde 2,1 aminomutase	24	gp|X53695|	*Synechococcus sp.*
2	NMHase	27	gp|A17961|	*Arthrobacter sp.*
3	ornithine aminotransferase	12	gp|X81802|	*Bacillus subtilis*
4	*S*-adenosyl-L-homocysteine hydrolase	81	gp|Z50174|	*Sulfolobus solfataricus*
5	2-isopropylmalate synthase	42	sp|P05342|	*Lactococcus lactis*
6	2-oxoisovalerate dehydrogenase, beta	43	sp|P37941|	*Bacillus subtilis*
7	3-phosphoshikimate 1-carboxyvinyltrans.	13	sp|P24497|	*Klebsiella pneumoniae*
8	anthranilate phosphoribosyltransferase	45	gp|M65060|	*Methanobacterium thermo.*
9	anthranilate synthase I	68	sp|Q08653|	*Thermotoga maritima*
10	anthranilate synthase II	52	gp|M98048|	*Sulfolobus solfataricus*
11	chorismate mutase	9	sp|P27603|	*Pseudomonas stutzeri*
12	chorismate synthase	34	gp|X67516|	*Synechocystis sp.*
13	dihydroxy-acid dehydratase	31	sp|P39522|	*Saccharomyces cerevisiae*
14	*hisF*	45	sp|P26721|	*Azospirillum brasilense*
15	*hisH*	11	sp|Q02132|	*Lactococcus lactis*
16	histidinol-phosphate transaminase	13	pir|A30270|	*Escherichia coli*
17	imidazoleglycerol-phosphate dehydratase	20	sp|P06633|	*Saccharomyces cerevisiae*
18	indole-3-glycerol-phosphate synthase	22	sp|Q06121|	*Sulfolobus solfataricus*
19	ATP phosphoribosyltransferase	15	sp|Q02129|	*Lactococcus lactis*
20	ketol-acid reductoisomerase	51	pir|A47037|	*Corynebacterium glutam.*
21	kynurenine/alpha-aminoadipate aminotrans.	22	gp|Z50144|	*Rattus norvegicus*
22	*N*-acetylornithine aminotransferase	47	pir|S44189|	*Anabaena sp.*
23	phosphoribosylanthranilate isomerase (*trpF*)	7	gp|X17149|	*Vibrio parahaemolyticus*
24	phosphoribosylpyrophosphate synthetase	15	gp|M76553|	*Leishmania donovani*
25	tryptophan synthase, alpha	5	gp|M65060|	*Methanobacterium thermo.*
26	tryptophan synthase, beta	6	sp|P19868|	*Bacillus stearothermophilus*
27	acetylglutamate kinase	8	sp|P36840|	*Bacillus subtilis*
28	argininosuccinate synthetase (Ass)	51	sp|P13256|	*Methanococcus vannielii*
29	aspartate semialdehyde dehydrogenase	35	pir|A44846|	*Leptospira interrogans*
30	aspartate aminotransferase	12	gp|D50624|	*Streptomyces virginiae*
31	aspartate beta-semialdehyde dehydrogenase	27	gp|M77500|	*Leptospira borgpetersenii*
32	aspartate kinase-homoserine dehydrogenase	15	gp|L33912|	*Zea mays*
33	ATP phosphoribosyltransferase	20	gp|U07830|	*Schizosaccharomyces pombe*
34	cystathionine gamma-synthase	17	sp|P00935|	*Escherichia coli*
35	L-isoaspartyl protein carboxyl methyltrans.	32	gp|M63493|	*Escherichia coli*
36	lysine-sensitive aspartokinase III	20	sp|P08660|	*Escherichia coli*
37	dihydrodipicolinate synthase (*mosA*)	15	sp|Q07607|	*Rhizobium meliloti*
38	succinyl-diaminopimelate desuccinylase	8	sp|P24176|	*Escherichia coli*
39	homoserine kinase (*thrB*)	16	gp|Y00522|	*Calothrix sp.*
40	3-isopropylmalate dehydrogenase	37	gp|U07980|	*Bos taurus*
41	acetolactate synthase, large subunit	41	sp|P37251|	*Bacillus subtilis*
42	branched-chain-amino-acid-transaminase	39	pir|S30668|	*Escherichia coli*
43	branched-chain amino acid transport BRAF	17	sp|P21629|	*Pseudomonas aeruginosa*
44	branched-chain amino acid transport BRAG	23	sp|P21630|	*Pseudomonas aeruginosa*
45	ketol-acid reductoisomerase	56	gp|L03181|	*Bacillus subtilis*
46	argininosuccinate lyase	14	sp|P11447|	*Escherichia coli*
47	argininosuccinate synthetase	16	gp|M21315|	*Methanococcus vannielii*
48	ferredoxin-dependent glutamate synthase	32	gp|U03006|	*Spinacia oleracea*
49	glutamate-1-semialdehyde 2,1 aminomutase	25	pir|A35789|	*Hordeum vulgare*
50	glutamate dehydrogenase	34	gp|L19995|	*Thermococcus litoralis*
51	glutamine synthetase	46	gp|X60160|	*Thermotoga maritima*
52	L-glu-D-fructose-6-phosphate amidotrans.	31	gp|U21932|	*Bacillus subtilis*
53	NADP-specific glutamate dehydrogenase	52	sp|P39475|	*Sulfolobus shibatae*
54	aspartate carbamoyltransferase	14	sp|P19936|	*Serratia marcescens*
55	hydroxymethyltrans.,3-methyl-2-oxobut.	46	sp|P31057|	*Escherichia coli*
56	serine hydroxymethyltransferase	24	gp|U02131|	*Mycoplasma genitalium*
57	serine *O*-acetyltransferase	12	sp|P05796|	*Escherichia coli*
58	thioredoxin reductase	19	pir|S38988|	*Eubacterium acidaminophilum*
Biosynthesis of cofactors, prosthetic groups, and carriers				
59	3-oxoacyl [acyl-carrier protein] reductase	21	sp|P27582|	
60	*adgA*	18	gp|X59399|	*Rhodobacter capsulatus*
61	cobyrinic acid a,c-diamide synthase	19	sp|P21632|	*Pseudomonas denitrificans*
62	dihydroflavonol 4-reductase	24	pir|S38474|	*Lycopersicon esculentum*
63	mRNA for dihydroflavonol-4-reductase	12	gp|X15536|	*Antirrhinum majus*
64	*pqqF/A/B/C*	9	gp|X87299|	*Pseudomonas fluorescens*
65	precorrin 3 methylase	43	sp|Q05590|	*Salmonella typhimurium*
66	precorrin-8W decarboxylase	10	sp|Q05632|	*Salmonella typhimurium*
67	GTP cyclohydrolase I	27	gp|X85954|	*Campylobacter jejuni*

Table 1b. *Continued*

ID #	Hit identification	Probability (-E)	Accession	Organism
68	GTP cyclohydrolase II (*ribA*)	14	sp|P17620|	*Bacillus subtilis*
69	porphobilinogen deaminase	23	gp|M57676|	*Bacillus subtilis*
70	uroporphyrin-III *C*-methyltransferase	14	sp|P29928|	*Bacillus megaterium*
71	lipoate biosynthesis A	19	gp|U32688|	*Haemophilus influenzae*
72	thiamine biosynthetic	7	gp|U17350|	*Zea mays*
73	*chlE/N*	22	gp|M21151|	*Escherichia coli*
74	molybdopterin biosynth. (*moeA*)	15	sp|P12281|	*Escherichia coli*
75	molybdenum cofactor biosynthesis B	32	pir|S31880|	*Escherichia coli*
76	riboflavin synthase, beta	14	gp|U32810|	*Haemophilus influenzae*
Cell envelope				
77	A10	26	gp|L21027|	*Mus musculus*
78	erythromycin biosynth. sensory transductn	19	sp|P39623|	*Bacillus subtilis*
79	gamma-aminobutyrate permease	9	gp|U31756|	*Bacillus subtilis*
80	laminin receptor homolog	25	gp|S35960|	*Homo sapiens*
81	transmembrane	24	gp|U00039|	*Escherichia coli*
82	ankyrin	17	gp|L35601|	*Drosophila melanogaster*
83	integral membrane phosphoprotein band 7.2b	41	gp|U17297|	*Mus musculus*
84	polysulfide reductase A	7	sp|P31075|	*Wolinella succinogenes*
85	proline/betaine transporter	37	sp|P30848|	*Escherichia coli*
86	oligopeptide permease	7	gp|X05491|	*Salmonella typhimurium*
87	membrane spanning	26	gp|X77636|	*Bacillus subtilis*
88	*n*-acetyl-gamma-glutamyl-phosphate reduct	12	sp|Q07906|	*Bacillus stearothermophilus*
Cellular processes				
89	carbon dioxide binding	25	gi|600730|	*Brucella melitensis*
90	carbonic anhydrase	11	sp|P17582|	*Escherichia coli*
91	ERV operon: *frvX*	15	sp|P32153|	*Escherichia coli*
92	*sua5*	43	gp|Z38002|	*Saccharomyces cerevisiae*
93	*ftsYEX* (cell division control)	30	gp|X04398|	*Escherichia coli*
94	(*fcr3*) *cpn60*	6	emb|X75420|	*Plasmodium falciparum*
95	heat-shock	54	gp|D29672|	*Pyrococcus sp.*
96	signal recognition particle	42	sp|P27414|	*Sulfolobus acidocaldarius*
Central intermediary metabolism				
97	2-hydroxyhepta-2,4-diene-1,7-dioate isom.	19	sp|P37352|	*Escherichia coli*
98	alpha-amylase (*amyA*)	12	gp|L13279|	*Escherichia coli*
99	enoyl-CoA hydratase	26	gp|X79899|	*Clostridium difficile*
100	glyoxysomal citrate synthase	16	gp|D38132|	*Cucurbita sp.*
101	inorganic pyrophosphatase	28	gp|X83728|	*Nicotiana tabacum*
102	malate oxidoreductase	39	sp|P16468|	*Bacillus stearothermophil.*
103	nodulation ATP-binding I	8	pir|S27496|	*Bradyrhizobium japonicum*
104	pyruvate-flavodoxin oxidoreductase (nifj)	10	sp|P03833|	*Klebsiella pneumoniae*
105	ribokinase	14	sp|P05054|	*Escherichia coli*
106	desulfoviridin, gamma	18	gp|L05610|	*Desulfovibrio vulgaris*
Energy metabolism				
107	3-hydroxybutyryl-CoA dehydrogenase	11	pir|A43723|	*Clostridium acetobutylicum*
108	3-hydroxyisobutyrate dehydrogenase	25	sp|P23523|	*Escherichia coli*
109	3-oxoacyl-[acyl-carrier-protein] reductase	14	sp|P28643|	*Cuphea lanceolata*
110	4-coumarate-CoA ligase	20	gp|U12012|	*Pinus taeda*
111	5-oxopent-3-ene1,2,5-tricarbox. decarbox.	20	gp|X75028|	*Escherichia coli*
112	acetyl-CoA synthetase	36	gp|M63968|	*Methanothrix soehngenii*
113	adenylsulfate reductase	28	pir|S18928|	*Archaeoglobus fulgidus*
114	subtilisin	61	sp|P29139|	*Bacillus polymyxa*
115	aminopeptidase P	25	gi|1046027|	*Mycoplasma genitalium*
116	*N*-acetylglutamate-g-semialdehyde DH (*argC*)	15	sp|Q07906|	*Bacillus stearothermophil.*
117	ATPase	87	gp|J04836|	*Methanosarcina barkeri*
118	*atrP*	13	gp|X86160|	*Escherichia coli*
119	beta-lactamase	9	sp|P00811|	*Escherichia coli*
120	carbamoylphosphate synthetase, large	31	gi|1304392|	*Sulfolobus solfataricus*
121	cellulase	26	gp|L32742|	*Caldocellum saccharolyticum*
122	cephalosporin C acylase	9	gp|A17015|	*Pseudomonas diminuta*
123	chloroplast ATPase	26	gp|X60752|	*chloroplast Odontella sine.*
124	dihydrolipoamide dehydrogenase	27	sp|P11959|	*Bacillus stearothermophil.*
125	DR-nm23	16	gp|U29656|	*Homo sapiens*
126	enolase	17	sp|P29201|	*Haloarcula*
127	*fadBA* operon (fatty acid oxidizing)	8	gb|M74164|	*Escherichia coli*
128	formate dehydrogenase	9	gp|X54057|	*Wolinella succinogenes*
129	fumarate reductase	25	pir|S34619|	*Thermoplasma acidophilum*
130	glucosamine-fruc.-6-phos. aminotransferase	24	sp|P39754|	*Bacillus subtilis*

Table 1b. *Continued*

ID #	Hit identification	Probability (-E)	Accession	Organism
131	glyceraldehyde 3-phosphate dehydrogenase	49	sp|P19315|	*Methanobacterium formic.*
132	indolepyruvate decarboxylase	19	sp|P2323|	*Enterobacter cloacae*
133	L-isoaspartyl protein carboxyl methyltrans.	24	sp|P24206|	*Escherichia coli*
134	lactaldehyde dehydrogenase	11	gp|M64541|	*Escherichia coli*
135	lipoic acid metabolism (*lipB*)	37	sp|P30976|	*Escherichia coli*
136	membrane-associated ATPase alpha	49	gp|J03218|	*Sulfolobus acidocaldarius*
137	membrane-associated ATPase beta	58	sp|P13052|	*Sulfolobus acidocaldarius*
138	*mmgC*	15	gp|U29084|	*Bacillus subtilis*
139	NADH oxidase	25	sp|P37061|	*Enterococcus faecalis*
140	NADH-plastoquinone oxidoreductase	26	sp|P27724|	*Synechocystis sp.*
141	*nap57*	35	gp|Z34922|	*Rattus norvegicus*
142	nitrate reductase, alpha (*narZ*)	68	sp|P19319|	*Escherichia coli*
143	nitrate reductase, beta (*narY*)	58	sp|P42176|	*Bacillus subtilis*
144	nitrate reductase, gamma	9	gp|Z49884|	*Bacillus subtilis*
145	nitrate transporter	9	sp|P38044|	*Synechococcus sp*
146	periplasmic divalent cation tolerance (*cutA*)	17	sp|P36654|	*Escherichia coli*
147	phospho-2-dehydro-3-deoxyheptonate aldol.	29	pir|S21418|	*Bacillus subtilis*
148	porphobilinogen deaminase	25	gp|M95623|	*Homo sapiens*
149	porphobilinogen synthase (*hem B*)	48	pir|S42531|	*Synechococcus sp.*
150	precorrin decarbocylase	9	sp|Q05632|	*Salmonella typhimurium*
151	precorrin methylase	22	s|p|Q05630|	*Salmonella typhimurium*
152	precorrin-3 methylase	17	sp|P21922|	*Pseudomonas denitrificans*
153	glucose-1-phosphate cytidylyltrans. (*Psa1p*)	11	gp|U19608|	*Saccharomyces cerevisiae*
154	thiosulfate sulfurtransferase	21	sp|P16385|	*Saccharopolyspora erythr.*
155	pyruvate phosphate dikinase	41	gp|U02529|	*Entamoeba histolytica*
156	pyruvate synthase	57	pir|S22397|	*Halobacterium halobium*
157	pyruvate, orthophosphate dikinase	44	sp|P11155|	*Zea mays*
158	ribose 5-phosphate isomerase A	10	gp|U10438|	*Caenorhabditis elegans*
159	sucrase-isomaltase	13	gp|L25926|	*Rattus norvegicus*
160	succinate dehydrogenase (ubiq.) iron-sulfur	15	sp|P21914|	*Drosophila melanogaster*
161	sulfite reductase	18	pir|S27479|	*Archaeoglobus fulgidus*
162	enoyl-CoA hydratase	19	gp|Z27079|	*Caenorhabditis elegans*
163	tungsten formylmethanofuran dehydrogenase	7	gp|X87970|	*Methanobacterium thermo.*
164	tartrate dehydratase	9	sp|P05847|	*Escherichia coli*
165	*ttuD*	16	gp|U32375|	*Agrobacterium vitis*
166	UDP-glucose 4-epimerase	29	sp|P45602|	*Klebsiella pneumoniae*
167	uroporphyrinogen III methyltransferase	11	gp|U05002|	*Bacillus megaterium*
168	x-aconitate hydratase	41	sp|P09339|	*Bacillus subtilis*
169	anaerobic dimethyl sulfoxide reductase A	13	gp|U32785|	*Haemophilus influenzae*
170	NADH dehydrogenase	45	sp|P42026|	*Bovine*
171	Carbamate kinase	31	sp|P13982|	*Pseudomonas aeruginosa*
172	cystathionine gamma-lyase	44	gp|S52028|	*Homo sapiens*
173	L-asparaginase	20	sp|P30362|	*Lupinus arboreus*
174	4-hydroxybutyrate dehydrogenase	14	gp|L36817|	*Alcaligenes eutrophus*
175	formate DH, nitrate-inducible, alpha	12	pir|S18213|	*Wolinella succinogenes*
176	iron sulfur	26	gp|M96826|	*Methanothermus fervidus*
177	L-lactate dehydrogenase	33	gp|M93720|	*Plasmodium falciparum*
178	adenosine triphosphatase	15	gp|L33259|	*Helicobacter pylori*
179	ATP sulfurylase	7	gp|U07353|	*Penicillium chrysogenum*
180	aldehyde:ferredoxin oxidoreductase	8	gp|X79777|	*Pyrococcus furiosus*
181	cytochrome *b*	14	sp|P39480|	*Sulfolobus acidocaldarius*
182	cytochrome *c* biogenesis	24	gp|Z22517|	*Bradyrhizobium japonicum*
183	cytochrome *c* oxidase assembly factor (*cyoE*)	30	gp|U00013|	*Mycobacterium leprae*
184	cytochrome *c* oxidase polypeptide I	35	sp|P08681|	*Mitochondria Chlamydomonas*
185	cytochrome oxidase I	57	gp|U08900|	*Mitochondria Trichoniscus pusil.*
186	cytochrome oxidase II	15	gi|155085|	*Thermus thermophilus*
187	D-lactate dehydrogenase (cytochrome)	14	gp|Z67750|	*Saccharomyces cerevisiae*
188	electron transfer flavoprotein alpha	40	gp|U17110|	*Clostridium acetobutylicum*
189	electron transfer flavoprotein beta	20	sp|P38975|	*Paracoccus denitrificans*
190	ferredoxin oxidoreductase	20	gp|X64521|	*Halobacterium halobium*
191	ferredoxin-nitrite reductase	14	pir|S30920|	*Nicotiana tomentosiformis*
192	formaldehyde:ferredoxin oxidoreductase	57	gp|X83963|	*Thermococcus litoralis*
193	molybdenum- iron-sulfur flavoprotein (*codH*)	12	gp|X82447|	*Oligotropha carboxidovora.*
194	polysulfide reductase B	43	sp|P31076|	*Wolinella succinogenes*
195	polysulfide reductase C	5	sp|P31077|	*Wolinella succinogenes*
196	quinol oxidase polypeptide I/II	22	sp|P39481|	*Sulfolobus acidocaldarius*
197	sulfide dehydrogenase	7	sp|Q06530|	*Chromatium vinosum*
198	thiosulfate reductase	40	gp|L32188|	*Salmonella typhimurium*
199	alcohol dehydrogenase	26	sp|P12311|	*Bacillus stearothermophil.*
200	alcohol dehydrogenase family member Ke 6	19	pir|A48154|	*Mus musculus*

Table 1b. *Continued*

ID #	Hit identification	Probability (-E)	Accession	Organism
201	aldehyde dehydrogenase, cytosolic	34	sp|P13601|	*Rattus norvegicus*
202	(salicyl)aldehyde dehydrogenase	31	gp|U19817|	*Arthrobacter globiformis*
203	3-phosphoglycerate kinase	21	gp|M55529|	*Methanothermus fervidus*
204	4-aminobutyrate aminotransferase	13	sp|P40829|	*Mycobacterium leprae*
205	fructose 1,6-biphosphate aldolase (class II)	21	gp|X14436|	*Escherichia coli*
206	*otsA*	27	gp|U15187|	*Mycobacterium leprae*
207	*pgk*	49	gp|X80178|	*Sulfolobus solfataricus*
208	phosphoenolpyruvate-utilizing	27	gp|S74619|	*Staphylothermus marinus*
209	phosphoglycerate kinase	17	sp|P20971|	*Methanothermus fervidus*
210	pyruvate kinase	12	gp|U12980|	*Saccharomyces cerevisiae*
211	pyruvate, water dikinase	59	gp|U08376|	*Pyrococcus furiosus*
212	fumarate dehydrogenase	46	gp|X75402|	*Sulfolobus solfataricus*
213	fumarate hydratase class II	41	sp|P39461|	*Sulfolobus solfataricus*
214	glucose dehydrogenase	7	gp|D90044|	*Bacillus megaterium*
215	transketolase	20	gp|X67688|	*Homo sapiens*
216	lipoamide dehydrogenase	15	sp|P35484|	*Acholeplasma laidlawii*
217	phosphonopyruvate decarboxylase	11	gp|D37809|	*Streptomyces hygroscopic.*
218	pyruvate dehydrobenase E1	24	sp|P35488|	*Acholeplasma laidlawii*
219	galactokinase	7	sp|Q01415|	*Homo sapiens*
220	aconitate hydratase	50	sp|P37032|	*Legionella pneumophila*
221	carboxyphosphonoenolpyruvate mutase	36	gp|X67953|	*Streptomyces hygroscopic.*
222	citrate synthase	34	gp|X55282|	*Thermoplasma acidophilum*
223	citrate synthase II	42	gp|U05257|	*Bacillus subtilis*
224	hydroxymethylglutaryl-CoA reductase	54	gp|X54658|	*Hevea brasiliensis*
225	malic acid	24	gb|M19485|	*Bacillus stearothermophil.*
226	phosphorylated isocitrate dehydrogenase	57	pdb|4ICD|	*Escherichia coli*
227	succinyl-CoA synthetase	49	sp|P25126|	*Thermus aquaticus*
Fatty acid and phospholipid metabolism				
228	acetyl-CoA synthetase	71	sp|P27095|	*Methanothrix soehngenii*
229	acyl-CoA dehydrogenase	19	sp|P15650|	*Rattus norvegicus*
230	acyl-coA ligase(luciferase)	49	sp|P29212|	*Escherichia coli*
231	beta-ketothiolase	33	gp|L37761|	*Acinetobacter sp.*
232	enoyl-coA hydratase, mitochondrial	14	gp|X15958|	*Rattus norvegicus*
233	glyoxysomal beta-ketoacyl-thiolase	7	gp|X93015|	*Brassica napus*
234	geranylgeranyl pyrophosphate synthase	29	gp|M87280|	*Sulfolobus acidocaldarius*
235	geranyltranstransferase	12	sp|Q08291|	*Bacillus stearothermophil.*
236	hexaprenyl pyrophosphate synthetase	10	gp|Z26494|	*Saccharomyces cerevisiae*
237	HMG-CoA reductase	25	sp|P34135|	*Haloferax volcanii*
238	isoprenyl diphosphate synthase	12	gp|S75695|	*Methanobacterium thermo.*
239	long chain fatty acid coA ligase	12	gp|U32686|	*Haemophilus influenzae*
240	medium chain fatty acid coA ligase	31	sp|Q00594|	*Pseudomonas oleovorans*
241	thiolase	28	pir|JC4032|	*Clostridium acetobutylicum*
Purines, pyrimidines, nucleosides, and nucleotides				
242	diadenosine tetraphosphatase (*apaH*)	12	gp|X04711|	*Escherichia coli*
243	mannosyltransferase B	6	gp|D43637|	*Escherichia coli*
244	deoxycytidine triphosphate deaminase	41	sp|Q02103|	*Desulfurolobus ambivalens*
245	thioredoxin reductase	36	gp|L04500|	*Eubacterium acidaminophi.*
246	CTP synthase (*pyrG*)	62	gp|U00021|	*Mycobacterium leprae*
247	rRNA (adenosine-*N*6,*N*6-)-dimethyltransf.	11	sp|P06992|	*Escherichia coli*
248	adenylate kinase	18	gp|U39882|	*Methanococcus jannaschii*
249	amidophosphoribosyltransferase	11	gp|D28868|	*Arabidopsis thaliana*
250	glu phosphoribosylpyrophosphate amidotrans.	7	sp|P00497|	*Bacillus subtilis*
251	GMP synthase (glutamine-hydrolyzing)	44	sp|P29727|	*Bacillus subtilis*
252	*guaA*	33	gp|U00015|	*Mycobacterium leprae*
253	phosphoribosylamine-glycine ligase	17	sp|P12039|	*Bacillus subtilis*
254	phosphoribosylformylglycinamidine synthase I	21	sp|P12041|	*Bacillus subtilis*
255	phosphoribosylformylglycinamidine synthase II	29	gp|M85265|	*Lactobacillus casei*
256	phosphoribosylglycinamide formyltransf.	11	pir|S37105|	*Arabidopsis thaliana*
257	pur operon encoding purine biosynthesis	20	gp|J02732|	*Bacillus subtilis*
258	*purL*	12	gp|U15182|	*Mycobacterium leprae*
259	carbamoyl-phosphate synthetase	47	gp|J05503|	*Mesocricetus auratus*
260	dihydroorotate synthetase	19	gp|U09990|	*Methanobacterium thermo.*
261	dihydroorotate dehydrogenase	6	sp|P32747|	*Schizosaccharomyces pom.*
262	orotate phosphoribosyltransferase	19	pir|A30492|	*Bacillus subtilis*
263	phosphoribosyl-amp 1,6 cyclohydrolase	25	gp|X82010|	*Rhodobacter sphaeroides*
264	carbamyl phosphate synthetase	31	gp|X87371|	*Saccharomyces cerevisiae*
265	*thiA*	17	gp|U26178|	*Bacillus subtilis*
266	UDP galactose 4-epimerase	21	gp|M94964|	*Klebsiella pneumoniae*
267	UMP synthase	24	gp|U22260|	*Caenorhabditis elegans*

Table 1b. *Continued*

ID #	Hit identification	Probability (-E)	Accession	Organism
268	uracil phosphoribosyl transferase	17	gp|U10246|	*Toxoplasma gondii*
269	cytidine deaminase	28	sp|P19079|	*Bacillus subtilis*
270	HPRT	8	gb|M88110|	*Plasmodium falciparum*
271	uridine phosphorylase	45	sp|P12758|	*Escherichia coli*
272	galactose-1-phosphate uridylyltransferase	21	sp|P31764|	*Haemophilus influenzae*
Regulatory functions				
273	adenosylhomocysteinase	36	sp|P10760|	*rat*
274	ATP synthase	11	gp|M22402|	*Sulfolobus acidocaldarius*
275	carbon starvation A	14	sp|P15078|	*Escherichia coli*
276	GTP binding (*hflX*)	18	gp|U00019|	*Mycobacterium leprae*
277	hydrobenase expression	9	sp|P31905|	*Alcaligenes eutrophus*
278	indoleacetamide hydrolase	12	sp|P06618|	*Pseudomonas syringae*
279	GTP-binding	10	gp|Z49068|	*Caenorhabditis elegans*
280	lactose operon repressor	6	sp|P03023|	*Escherichia coli*
281	mg11	14	gp|U15635|	*Mus musculus*
282	pleiotropic regulatory	44	gp|M29002|	*Bacillus stearothermophilus*
283	repressor	7	gp|M81646|	*Agrobacterium tumefacie.*
284	signal recognition particle (SRP54)	11	sp|P37106|	*Arabidopsis thaliana*
Replication				
285	DMC1/*rad51*/*recA*	43	sp|P25453|	*yeast*
286	*recF*	9	sp|P24900|	*S. typhimurium*
287	*recQ*	25	sp|P30015|	*E coli*
288	DNA repair helicase (*rad25*)	7	sp|P19447|	*yeast*
289	DNA repair (*rad2*)	8	sp|P39750|	*Schizosaccharomyces pom.*
290	DNA repair, ionizing radiation (XRCC1)	5	sp|P18887|	*human*
291	*mutT*	8	gp|U00021|	*Mycobacterium leprae*
292	O6-methyltransferase	9	sp|P16455|	*human*
293	methyltransferase (*uvrC*)	5	gp|L29642|	*Pseudomonas fluorescens*
294	cytosine sp. DNA methyltransferase	20	gp|S86113|	*Neisseria gonorrhoeae*
295	possible G-T mismatches repair	16	sp|P29588|	*Methanobacterium thermo.*
296	*umuD*/*lexA*-type , UV SOS operon	9	sp|P04153|	*Escherichia coli*
297	SIR2 (silent information regulator)	7	sp|P06700|	*yeast*
298	superoxide dismutase (Mn)	19	sp|Q08713|	*Sulfolobus acidocaldarius*
299	reverse gyrase	41	gp|L10651|	*Sulfolobus acidocaldarius*
300	DNA replication licensing factor	19	gp|L41762|	*Drosophila melanogaster*
301	DNA topoisomerase I	11	gp|L27797|	*Bacillus subtilis*
302	DNA topoisomerase III	21	sp|P13099|	*Saccharomyces cerevisiae*
303	DNA topoisomerase (ATP-hydrolysing)	42	pir|S47332|	*Erwinia carotovora*
304	modification methylase FNUDI	14	sp|P34906|	*Fusobacterium nucleatum*
305	heat shock	25	gi|473965|	*Pyrococcus sp.*
306	TF55 heat shock	80	sp|P28488|	*Sulfolobus shibatae*
307	TF56	57	gp|L34691|	*Sulfolobus shibatae*
308a	DNA polymerase	34	gp|M74198|	*Thermococcus litoralis*
308b	DNA polymerase	59	gi|807828|	*Pyrodictium occultum*
308c	DNA polymerase	63	gp|D38573|	*Pyrodictium occultum*
308d	DNA polymerase	60	gp|D38574|	*Pyrodictium occultum*
308e	DNA polymerase I	22	sp|P26811|	*Sulfolobus solfataricus*
309	replication factor C, small subunit	23	sp|P35249|	*Homo sapiens*
310	replication factor C, large subunit	12	sp|P35601|	*Mus musculus*
311	adenylylsulfate reductase	36	gp|X63435|	*Archaeoglobus fulgidus*
312	alkyl hydroperoxide reductase	49	gp|U36479|	*Sulfolobus sp.*
313	diadenosine tetraphosphatase	19	gi|1054947|	*Sus scrofa*
314	DNA-ligase	122	sp|Q02093|	*Desulfurolobus*
315	endonuclease III	19	gp|U11289|	*Bacillus subtilis*
316	Holliday junction DNA helicase	5	gp|U32716|	*Haemophilus influenzae*
317	minichromosome maintenance	6	sp|P29469|	*Saccharomyces cerevisiae*
318	protein-L-isoaspartate (D-asp) *O*-methyltrans.	13	gp|U09669|	*Caenorhabditis elegans*
319	structure specific endonuclease	10	sp|P26793|	*Saccharomyces cerevisiae*
320	XPBara	9	gp|U29168|	*Arabidopsis thaliana*
321	YSA1	7	sp|Q01976|	*Saccharomyces cerevisiae*
Transcription				
322	RNA polymerase sigma factor (*ntrA*)r	11	gp|X69959|	*Azorhizobium caulinodans*
323	TATA-binding	57	gp|U23419|	*Sulfolobus shibatae*
324	RNase PH (tRNA nucleotidyltransferase)	15	gp|L10328|	*Escherichia coli*
325a	DNA-directed RNA polymerase	41	pir|A28213|	*Methanobacterium thermo.*
325b	DNA-directed RNA polymerase II, RpB10	16	sp|P29199|	*Haloarcula marismortui*
325c	DNA-directed RNA polymerase II, RpB3	17	pir|D44126|	*Halobacterium marismort.*
326	DNA-directed RNA polymerase A	81	sp|P31813|	*Thermococcus celer*

Table 1b. *Continued*

ID #	Hit identification	Probability (-E)	Accession	Organism
327	DNA-directed RNA polymerase B	65	sp|P11513|	*Sulfolobus acidocaldarius*
328	DNA-directed RNA polymerase C	38	gp|X14818|	*Sulfolobus acidocaldarius*
329	DNA-directed RNA polymerase E	29	sp|P39466|	*Sulfolobus acidocaldarius*
330	RNA helicase	20	gp|L01622|	*Escherichia coli*
331	transcription activator	11	gp|M73546|	*Bacillus subtilis*
332	transcription elongation factor S-II	26	sp|Q07271|	*Sulfolobus acidocaldarius*
333	transcription factor IIB	18	gp|U20899|	*Sulfolobus shibatae*
334	transcriptional activator (*tenA*)	11	sp|P25052|	*Bacillus subtilis*
Translation				
335	5S rRNA	41	gb|M16530|	*Sulfolobus sp.*
336	16S rRNA	43	gb|M38637|	*Thermoplasma acidophilum*
337	23S rRNA	65	emb|X05480|	*Desulfurococcus mobilis*
338	tRNA/5S rRNA gene cluster	12	emb|X00916|	*Methanococcus vannielii*
339	ala-tRNA	16	emb|X05069|	*Thermoproteus tenax*
340	ala-tRNA synthetase	64	sp|P35029|	*Sulfolobus solfataricus*
341	asn-tRNA synthetase	29	sp|P10723|	*Brugia malayi (nematode)*
342	asp-tRNA	26	emb|X07692|	*Methanobacterium thermo.*
343	asp-tRNA synthetase	38	gp|U32810|	*Haemophilus influenzae*
344	cys-tRNA synthetase	59	gp|X56234|	*Escherichia coli*
345	glu-tRNA	7	gb|L36898|	*Saccharomyces cerevisiae*
346	glu-tRNA synthetase	17	gp|X07466|	*Homo sapiens*
347	gly-tRNA	96	emb|X14835|	*Thermofilum pendens*
348	gly-tRNA synthetase	27	gp|X78993|	*Saccharomyces cerevisiae*
349	his-tRNA synthetase	19	gp|L36863|	*Sulfolobus shibatae*
350	ile-tRNA synthase	43	pir|S21569|	*Methanobacterium thermo.*
351	Iso-tRNA synthetase	27	sp|P41368|	*Staphylococcus aureus*
352	leu-tRNA	25	emb|X05071|	*Thermoproteus tenax*
353	leu-tRNA synthetase	22	sp|P11325|	*mitochondrial – yeast*
354	met-tRNA	9	gb|M26978|	*Methanothermus fervidus*
355	met-tRNA synthetase	13	pir|A25424|	*Saccharomyces cerevisiae*
356	pro-tRNA	8	gb|J01365|	*Saccharomyces cerevisiae*
357	pro/glu-tRNA synthetase	28	gp|M74104|	*Drosophila melanogaster*
358	ser-tRNA	10	gb|M97644|	*Methanopyrus kandleri*
359	ser-tRNA synthetase	49	sp|P26636|	*Cricetulus griseus*
360	thr-tRNA synthetase	22	sp|P18256|	*Bacillus subtilis*
361	trp-tRNA	11	gb|K02528|	*Halobacterium volcanii*
362	tyr-tRNA synthetase	15	gp|L12221|	*Saccharomyces cerevisiae*
363	val-tRNA synthetase	18	sp|P36420|	*Lactobacillus casei*
364	elongation factor, EF 1 alpha	72	sp|P35021|	*Sulfolobus solfataricus*
365	elongation factor, EF 2	75	gp|X69297|	*Sulfolobus solfataricus*
366	Initiation factor, IF 4A	5	sp|P10081|	*Saccharomyces cerevisiae*
367	initiation factor, IF 5A	19	sp|P28461|	*Sulfolobus acidocaldarius*
368	L2, 50S ribosomal	16	sp|P21479|	*Methanococcus vannielii*
369	L3, 60/50S ribosomal	14	sp|P35684|	*Oryza sativa*
370	L4, 50S ribosomal	12	sp|Q06845|	*Halobacterium halobium*
371	L6,50S ribosomal	21	sp|P14030|	*Methanococcus vanelli*
372	L10, ribosomal, acidic	11	sp|P35023|	*Sulfolobus solfataricus*
373	L11, 50S ribosomal	21	sp|P35025|	*Sulfolobus solfataricus*
374	L12, ribosomal	24	gp|X59038|	*Sulfolobus solfataricus*
375	L13, 50S ribosomal	22	sp|P29198|	*Haloarcula marismortui*
376	L15, 50S ribosomal	28	gp|U16148|	*Thermoplasma acidophilum*
377	L18, 50S ribosomal	50	sp|P14033|	*Methanococcus vannielii*
378	L23, ribosomal	19	gp|Y00772|	*Methanococcus vannielii*
379	L24, 50S ribosomal	26	sp|P14034|	*Methanococcus vannielii*
380	L29, 50S ribosomal	10	sp|P04457|	*Bacillus stearothermophil.*
381	L30, 50S ribosomal	18	sp|P14035|	*Methanococcus vanelli*
382	L32e, 50S ribosomal	17	sp|P14549|	*Methanococcus vannielii*
383	L37a, 60S ribosomal	16	pir|S24170|	*Gallus gallus*
384	S3, 30S ribosomal	15	sp|P20281|	*Haloarcula marismortui*
385	S5, 30S ribosomal	40	sp|P14036|	*Methanococcus vannielii*
386	S6, ribosomal	10	sp|P17116|	*Escherichia coli*
387	S7, 30S ribosomal	43	sp|P35026|	*Sulfolobus solfataricus*
388	S8, 30S ribosomal	35	sp|P14038|	*Methanococcus vannielii*
389	S10, 30S ribosomal	12	sp|P23357|	*Haloarcula marismortui*
390	S11, 30S ribosomal	57	sp|P39469|	*Sulfolobus acidocaldarius*
391	S12, 30S ribosomal	62	sp|P29161|	*Thermococcus celer*
392	S13, 40S ribosomal	46	sp|Q05761|	*Zea mays*
393	S15, ribosomal	21	pir|A35908|	*Homo sapiens*
394	S17, ribosomal	14	pir|A24028|	*Rattus norvegicus*

Table 1b. *Continued*

ID #	Hit identification	Probability (-E)	Accession	Organism
395	S18, 40S ribosomal	20	gp|Z46260|	*Saccharomyces cerevisiae*
396	S26e, 40S ribosomal	13	pir|S47942|	*Saccharomyces cerevisiae*
397	ribosomal alanine acetyltransferase	10	sp|P09453|	*Escherichia coli*
398	trmlp	26	gp|Z48758|	*Saccharomyces cerevisiae*
399	tRNA-splicing endonuclease ß	6	sp|P16658|	*Saccharomyces cerevisiae*
400	ARD1 *N*-acetyl transferase	12	gp|X77588|	*Homo sapiens*
401	multifunctional aminoacyl-tRNA synthetase	26	sp|P07814|	*Homo sapiens*
402	ATP-dependent protease	10	gp|L19301|	*Myxococcus xanthus*
403	beta-type proteasome	25	gp|U22157|	*Methanosarcina thermophi.*
404	dipeptidase	17	gp|D13142|	*Sus scrofa*
405	*leuA*	11	gp|U18657|	*Haemophilus influenzae*
406	lysosomal alpha glucosidase	28	gp|Y00839|	*Homo sapiens*
407	prolidase	11	gp|Z34896|	*Lactobacillus delbrueckii*
408	proteasome, beta	18	sp|P28061|	*Thermoplasma acidophilum*
409	protein serine/threonine phosphatase 2A	14	gp|X56261|	*Saccharomyces cerevisiae*
410	TVG	29	gp|D13178|	*Thermoactinomyces vulga.*
411	sulfite reductase	10	pir|S27478|	*Archaeoglobus fulgidus*
Transport and binding proteins				
412	2′–5′ oligoadenylate binding	25	pir|S52166|	*Homo sapiens*
413	*appB*	11	gp|U20909|	*Bacillus subtilis*
414	ATP-binding	25	gp|Z25798|	*Bacillus subtilis*
415	*cysA*	20	gp|M32101|	*Escherichia coli*
416	maltose-binding periplasmic	11	sp|P19576|	*Salmonella typhimurium*
417	*natA/B*	31	gp|U30873|	*Bacillus subtilis*
418	p87	16	gp|S47919|	*Bos sp.*
419	alkylphosphonate uptake (*phn*)A-Q	16	gp|J05260|	*Escherichia coli*
420	selenium-binding; ap56	36	sp|P17563|	*Mus musculus*
421	transport sec61, alpha	15	sp|P38379|	*Pyrenomonas salina*
422	arginine permease, substrate-binding	17	gp|U08865|	*Listeria monocytogenes*
423	D-lactate dehydrogenase	18	sp|P32891|	*Saccharomyces cerevisiae*
424	*dciAD*	17	gp|X56678|	*Bacillus subtilis*
425	dipeptide transport system permease *dppB*	24	sp|P37316|	*Escherichia coli*
426	dipeptide transport system permease *dppC*	32	gp|U17295|	*Haemophilus influenzae*
427	glutamine permease	27	gp|X14180|	*Escherichia coli*
428	glutamine-binding periplasmic	19	sp|P10344|	*Escherichia coli*
429	branched-chain amino acid transport	46	sp|P30294|	*Salmonella typhimurium*
430	spoOk operon	26	gp|M57689|	*Bacillus subtilis*
431	oligopeptide perm. (*oppC*, sporulation initiation)	20	sp|P24139|	*Bacillus subtilis*
432	oligopeptide transport permease *appB*	10	sp|P42062|	*Bacillus sbutilis*
433	oligopeptide transport permease *appC*	8	sp|P42063|	*Bacillus subtilis*
434	oligopeptide transport ATP binding *appF*	32	sp|P42065|	*Bacillus subtilis*
435	oligopeptide transport ATP-binding *oppF*	51	sp|P08007|	*Salmonella typhimurium*
436	phosphonates transport ATP-binding *phnC*	21	sp|P16677|	*Escherichia coli*
437	phosphonates transp. sys. permease (*phnE*)	9	sp|P16683|	*Escherichia coli*
438	translocase SecY	23	gp|X85020|	*Sulfolobus acidocaldarius*
439	proline/betaine transporter	24	gp|U14003|	*Escherichia coli*
440	putrescine transport system permease (*potl*)	10	sp|P31136|	*Escherichia coli*
441	spermidine/putrescine-binding periplasmic	10	gp|U32731|	*Haemophilus influenzae*
442	phosphate transport ATP-binding *pstB*	45	sp|P07655|	*Escherichia coli*
443	lactose transport ATP-binding *lack*	13	sp|Q01937|	*Agrobacterium radiobacter*
444	sterol carrier -2	24	gp|M75884|	*Homo sapiens*
445	sugar-binding	7	gp|M77351|	*Streptococcus mutans*
446	cation efflux system CZCD	11	sp|P13512|	*Alcaligenes eutrophus*
447	copper transport atp-binding *nosf*	17	sp|P19844|	*Pseudomonas stutzeri*
448	mercuric reductase	14	gp|X73112|	*Pseudomonas fluorescens*
449	*n*-ferritin repressor (FRP)	48	gb|M95815|	*Oryctolagus cuniculus*
450	$Na^{(+)}$-ATPase B	17	sp|Q08637|	*Enterococcus hirae*
451	periplasmic divalent cation tolerance	14	gp|Z36905|	*Escherichia coli*
452	potassium/copper transporting ATPase	32	sp|P32113|	*Enterococcus hirae*
453	AP56	39	gp|S56599|	*Mus sp.*
454	rhodanese-like (*cysA*)	41	gp|M29612|	*Saccharopholyspora erythr.*
455	*SN*-glycerol-3-phos. transport ATP-binding (*ugpC*)	56	sp|P10907|	*Escherichia coli*
456	sulfate permease (*cysA*)	19	sp|P14788|	*Anacystis nidulans*
Other categories				
457	7-alpha-hydroxysteroid hydrogenase	8	pir|A42468|	*Eubacterium sp.*
458	acylphosphatase	13	sp|P07032|	*Gallus gallus*
459	cheA/W/Y	22	gp|U30501|	*Thermotoga maritima*

Table 1b. *Continued*

ID #	Hit identification	Probability (-E)	Accession	Organism
460	FixA (nitrogen fixation)	8	sp|P31573|	*Escherichia coli*
461	FixB (nitrogen fixation)	41	gp|M91817|	*Clostridium acetobutylic.*
462	FixC (nitrogen fixation)	30	sp|P31575|	*Escherichia coli*
463	FixX (nitrogen fixation)	33	pir|S49190|	*Azotobacter vinelandii*
464	homocitrate synthase (*nifV*, nitrogen fixation)	39	sp|P05342|	*Azotobacter vinelandii*
465	phosphomannomutase	15	gp|M60873|	*Pseudomonas aeruginosa*
466	ABC transporter (*pstC*-1)	14	gp|Z47982|	*Mycobacterium tuberculos.*
467	*sec61*	12	gp|X77805|	*Pyrenomonas salina*
468	threonine dehydratase biosynthetic	6	sp|Q02145|	*Lactococcus lactis*
469	vacuolar H^+ phosphatase	49	gp|M81892|	*Arabidopsis thaliana*
470	multidrug resistance	18	gp|Z49126|	*Caenorhabditis elegans*
471	phenylacrylic acid decarboxylase	38	gp|L09263|	*Saccharomyces cerevisiae*
472	daunorubicin-doxorubicin polyketide synth.	16	gp|L35560|	*Streptomyces peucetius*
473	pyrroline-5-carboxylate reductase	6	pir|JC2078|	*Thermus aquaticus*
474	amylomaltase	43	gp|J01796|	*Streptococcus pneumoniae*

quence contigs. The distance between these sequence markers is 11.5 Kb on average, while the average distance between the 200 sequence gaps will be 8.5 Kb, or in other words the average contig size will be 8.5 Kb. Thus, a randomly placed contig is likely to fall across a sequence marker which consequently maps that contig to particular fosmid clones. The few contigs which do not cross a sequence marker will be mapped by hybridization to the 96 fosmid clones.

Discussion

The mapping and sequencing of the *Pyrobaculum aerophilum* genome will provide valuable resources to researchers in many disciplines. We report here a list of 474 putative genes and a genome map. Access to these gene sequences now will provide researchers with a valuable tool for a multitude of studies.

A few whole genome sequencing projects have already been completed (Bult et al. 1996; Fleischmann et al. 1995; Fraser et al. 1995) and several others are in progress (Burland et al. 1995; Charlebois et al. 1996; Coulson 1996; Levy 1994). These efforts are concentrated on key organisms dispersed throughout the domains of life and will be important in elucidating the evolutionary path of modern organisms.

Pyrobaculum aerophilum is a member of the crenarchaea, which were thought to be a branch of the archaea but whose position in the phylogenetic tree of life has been a subject of controversy. It is uncertain whether the archaea are truly monophyletic (sharing a common ancestor distinct from other groups) or whether the crenarchaea in fact only share a distinct common ancestor with the eukaryotes and not with the rest of the archaea (Rivera and Lake 1992). Tracing the evolution of a few single genes is not sufficient to resolve these questions since many factors complicate the analyses including gene duplications, exon shuffling, and horizontal transfers (i.e., transfer of a gene from one organism to another). Accumulating large amounts of sequence data from representative organisms is the most efficient way to unscramble these evolutionary puzzles.

Complete genome sequences from representative organisms also provide researchers with the information needed to quickly isolate their gene of interest from virtually any organism. Furthermore, researchers will frequently be able to trace the evolutionary history of their favorite gene directly from the sequence databases. Quickly determining the absence of a particular gene in an organism is in itself a powerful tool. Currently, this takes weeks of hybridization experiments and still leaves the researcher uncertain of the gene's absence.

Although a detailed analysis comparing this sequence to other completed sequences will wait for the final edited sequence, there are already some surprises. One example is in the area of repair systems. We have identified tags to members of many of the major repair systems found in *E.coli* and the human, including those involved in excision repair, repair of oxidative damage, and methylated bases. *Pyrobaculum aerophilum* is a particularly interesting organism for the study of repair systems because of its remarkable similarity, at the amino acid sequence level, to eukaryotes. For instance, we have found homologs to genes involved in human repair disorders, such as xeroderma pigmentosum. Interestingly, despite the fact that we have sequenced more than 95% of the *Pyrobaculum* genome, we have not detected any genes involved in either the mismatch repair systems or deamination of cytosines. The latter is expected to be enhanced at high temperatures. The reported complete genome sequence of *Methanococcus jannaschii* (Bult et al. 1996) also fails to list genes involved in either of these repair pathways. Therefore, either the Archaea have a different way of repairing replication mismatches, or we must entertain the fascinating possibility that they have high mutation rates in vivo.

Materials and methods

Laboratory procedures were performed as described in the laboratory manual (Sambrook et al. 1989) unless otherwise mentioned.

Genomic DNA extraction

Pyrobaculum aerophilum (type strain IM2; DSM 7523) was grown anaerobically and pelleted at Karl Stetter's laboratory, University of Regensburg, Germany, under the previously published conditions (Völkl et al. 1993). Approximately 0.5 g (wet weight) of the cell pellet was chilled with liquid nitrogen, ground with a mortar and pestle, and resuspended in 500 ml TE buffer. Lysozyme (5 ml, to give 50 mg/ml) was added to the cell suspension, followed by the addition of 30 ml 20% w/v SDS, and the lysis mixture was incubated at 37°C for 1 h. Lysed cells were phenol extracted once, ethanol precipitated, then resuspended in TE at a final concentration of 0.5 mg/ml.

Fosmid library construction

For the construction of the genomic fosmid library, 10 µg of genomic DNA was resuspended in 500 µl *Sau*3AI buffer by gently swirling on a rotating platform for 3 h at 4°C followed by 10 min at 37°C. The DNA was restriction digested by adding 0.1 unit *Sau*3AI enzyme (New England Biolabs, Beverly, MA, USA) and incubating at 37°C for 15 minutes, followed by addition of another 0.1 U *Sau*3AI enzyme with additional incubation at 37°C for 15 minutes. Then, the DNA was extracted by chloroform, ethanol precipitated, and resuspended in 20 µl TE. Analysis on an agarose gel showed that most fragments were within 20–40 Kb. From this DNA stock a 3-µl aliquot was taken and dephosphorylated in a 25-µl reaction with 2 units HK-Phosphatase (Epicentre Technologies, Madison, WI, USA) at 30°C for 1 h, followed by heat inactivation of the phosphatase at 65°C for 15 min. Fosmid arms were prepared as described previously (Kim et al. 1992). Ligations were done with a large molar excess of vector DNA (50–100×) to minimize the probability of chimera formation. *Bam*HI digested fosmid arms (1 µg) were ligated with approximately 15 ng of the genomic DNA in a 15-µl reaction with 1 unit of T4 DNA ligase (Epicentre Technologies). The ligation product thus generated (3 µl) was packaged into phage particles using GigaPack XL in-vitro packaging extract (Stratagene, La Jolla, CA, USA) as instructed by the vendor. The packaged particles were suspended in 500 µl SM buffer. *E coli* DH10B cells were grown in LB + 0.2% maltose + 10 mM $MgSO_4$ to an optical density at 600 nm (OD_{600}) of 0.5–0.9, centrifuged at 4°C, and resuspended in 10 mM $MgSO_4$ to a final concentration of OD_{600} = 1.0. Undiluted packaging lysate (25 µl) was mixed with 25 µl resuspended cells and kept at room temperature for 30 min. LB (200 µl at 37°C) was then added and kept at 37°C for 1 h with gentle shaking every 15 min. Cells were centrifuged and resuspended in 50 µl LB, which was subsequently plated on LB plates containing 10 µg/ml chloramphenicol. Emerging colonies from the plates were inoculated into eight 96-well microtiter plates containing LB + 10% glycerol, which were then grown at 37°C overnight and kept frozen at –70°C until use.

Construction of pUC18 library

Genomic DNA (10 µg) was diluted in 300 µl TE and sheared by sonicating with a Branson cell disrupter 200 sonicator at setting 3.5 for 2 s. This sheared DNA (180 µl) was mixed with 180 µl 2× *Bal*31 buffer and kept at 30°C for 15 min. Then, 0.015 units of *Bal*31 enzyme (New England Biolabs) was added and the mixture incubated at 30°C for 10 min. The mixture was then diluted to 660 µl with TE, extracted with phenol: chloroform and chloroform, and precipitated with ethanol. One fifth of the remaining DNA was run on a preparative 1.2% SeaKem GTG (FMC Bioproducts, Rockland, ME, USA) agarose gel. After brief staining with ethidium bromide, fragments of 1.5–2 Kb were excised and recovered by QiaexII (Qiagen, Chatsworth, CA, USA) binding following manufacturers instructions. Elution was in 20 µl 10 mM Tris pH 8.5. This entire volume was used for subcloning into the *Sma*I site of pUC18 (Ready-To-Go subcloning kit, Pharmacia Biotech, Piscataway, NJ, USA). These subclones were randomly sequenced from both ends (unpublished) using fluorescent multiplex automated sequencers (ABI 373, Applied Biotechnology, Foster City, CA, USA), and the public databases were searched with these sequences by BLAST (Altschul et al. 1990; States et al. 1991) to find similarities to previously known sequences.

Generation of sequence marker probes by PCR amplification

Typically, 0.005 pmoles of template DNA, 20 pmoles of each primer, and 1 unit of TaqI polymerase (Gibco BRL, Gaithersburg, MD, USA) were combined in a 25-µl reaction. Polymerase chain reaction (PCR) products were isolated by running the PCR reactions on 1% SeaPlaque low melting point agarose gel (FMC Bioproducts, Rockland, ME, USA) in 1× TAE and excising DNA bands. The DNA fragments, in low melting point agarose, were radiochemically labeled using the DecaPrime kit (Ambion, Austin, TX, USA) as instructed by the vendor except that each reaction was scaled down to half the instructed values.

Generation of riboprobes

Fosmid clone templates were prepared from 3 ml LB cultures using an AutoGen 740 automated miniprep machine (Integrated Separation Systems, Natick, MA, USA). Half of the DNA isolated from a 3-ml culture was used in a 100-µl digestion reaction with 5 units *Hin*cII restriction enzyme. Digested products were ethanol precipitated and resuspended in 5 µl TE or water. This concentrated template (1.5 µl) was used in the standard transcription reaction de-

scribed in the RiboScribe kit protocol (Epicentre Technologies), except that all volumes were scaled down to half, giving a final reaction size of 10 µl. After the 2-h labeling period the entire contents of the reactions were used for hybridization.

Hybridization

Fosmid library colony membranes were prepared by gridding the entire library, represented by eight 96-well microtiter plates, at high density onto single 8 cm × 12 cm nylon membranes using a Biomek 1000 laboratory workstation (Beckman Instruments, Fullerton, CA, USA). Labeled probes (PCR products or riboprobes, as described in the previous section) were transferred directly into hybridization chambers without purification. Colony hybridizations were incubated overnight in 1 M NaCl, 50 mM Tris-HCl pH 8.0, 5 mM EDTA pH 8.0, 10% PEG 8000, at 65°C. The membranes were washed in 2× standard saline citrate (SSC), 0.1% sodium dodecyl sulfate (SDS) at 65°C for 30 min, followed by another wash in 0.2× SSC, 0.1% SDS at 65°C for 30 min. The membranes were exposed to X-ray film from 3 h to overnight.

Fingerprinting

Restriction fingerprint analysis of fosmid clones was done essentially according to the previously described modified protocol that was adapted from the original cosmid fingerprinting procedure (Coulson et al. 1986; Sulston et al. 1988). The fingerprint gels were digitized using a Phosphorimager (Molecular Dynamics, Foster City, CA, USA) and analyzed using the contigc program (http://www.sanger.ac.uk).

Assembly of fosmid contig

Ordering and alignment of contigs was performed manually and based primarily on the results of hybridization of the markers to the fosmid clones. Two clones are considered overlapping if they are hit by one or more common probes. Ambiguities in the overlapping relations were resolved by fingerprint analysis or additional hybridization with riboprobes that are specific to clone ends.

Acknowledgments We thank Dr. Phil Green and Dr. David Gordon for use of their Phred and Phrap programs, Dr. David Mathog for helping us use the GCG DNA sequence analysis programs and Jack Gerson for critical reading of the manuscript. This work was supported by funding from the United States Department of Energy to M.I.S. and from the Office of Naval Research to J.H.M. (N00014-95-1-0938).

References

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

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 JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NSM, Weidman JF, Fuhrmann JL, Venter JC et al. (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii* Science 273:1058–1073

Burland V, Plunkett GR, Sofia HJ, Daniels DL, Blattner FR (1995) Analysis of the *Escherichia coli* genome VI: DNA sequence of the region from 92.8 through 100 minutes. Nucleic Acids Res 23:2105–2119

Charlebois RL, Gaasterland T, Ragan MA, Doolittle WF, Sensen CW (1996) The *Sulfolobus solfataricus* P2 genome project. FEBS Lett 389:88–91

Coulson A (1996) The *Caenorhabditis elegans* genome project. *C. elegans* Genome Consortium. Biochem Soc Trans 24:289–291

Coulson A, Sulston J, Brenner S, Karn J (1986) Toward a physical map of the genome of the nematode *Caenorhapditis elegans*. Proc Natl Acad Sci USA 83:7821–7825

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd Science 269:496–512

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM et al. (1995) The minimal gene complement of *Mycoplasma genitalium* Science 270:397–403

Kim UJ, Shizuya H, de Jong PJ, Birren B, Simon MI (1992) Stable propagation of cosmid sized human DNA inserts in an F factor based vector. Nucleic Acids Res 20:1083–1085

Levy J (1994) Sequencing the yeast genome: an international achievement. Yeast 10:1689–1706

Riley M (1993) Microbiol Rev 57:862

Rivera MC, Lake JA (1992) Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. Science 257:74–76

Robb FT, Place AR (eds) (1995) Archaea : A laboratory manual. Thermophiles. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 3 vol

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning : a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 3 vol

States DJ, Gish W, Altschul SF (1991) Improved sensitivity of nucleic acid databases searches using application-specific scoring matrices. Methods 3:66–70

Stetter KO (1992) Life at the upper temperature border. In: Tran Than Van JK, Mounolou JC, Schneider J, McKay C (eds) Colloque Interdisciplinaire du Comite National de la Recherche Scientifique, Frontiers of Life

Sulston J, Mallett F, Staden R, Durbin R, Horsnell T, Coulson A (1988) Software for genome mapping by fingerprinting techniques. Comput Appl Biosci 4:125–32

Völkl P, Huber R, Drobner E, Rachel R, Burggraf S, Trincone A, Stetter KO (1993) *Pyrobaculum aerophilum* sp. nov., a novel nitrate-reducing hyperthermophilic archaeum. Appl Environ Microbiol 59:2918–2926

Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–271