DNA Research **8**, 123–140 (2001)

Complete Genome Sequence of an Aerobic Thermoacidophilic Crenarchaeon, *Sulfolobus tokodaii* strain7

Yutaka Kawarabayasi,[1,2,*] Yumi Hino,[1] Hiroshi Horikawa,[1] Koji Jin-no,[1] Mikio Takahashi,[1] Mitsuo Sekine,[1] Sin-ichi Baba,[1] Akiho Ankai,[1] Hiroki Kosugi,[1] Akira Hosoyama,[1] Shigehiro Fukui,[1] Yoshimi Nagai,[1] Keiko Nishijima,[1] Rie Otsuka,[1] Hidekazu Nakazawa,[1] Minako Takamiya,[1] Yumiko Kato,[1] Takio Yoshizawa,[1] Toshihiro Tanaka,[1] Yutaka Kudoh,[1] Jun Yamazaki,[1] Norihiro Kushida,[1]Akio Oguchi,[1] Ken-ichi Aoki,[1] Sayaka Masuda,[1] Masao Yanagii,[1] Masami Nishimura,[1] Akihiko Yamagishi,[3] Tairo Oshima,[3] and Hisasi Kikuchi[1]

National Institute of Technology and Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan,[1] *National Institute of Advanced Industrial Science and Technology, 1-1 Higashi, Tsukuba, Ibaraki 305-0046, Japan,*[2] *and Tokyo University of Pharmacy and Life Science, Horinouchi, Hachioji, Tokyo 192-0392, Japan*[3]

(Received 23 July 2001; revised 2 August 2001)

Abstract

The complete genomic sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain7 which optimally grows at 80°C, at low pH, and under aerobic conditions, has been determined by the whole genome shotgun method with slight modifications. The genomic size was 2,694,756 bp long and the G+C content was 32.8%. The following RNA-coding genes were identified: a single 16S–23S rRNA cluster, one 5S rRNA gene and 46 tRNA genes (including 24 intron-containing tRNA genes). The repetitive sequences identified were SR-type repetitive sequences, long dispersed-type repetitive sequences and Tn-like repetitive elements. The genome contained 2826 potential protein-coding regions (open reading frames, ORFs). By similarity search against public databases, 911 (32.2%) ORFs were related to functional assigned genes, 921 (32.6%) were related to conserved ORFs of unknown function, 145 (5.1%) contained some motifs, and remaining 849 (30.0%) did not show any significant similarity to the registered sequences. The ORFs with functional assignments included the candidate genes involved in sulfide metabolism, the TCA cycle and the respiratory chain. Sequence comparison provided evidence suggesting the integration of plasmid, rearrangement of genomic structure, and duplication of genomic regions that may be responsible for the larger genomic size of the *S. tokodaii* strain7 genome. The genome contained eukaryote-type genes which were not identified in other archaea and lacked the CCA sequence in the tRNA genes. The result suggests that this strain is closer to eukaryotes among the archaea strains so far sequenced.

The data presented in this paper are also available on the internet homepage (http://www.bio.nite.go.jp/ E-home/genome_list-e.html/).

Key words: aerobic thermoacidophilic crenarchaeon; genome sequencing; whole genome shotgun method; comparative analysis; plasmid

1. Introduction

Among the complete genome sequences of thermophilic archaea reported,[1–5] two genomes were sequenced by our group: *Pyrococcus horikoshii* OT3[1] and *Aeropyrum pernix* K1.[2] These two species are hyper-thermophilic and optimally grow at over 95°C. While *P. horikoshii* OT3 is an anaerobic euryarchaeon, *A. pernix* K1 is an aerobic crenarchaeon. The entire genomic sequence of *A. pernix* K1 was the first and the only complete genomic data of crenarchaeota. It is also the only known genomic sequence of a strictly aerobic hyperthermophile. To obtain more information about the genomic sequence data of *A. pernix* K1, it is useful to compare the data with data from closely related species.

We therefore selected *Sulfolobus tokodaii* strain7, a species of genus *Sulfolobus* in crenarchaeon for sequencing. This strain was isolated from Beppu hotsprings

Communicated by Michio Oishi

* To whom correspondence should be addressed. Tel. +81-3-3481-8951, Fax. +81-3-3481-1962, E-mail: kyutaka@nite.go.jp

† The entire genome sequence has been deposited in DDBJ/GenBank/EMBL databases under accession numbers AP000981–AP000990.

in Kyushu, Japan in 1983.[6] This species grows under aerobic conditions, as does *A. pernix* K1. The optimal growth temperature of *S. tokodaii* strain7 is 80°C and the optimal pH is between 2 and 3. The genomic data of this strain is expected to provide the information on not only the thermostability of proteins but also the characteristics of cells living in an acidic environment. The genomic size of this strain is approximately 2.7 Mbp,[7] 1 Mbp larger than the other two species previously determined by our group: *P. horikoshii* OT3 and *A. pernix* K1. In addition to above genomic features, *S. tokodaii* strain7 has no extra-chromosomal genetic unit[7] and is able to convert hydrogen sulfide to sulfate (personal communication from A. Yamagishi and T. Oshima).

To determine the entire sequence of this genome, the shotgun libraries with short and long inserts were constructed both from the entire genomic DNA and from the restriction fragments of genomic DNA digested with *Bss*H II. The raw sequencing data from shotgun clones were assembled using the software PhredPhrap and Consed.[8–10] The remaining sequencing gaps were filled by walking of long insert shotgun clones.

2. Materials and Methods

2.1. Bacterial strains and genomic DNA

The strain, *S. tokodaii* strain7 deposited in the Japan Collection of Microorganisms (JCM number 10545) was used for genome sequencing in this study. These cells were inoculated into 100 ml of the Sulfolobus culture medium[11] which was prepared in 500-ml Erlenmeyer flasks, and cultured at 80°C with vigorous shaking. The genomic DNA was isolated principally based on the method of Yamagishi and Oshima.[12] *Escherichia coli* DH10B was used for the preparation of plasmid clones.

2.2. Construction of shotgun clones

The *S. tokodaii* strain7 genomic DNA was sonicated for 5, 10, and 20 sec at the L position of a sonicator Biorupter (Cosmo Bio, Tokyo, Japan), followed by size-fractionation by agarose gel electrophoresis. The fractions from 0.8 to 1.2 kb and from 2.0 to 2.5 kb were independently cloned into the *Hinc* II site of pUC118 (Takara Shuzo, Kyoto, Japan). They are referred to as short-fragment and long-fragment shotgun libraries, respectively.

For the construction of the shotgun libraries from restriction fragments, the genomic DNA was prepared in the agarose-plug as described by Kondo et al.,[7] and after complete digestion of DNA in the plugs with *Bss*H II, the digests were resolved in 0.4% low-melting-point agarose gel by electrophoresis. The bands were cut out and the gel was dissolved by incubation at 50°C with Agarase (FMC, Rockland, ME, USA) overnight. The resultant solution containing DNA fragments was sonicated, and the short-fragment shotgun library was prepared as above, except for *Bss*H II fragment A (945 kb) from which both short- and long-fragment shotgun libraries were constructed.

2.3. Sequencing

Plasmid DNA was prepared by an Autogen 740 automatic DNA preparation system (Autogen, Framington, MA, USA). The sequencing reaction was performed using two kinds of cycle sequencing kits, a dye-primer cycle sequencing kit and a dye-terminator cycle sequencing kit. The sequence data were detected by ABI-DNA sequencers (377XL; Perkin-Elmer ABI, Foster City, CA, USA). Also the dye-terminator cycle sequencing kit was used for filling the sequencing gaps. The long-fragment shotgun clones with inserts covering gap regions were used for filling of gaps by walking with synthesized primers.

The raw sequence data were first analyzed with the software Phred.[7] After the elimination of contaminated sequences derived from *E. coli* or vector DNAs, the treated data were assembled into contigs using the software Phrap[8] and Consed.[9] The assembled sequences were split into 30-kbp segments and the sequence in each segment was re-assembled and edited by Sequencher (Gene Codes, Ann Arbor, MI, USA).

2.4. Computational analysis

The criteria used for the assignment of potential protein-coding regions were similar to those used in the previous paper.[1] However, the open reading frames (ORFs) which had neither similarity nor motif sequences completely included within longer ORFs were not assigned as potential protein-coding regions in this work. Similarity search of the assigned ORFs was mainly performed by the Smith-Waterman algolithm.[13] The databases used for similarity search were GenBank release 109, EMBL release 56.0, Swiss-Prot release 38.0, PIR release 62.0, and Owl release 31.4.

3. Results and Discussion

3.1. Determination of entire genome sequence

The physical map of circular genomic DNA of *S. tokodaii* strain7 has been reported by Kondo et al.[7] According to their observation that the genomic DNA generates six fragments upon digestion with *Bss*H II, these six fragments were used for the construction of fragment-specific shotgun libraries.

The entire genome sequence was determined by the whole genome shotgun method[14] with slight modifications. The two kinds of shotgun libraries, one with approximately 1-kb inserts (short-fragment shotgun library) and the other with approximately 2.2-kb inserts

(long-fragment shotgun library), were prepared from the genomic DNA and from *Bss*H II fragment A. From the other *Bss*H II fragments, only short-fragment shotgun libraries were constructed. The clones from the short-fragment shotgun libraries were sequenced from one end whereas the clones from the long-fragment shotgun libraries were sequenced from both ends. A total of 61,000 readings of raw sequencing data were accumulated and assembled using PhredPhrap.[8–10]

As indicated in the following section, a large number of long dispersed repetitive sequences were identified from the consensus sequences of primary contigs generated by PhredPhrap. To avoid the confusion of assembling by these repetitive sequences, the nucleotide sequence of these repetitive units were added in the cross-match database of PhredPhrap and were not used for contig assembling. The resulting contigs were separated into 30-kbp regions and the raw data contained in each region was re-assembled and edited by Sequencher.

To fill the remaining sequencing gaps, the clones carrying the end sequences of contigs were selected from the long-fragment shotgun libraries and sequenced by walking with synthesized primers. The nucleotide sequences of long dispersed repetitive units were also determined by walking the long-fragment shotgun clones which included each repetitive unit with synthesized primers. All the sequences were determined by co-incidence of sequences of at least two clones per base.

To confirm the authenticity of the genomic sequence constructed, the restriction pattern of each 15-kb fragment directly amplified from the genomic DNA by long PCR were compared with those deduced from the sequence data.

The total length of the genome finally confirmed was 2,694,756 bp. The nucleotide position was numbered from the one end of the *Bss*H II restriction site located on the *Not* I B and *Rsr* II F fragments in the physical map shown by Kondo et al.[7]

Distribution of ACGT along the strand of the entire genome was 33.4% A, 16.3% C, 16.5% G, and 33.8% T, resulting in a G + C content of 32.8%. Through the processes of genomic sequencing, no evidence was obtained for the presence of an extra-chromosomal unit.

3.2. RNA coding genes

The entire genomic sequence was subjected to similarity search against the rRNA sequences registered in the databases. It was found that the *S. tokodaii* strain7 genome contained a single 16S–23S rRNA cluster and one 5S rRNA. This organization is similar to those of *P. horikoshii* OT3[1] and *A. pernix* K1.[2]

Forty-five tRNA genes were identified by searching with tRNA scan,[15] and one tRNA gene for Leu (GAG) was identified by similarity search. The species of tRNA genes identified are shown in Table 1. Forty-one of these genes were discretely mapped, while the remaining six tRNA genes were mapped as clusters of two tRNA genes. As noted in other microorganisms,[1,2,16] no tRNA genes containing A at the first position of the anticodon were identified. Similar to the archaea genomes already sequenced, the Met-tRNA gene occurred in triplicate. However, the sequences had no similarity to one another, suggesting that the different Met-tRNA species may be used for the translation initiation of different classes of genes.

Twenty-four tRNA genes were found to contain 11- to 57-bp long introns. These genes containing introns and the organization of the introns in each gene are summarized in Table 2. In all of the intron-containing tRNA genes except for the tRNA-Leu (GAG), tRNA-Glu (UUC) and tRNA-Glu (CUC) genes, the introns were identified 1 bp downstream from the 3′ end of the anticodon triplet. The intron in the tRNA-Leu (GAG) gene had been inserted 4-bp upstream from the 5′ end of anticodon region. The 17 bp-long intron portion was assigned within the D-loop of two tRNA-Glu genes. The Insertion position of these introns was similar to that of the tRNA-Thr (UGU) gene identified in the *A. pernix* K1 genome.[2]

In contrast to tRNA genes in other archaea, most of the tRNA genes identified did not contain the 3′ terminal CCA sequence. On the other hand, tRNA nucleotidyltransferase,[17] that catalyzed the addition of the CCA sequence to tRNA transcripts, was found on this genome (ST0952). In bacteria, the 3′ CCA is generally encoded on the tRNA genes. The role of *E. coli* tRNA nucleotidyltransferase is the repair of tRNA 3′ ends.[18–20] In eukaryotes, where CCA is rarely encoded by tRNA genes, tRNA nucleotidyltransferase is essential.[21] Although the role of tRNA nucleotidyltransferase is unclear in *S. tokodaii* strain7, the structure of tRNA genes and the presence of a CCA-adding enzyme are similar to eukaryotic cells.

The genes for all tRNA synthetases have been identified in the assigned ORFs, except for those for $tRNA^{Gln}$ and $tRNA^{Asn}$. Instead, similar to *Bacillus subtilis*[22] and *Deinococcus radiodurans*,[23] the three subunits of Glu-$tRNA^{Gln}$ amidotransferase (ST1283, ST0282, and STS140), which were necessary for transamidation of Glu-$tRNA^{Gln}$ and Asp-$tRNA^{Asn}$, were identified.

3.3. Repetitive sequences identified

Three types of repeating units including Tn-like elements were found on the genome of *S. tokodaii* strain7.

The first type, named type A, is comprised of repetitive sequences composed of LR and SR segments, and is divided into two subtypes based on the sequences of SR segments. The subtype A-I repetitive unit contained 24 bp-long well-conserved SR segments (GMTAATCCWTAATGGAATTGAAAG) and a

Table 1. Summary of assigned tRNA genes.

UUU (Phe)		UCU (Ser)		UAU (Tyr)		UGU (Cys)	
UUC (Phe)	◎	UCC (Ser)	○	UAC (Tyr)	◎	UGC (Cys)	◎
UUA (Leu)	○	UCA (Ser)	◎	UAA (End)		UGA (End)	
UUG (Leu)	◎	UCG (Ser)	◎	UAG (End)		UGG (Trp)	◎
CUU (Leu)		CCU (Pro)		CAU (His)		CGU (Arg)	
CUC (Leu)	◎	CCC (Pro)	◎	CAC (His)	○	CGC (Arg)	◎
CUA (Leu)	◎	CCA (Pro)	○	CAA (Gln)	○	CGA (Arg)	○
CUG (Leu)	◎	CCG (Pro)	○	CAG (Gln)	○	CGG (Arg)	○
AUU (Ile)		ACU (Thr)		AAU (Asn)		AGU (Ser)	
AUC (Ile)	○	ACC (Thr)	○	AAC (Asn)	○	AGC (Ser)	○
AUA (Ile)		ACA (Thr)	◎	AAA (Lys)	◎	AGA (Arg)	◎
AUG (Met)	◎ *	ACG (Thr)	◎	AAG (Lys)	◎	AGG (Arg)	◎
GUU (Val)		GCU (Ala)		GAU (Asp)		GGU (Gly)	
GUC (Val)	○	GCC (Ala)	○	GAC (Asp)	◎	GGC (Gly)	○
GUA (Val)	○	GCA (Ala)	○	GAA (Glu)	◎	GGA (Gly)	○
GUG (Val)	○	GCG (Ala)	○	GAG (Glu)	◎	GGG (Gly)	○

tRNA genes assigned are indicated by ○ and those with introns by ◎
*; All three $tRNA^{Met}$ possessed the introns

Table 2. The size and position of introns in the intron-containing tRNA genes.

tRNA-ID	tRNA species	anticodon	gene size (nt)	Intron position* start	end	Intron size (nt)	mature tRNA (nt)
STtRNA02	tRNA-Glu	TTC	91	**22****	38	17	74
STtRNA03	tRNA-Lys	TTT	99	39	63	25	74
STtRNA04	tRNA-Lys	CTT	101	39	65	27	74
STtRNA06	tRNA-Cys	GCA	88	38	51	14	74
STtRNA08	tRNA-Met	CAT	98	39	62	24	74
STtRNA11	tRNA-Thr	CGT	98	39	62	24	74
STtRNA15	tRNA-Met	CAT	93	39	57	19	74
STtRNA16	tRNA-Tyr	GTA	93	39	57	19	74
STtRNA18	tRNA-Ser	CGA	111	39	64	26	85
STtRNA19	tRNA-Leu	CAG	101	41	56	16	85
STtRNA20	tRNA-Glu	CTC	91	**22****	38	17	74
STtRNA21	tRNA-Ser	TGA	95	39	49	11	84
STtRNA23	tRNA-Phe	GAA	91	39	55	17	74
STtRNA25	tRNA-Leu	TAG	101	41	56	16	85
STtRNA26	tRNA-Leu	GAG	104	**34****	52	19	85
STtRNA27	tRNA-Pro	GGG	94	41	58	18	76
STtRNA29	tRNA-Ile	GAT	90	39	54	16	74
STtRNA31	tRNA-Thr	TGT	92	40	56	17	75
STtRNA32	tRNA-Arg	GCG	88	40	52	13	75
STtRNA33	tRNA-Leu	CAA	97	40	52	13	84
STtRNA34	tRNA-Met	CAT	87	40	51	12	75
STtRNA36	tRNA-Arg	CCT	99	40	63	24	75
STtRNA40	tRNA-Trp	CCA	131	39	95	**57****	74
STtRNA46	tRNA-Arg	TCT	91	40	55	16	75

*: Nucleotide positions from the 5′ end of tRNA genes.
**: These numerals indicated the unusual position or size of introns.

variable 36- to 169-bp sequence, followed by 223 bp-, 244 bp-, or 310 bp-long LR segments. This type of repetitive unit was present at three different positions with 48, 104, and 119 repeats of SR segments (coordinates: 2,667,399–2,664,097; 2,678,286–2,671,319; and 2,693,179–2,685,053).

The subtype A-II repetitive unit was composed of 73 and 113 repeats of 25-bp-long well-conserved SR segments (GATGAATCCCAAAAGGAATTGAAAG) and a variable 33- to 45-bp spacer sequence, followed by 228-bp-long LR segments. These repetitive units were identified at two different positions (coordinates: 30,837–25,997 and 32,475–39,835).

The second type of repetitive sequences was Tn-like elements. This was extracted by analysis of ORFs that are present in repetitive sequences. Out of nine ORFs

Table 3. The lengths and coordinations of the Tn-like elements and dispersed repetitive sequences.

type of repetitive sequence	length (nt)	coordinates			direction
Tn-like element	**1779**	57900	-	59678	<
		660704	-	662475	<
		1034923	-	1036702	>
		1077993	-	1079772	>
		1157568	-	1159339	>
		1907731	-	1909502	>
		2562560	-	2564337	<
truncated Tn-like element	**345**	120380	-	120715	>
		715738	-	716074	>
		740655	-	740991	<
		760848	-	761184	<
		1115939	-	1116276	<
		1146090	-	1146426	>
		1623632	-	1623968	<
		1777801	-	1778137	>
		1784191	-	1784527	<
dispersed repetetive unit					
subtype I	**1459**	1309109	-	1310567	>
		1367448	-	1368906	<
subtype II	**1322**	823496	-	824817	<
		885027	-	886348	>
		1164946	-	1166267	<
		1922728	-	1924049	<
subtype III	**844**	163506	-	164343	<
		269897	-	270739	<
		833260	-	834103	<
		1164031	-	1164874	<
		2269440	-	2270271	<
subtype IV	**518**	2394814	-	2395331	>
		2396410	-	2396927	<

assigned as transposases, seven ORFs with 136 amino acid residues were located within the highly conserved 1779-bp-long sequences. Thus, these regions were assigned as a Tn-like element. However, the inverted repeats, which were generally identified at the border of Tn elements were not found. In addition to this complete Tn-like element, its truncated form was also identified. The end region of each truncated element was identical to the end sequence of the Tn-like element. The size of this element was 352 bp long and the transposase was not found in this truncated element. The position and direction of these Tn-like repetitive elements and truncated form of the Tn-like element are summarized in Table 3.

The third type of repetitive units comprised four subtypes of different repeating units from 518 bp to 1459 bp, and appeared dispersed along the entire genome. The length and coordinate of these four different repetitive sequences are summarized in Table 3. These long repetitive sequences often interfered with the assembly of raw sequencing data, but the problem could be minimized by registration of repetitive sequences into a cross-match database in PhredPhrap. The biological significance of these repeats, except for the Tn-like element, is not known.

3.4. Assignment and similarity search of potential protein-coding regions

The assignment of potential protein-coding regions was performed as the previous paper[1] with some modifications. The ORFs which consist of longer than 100 codons starting with ATG or GTG were designated by a two-letter code (ST) plus a four-digit number indicating the ORF position. The ORFs which were entirely included within longer ORF on either strand and showed neither similarity nor motif sequences were not assigned as potential protein-coding regions. There were a total of 2558 ST-class ORFs.

Subsequently, shorter ORFs consisting of 50 to 99 codons were extracted from the regions where no ST-class ORF was assigned plus a neighboring 150-bp region overlapping the next ST-class ORFs. Among these short ORFs, those which possessed significant similarity to either the registered sequences in databases or to protein motifs were taken as potential protein-coding regions and named with a three-letter code (STS) plus a three-digit number indicating the ORF position. The organization of all ORFs assigned among the entire genome is shown in Fig. 1.

Thus, a total of 2826 ORFs were assigned along the entire genomic sequence (ST-class ORFs = 2558, STS-class ORFs = 268). The average size of the ORFs was

Table 4. Summary of functional classification of ORFs.

functional categories	number of ORFs
Amino acid biosynthesis	77
Purines, pyrimidines, nucleosides, and nucleotides	55
Fatty acid and phospholipid metabolism	41
Biosynthesis of cofactors, prosthetic groups, and carriers	53
Central intermediary metabolism	34
Energy metabolism	184
Transport and binding proteins	92
DNA metabolism	45
Transcription	47
Protein synthesis	109
Protein fate	42
Regulatory functions	12
Cell envelope	21
Cellular processes	34
Other categories	23
Unknown function	42
Conserved hypothetical protein	921
total	1832

267 amino acid residues, and the longest one consisted of 1933 residues (ST0620). The 2826 assigned ORFs occupy 83.9% of the entire genome. The G + C content in these coding regions was 33.52%, and that in non-coding regions was 28.51%.

For the assignment of gene function, the products name in public databases with a Zscore greater than or equal to 20, or with over 30% amino acid sequence identity along the entire coding region, were taken for ST-class ORFs. The name of gene products with a Zscore greater than or equal to 12 were also used for STS-class ORFs. Among ST- and STS-class ORFs, there were 911(32.2%) ORFs with assigned function, 921 (32.6%) showed significant similarity to registered sequences with unknown function, and 145 (5.1%) contained some motifs. The remaining 849 ORFs showed no significant similarity to the sequences in public databases and protein motif sequence. The predicted products of ORFs with known functions are summarized in Table 11, in which all products predicted are classified according to functional categories.

The ORFs showing similarities to the registered genes with known functions were classified according to the functional categories and are provided in Table 4; the product names are listed in Table 11. As indicated in Table 4, this genome contained a large number of genes related to energy metabolism. This may be due to the heterotrophic feature of this microorganism, which use a large number of compounds as the energy source.

The codon usages of the proteins encoded by the 2826 ORFs are summarized in Table 5. The codons with A or U at the third position appeared to be more frequently used, probably reflecting the relatively low G+C content of this strain.

3.5. Other features

3.5.1. Genes related to respiration

Since *S. tokodaii* strain7 is aerobic, the genes involved in the TCA cycle were searched, and all of the genes in this cycle were identified on the genome. Two copies of the genes coding for citrate synthase and two subunits of 2-oxoacid:ferredoxin oxidoreductase,[24] which plays the same role in archaea as alpha-ketoglutarate dehydrogenase, were identified in the genome. This result suggests that *S. tokodaii* strain7 has a complete TCA cycle system similar to that found in mitochondria of eukaryotes.

On the other hand, some genes in the respiratory chain which participate in the production of ATP were not identified by similarity searching. In particular, cytochrome *c* was not identified in this microorganism. This feature is similar to that in *A. pernix* K1. As cytochrome *c* has an important role in electron transfer to oxygen in eukaryotes, it is likely that this microorganism uses a different molecule for the same function or possesses a different pathway. Participation of Rieske iron-sulfur protein and/or sulfocyanin in place of cytochrome *c* has been suggested in *Sulfolobus* species.

3.5.2. ORFs related to sulfide metabolism

S. tokodaii strain7 is known to oxidize hydrogen sulfide to sulfate intracellularly, and this feature has been applied for the treatment of industrial waste water (personal communication from A. Yamagishi and T. Oshima). By similarity search of the genes relating to this pathway, a total of eight ORFs were detected as genes related to sulfide metabolism (from hydrogen sulfide to sulfate). These ORFs and their products are summarized in Table 6. These enzymes seem to be enough to oxidize sulfur from hydrogen sulfide to sulfate. Although

Table 5. Codon usages of the predicted proteins coded for by the 2826 ORFs.

UUU	(Phe)	24003	31.53	UCU	(Ser)	14563	19.13	UAU	(Tyr)	27402	36.00
UUC	(Phe)	10453	13.73	UCC	(Ser)	3715	4.88	UAC	(Tyr)	9701	12.74
UUA	(Leu)	36642	48.14	UCA	(Ser)	12912	16.96	UAA	(Stop)	1532	2.01
UUG	(Leu)	7685	10.10	UCG	(Ser)	2509	3.30	UAG	(Stop)	482	0.63
CUU	(Leu)	14179	18.63	CCU	(Pro)	10889	14.31	CAU	(His)	7283	9.57
CUC	(Leu)	4243	5.57	CCC	(Pro)	3211	4.22	CAC	(His)	2636	3.46
CUA	(Leu)	12512	16.44	CCA	(Pro)	13048	17.14	CAA	(Gln)	11842	15.57
CUG	(Leu)	2916	3.83	CCG	(Pro)	2622	3.44	CAG	(Gln)	3964	5.21
AUU	(Ile)	30581	40.17	ACU	(Thr)	16723	22.00	AAU	(Asn)	26487	34.80
AUC	(Ile)	6932	9.11	ACC	(Thr)	3368	4.42	AAC	(Asn)	10515	13.81
AUA	(Ile)	37755	49.60	ACA	(Thr)	12969	17.04	AAA	(Lys)	39298	51.63
AUG	(Met)	15580	20.47	ACG	(Thr)	3096	4.07	AAG	(Lys)	21379	28.09
GUU	(Val)	22315	29.32	GCU	(Ala)	19352	25.42	GAU	(Asp)	27892	36.64
GUC	(Val)	4210	5.53	GCC	(Ala)	3927	5.16	GAC	(Asp)	7224	9.49
GUA	(Val)	22471	29.52	GCA	(Ala)	15919	20.91	GAA	(Glu)	35726	46.93
GUG	(Val)	6357	8.35	GCG	(Ala)	2969	3.90	GAG	(Glu)	17583	23.10
UGU	(Cys)	3569	4.69								
UGC	(Cys)	1395	1.83								
UGA	(Stop)	812	1.07								
UGG	(Trp)	7716	10.14								
CGU	(Arg)	1103	1.45								
CGC	(Arg)	355	0.47								
CGA	(Arg)	700	0.92								
CGG	(Arg)	271	0.36								
AGU	(Ser)	12679	16.66								
AGC	(Ser)	4328	5.69								
AGA	(Arg)	19962	26.22								
AGG	(Arg)	9091	11.94								
GGU	(Gly)	18010	23.66								
GGC	(Gly)	4105	5.39								
GGA	(Gly)	19959	26.22								
GGG	(Gly)	5585	7.34								

Numerals in the second column are the sum of codons present, and those in the third column are the frequency of occurrence per thousand in ORFs of *S. tokodaii* strain7

Table 6. List of ORFs related to H_2S oxidizing reactions.

ORF-ID	Length of ORF (a. a.)	predicted product	predicted function
ST0615	384	**sulfide dehydrogenase**	**H_2S = Sulfar + $\dot{H}_2$**
ST0971	390	**sulfide dehydrogenase** (flavocytochrome c) flavoprotein chain	**H_2S = Sulfar + H_2**
ST1010	208	**sulfite oxidase**	**$SO_3^{2-} + O_2 + H_2O = SO_4^{2-} + H_2O_2$**
ST1839	270	**thiosulfate reductase electron transport protein**	**$H_2S = S_2O_3^{2-}$**
ST2564	293	**thiosulfate sulfurtransferase**	**$S_2O_3^{2-}$ + cyanide = SO_3^{2-} + thiocyanate**
ST2566	628	**sulfite reductase**	**H_2S + 3** Oxidized Ferredoxin + **3 $H_2O = SO_3^{2-}$ + 3** Reduced Ferredoxin
ST2567	239	**phosphoadenosine phosphosulfate reductase**	**$SO_4^{2-} + H_2 = SO_3^{2-} + H_2O$**
ST2568	412	**sulfate adenylyltransferase**	**ATP + SO_4^{2-} = PPi + AMP-SO_4^{2-}**

the products of these ORFs have not yet been isolated, confirmation of their activities would provide the valuable information for sulfide metabolism in this microorganism and facilitate the improvement of this system for industrial applications.

3.5.3. Possible integration of plasmid into chromosome

It has previously been reported that *Sulfolobus* sp. NOB8H2 contains 41,229 bp long plasmid, pNOB8.[25,26] However, no extra-chromosomal genetic unit was detected in the analysis in the present study, whereas these two strains belong to the same genus. To clarify the possibility that a plasmid had been integrated into this genome, we searched ORF homologues of pNOB8. As indicated in Table 7, a total of 37 gene families, consisting of 66 ORFs, were identified in the *S. tokodaii* strain7 genome, although their organization and the number of genes included in each gene family are variable in the genome. These results imply that a plasmid from an ancestral species was integrated into the *S. tokodaii* strain7 genome, and following rearrangement and duplication of the genome resulted in the variation of the order and copy number of genes originally encoded by the plasmid. Peng et al. previously showed that pHEN7, a plasmid isolated from *S. islandicus*, was integrated into the host chromosome by integrase encoded by the plasmid itself.[27] The length of the plasmid is very short, 7.83 kb, and the integrase is highly similar to that found in SVV1 virus.[28] Two kinds of integrases were identified in the genome of *S. tokodaii* strain7, one was partially similar to that in SSV1 virus and the other was highly similar to the integrase/recombinase possessed in bacteria. The mechanism of integration in *S. tokodaii* strain7 may be different from that in *S. islandicus*.

Out of the 66 ORFs identified, the functions of only three genes were assigned, and the remaining 63 ORFs were assigned as genes with unknown function. These genes with unknown function may be indispensable for *S. tokodaii* strain7, as these genes have been maintained in the genome until now.

3.5.4. ORFs related to eukaryote-type gene families

Many genes assigned on archaeal genomes are often identified in eukaryotes or eubacteria.[29–31] During sim-

Table 7. *S. tokodaii* ORFs orthologous to genes coded in plasmid pNOB8.

genes in pNOB8	Length (a. a.)	number of identified ORFs	*S. tokodaii* ORF-ID orthologous to pNOB8 genes						
ORF01	116	**3**	ST1310	ST0251	ST1317				
ORF02	188	**1**	ST0159						
ORF03	81	**5**	ST0758	ST0236	STS156	STS229	ST1340		
ORF04	422	**2**	ST1308	ST1318					
ORF05	537	**2**	ST1307	ST2280					
ORF06	72	**1**	STS202						
ORF07	50	**3**	STS147	STS180	STS202				
ORF08	406	**3**	ST1056	ST2325	ST2008				
ORF09	87	**1**	STS029						
ORF10	1025	**1**	ST1326						
ORF17	253	**2**	ST2505	ST1338					
ORF18	94	**7**	STS156	STS035	STS149	STS148	ST1340	STS157	ST0236
ORF19	97	**4**	STS156	STS149	ST0173	STS035			
ORF20	108	**1**	ST1342						
ORF22	164	**1**	ST1344						
ORF23	69	**1**	STS162						
ORF24	72	**2**	STS163	STS072					
ORF25	92	**1**	STS164						
ORF26	101	**3**	ST1345	STS045	ST2503				
ORF27	439	**4**	ST1346	ST0313	ST2498	ST0255			
ORF28	80	**4**	STS146	STS224	ST0837	ST1890			
ORF30	248	**3**	ST1850	ST1311	ST0254				
ORF31	630	**1**	ST1312						
ORF34	86	**1**	ST1479						
ORF37	52	**1**	STS159						
ORF38	604	**1**	ST1315						
ORF39	165	**1**	ST1316						
ORF40	65	**2**	STS155	STS151					
ORF41	110	**3**	ST0251	ST1310	ST1317				
ORF42	205	**1**	ST2077						
ORF43	74	**1**	ST1343						
ORF44	93	**7**	STS033	ST1486	STS229	STS148	ST1060	ST1339	ST0766
ORF45	470	**2**	ST1308	ST1318					
ORF46	315	**1**	ST1320						
ORF48	142	**1**	ST1321						
ORF50	152	**1**	ST1316						
ORF52	81	**2**	STS154	STS152					
	total	**81**							

Table 8. *S. tokodaii* ORFs with similarity to the eukaryote-type genes.

ORF-ID	Length of ORFs (aa)	predicted product	distribution
ST0059	462	**selenium-binding protein 2**	human, mouse, plant, *C. elegans*
ST0155	353	**flug protein.**	*A. nidulans*
ST0233	288	**hypothetical protein.**	*S. cerevisiae*
ST0467	548	**DNA replication licensing factor**	human, mouse, frog, fruit fly, *S. pombe*
ST0779	583	**acylamino-acid-releasing enzyme.**	human, rat, pig, *C. elegans*
ST0940	767	**oligosaccharyl transferase stt3 subunit.**	human, mouse
ST0945	254	**hypothetical protein.**	*S. cerevisiae* , *C. elegans*
ST1110	386	**nonspecific lipid-transfer protein.**	mouse, rat
ST1257	108	**s100 calcium-binding protein a3.**	human, mouse
ST1350	363	**nonspecific lipid-transfer protein.**	human
ST1603	125	**hypothetical protein.**	*S. cerevisiae*
ST1745	565	**acylamino-acid-releasing enzyme.**	human, rat, pig, *C. elegans*
ST2082	360	**sterol-regulatory element-binding protein.**	human, *C. griseus*
ST2367	297	**serine/threonine protein phosphatase pp2a catalytic subunit.**	plant, human, mouse, fruit fly, *S. cerevisiae*

ilarity search of the *S. tokodaii* strain7 ORFs against public databases, we found that 14 ORFs show similarity to gene families that are only identified in eukaryotes. The homologues of these ORFs are not detected either in other archaea or bacteria. These ORFs and predicted gene products are given in Table 8. In addition to these eukaryotic-type genes, this strain contained 37 archaea-specific ORFs and 53 ORFs which are present both in archaea and eukaryotes. Identification of eukaryote-specific genes together with the absence of the CCA sequence in tRNA genes suggests that this strain is closer to eukaryote among archaea that have had their genomes completely sequenced.

Table 9. Homologous ORFs identified by sequence comparison among the 2826 ORFs of *S. tokodaii* strain7.

number of homologous ORFs* in one group	number of groups	total ORFs
12	1	12
11	1	11
10	1	10
9	4	36
8	6	48
7	11	77
6	14	84
5	25	125
4	56	224
3	94	282
2	281	562
total	494	1471

*: ORFs with Zscore higher than 8 in SW search and amino acid identity higher than 30% in 70% of the entire region were taken.

Table 10. List of the same ORFs identified on the *S. tokodaii* strain7 genome.

predicted function	length of ORFs (a. a.)	ORFs included in one same gene families
transposase	136	ST0043 ST2553 ST1048 ST1096 ST1904 ST0680 ST1158
insertion element protein	242	ST0142 ST1907 ST1139 ST1082 ST0671
not assigned	235	ST0143 ST0857 ST1165 ST2259 ST1981
not assigned	330	ST0847 ST0907 ST1167 ST1916 ST2230
not assigned	240	ST0162 ST1719 ST0798 ST1987
not assigned	347	ST0041 ST1098 ST1050
not assigned	108	ST0042 ST1097 ST1049
not assigned	350	ST0055 ST1957 ST1769
not assigned	381	ST0152 ST2431 ST2008
not assigned	168	ST0254 ST1311 ST1850
not assigned	434	ST0678 ST1160 ST1906
not assigned	144	ST0679 ST1159 ST1905
IS hypothetical protein	202	ST1092 ST1939 ST2495
not assigned	328	ST0046 ST0099
ferredoxin	104	ST0163 ST1175
not assigned	123	ST0856 ST1718
IS element DNA-binding protein	314	ST1122 ST1908
insertion element protein	100	ST1156 ST2015
not assigned	470	ST1304 ST1360
transposase	340	ST1779 ST2430
not assigned	353	ST1952 ST2181
sulfocyanin	220	ST2393 ST2394

3.5.5. Duplication of ORFs

By sequence comparison among the ORFs, 1471 ORFs were grouped into 494 families (Table 9). The result can be interpreted that ORFs in each group were generated by duplication of an ancestral sequence. Such homologous ORFs were present in the genome as either tandem repeats of a single ORF or the repeats of a single or a cluster of ORFs at different locations.

Within these families, we detected 22 families that are composed of ORFs which have over 95% length identity and over 95% amino acid identity. The gene families detected are summarized in Table 10. In addition to transposase and IS-element related proteins, sixteen kinds of proteins which possessed the same sequence were detected at different positions in the genome. This result suggests that rearrangement or duplication of the genome may continue to occur.

3.5.6. Other notable genes

The structure of the operon encoding H^+-ATPase (ST1435, ST1436, ST1437, ST1438, ST1439, STS172) of this strain has already been reported by Denda et al.[32–35] Data comparison indicated that the length of the delta subunit is 80 amino acids shorter than that of ST1435. Because some subunits which were present in *E. coli*[36] or *Synecococcus*[37] were not identified in the operon cloned from *S. tokodaii* strain7, Denda et al. suggested two possibilities: One is that the operon encodes all of the genes necessary for the whole ATPase complex and the other is that the operon encoding the subunits of

ATPase were split into two (or more) independent operons. Our data suggest that one more ORF (ST1434) upstream of ST1435 may be included in this operon as the subunit of the H^+-ATPase. It seems reasonable to conclude that this operon in *S. tokodaii* strain7 contains all of the genes of the whole ATPase complex.

The genes containing inteins, defined as the self-splicing portions of a polypeptide sequence,[38,39] have been identified in the genome of *P. horikoshii* OT3[1] and *A. pernix* K1.[2] However, no genes with the intein portion were identified in the *S. tokodaii* strain7 genome. Like all other aerobic organisms, the gene for superoxide dismutase[2,40–42] was also present in this organism (ST2283).

The entire genomic sequence of *S. tokodaii* strain7 is the third sequence from aerobic thermophilic crenarchaeota. The comparison of this genome with other archaea or thermophilic microorganisms may provide important information about the difference between euryarchaeota and crenarchaeota, as well as differences of the thermostability of proteins and the origin or evolution of eukaryotes. More detailed analysis of gene organization and gene structure in comparison with other archaeal genomes is under investigation.

Acknowledgements: We thank Dr. P. Green for donation of the assembling software PhredPhrap and Mr. S. Suharnan for helpful discussions. This work was supported by the Ministry of Economy, Trade and Industry.

References

1. Kawarabayasi, Y., Sawada, M., Horikawa, H. et al. 1998, Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3, *DNA Res.*, **5**, 55–76 & 147–155.
2. Kawarabayasi, Y., Hino, Y., Horikawa, H. et al. 1999, Complete genome sequence of an aerobic hyperthermophilic crenarchaeota, *Aeropyrum pernix* K1, *DNA Res.*, **6**, 84–76 & 147–155.
3. Bult, C. J., White, O., Olsen, G. J. et al. 1996, Complete genome sequence of the methanogenic archaon, *Methanococcus jannaschii*, *Science*, **273**, 1058–1073.
4. Smith, D. R., Doucette-Stamm, L. A., Deloughery, C. et al. 1997, Complete genome sequence of *Methanobacterium thermoautotrophicum* ΔH: Functional analysis and comparative genomics, *J. Bacteriol.*, **179**, 7135–7155.
5. Klenk, H. P., Clayton, R. A., Tomb, J. F. et al. 1997, The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*, *Nature*, **390**, 364–370.
6. Suzuki, T., Iwasaki, T., Uzawa, T. et al. *Sulfolobus tokodaii* sp. nov. (f. *Sulfolobus* sp. strain7), a new member of the genus *Sulfolobus* isolated from Beppu hot springs, Japan, *Extremophiles*, in press.
7. Kondo, S., Yamagishi, A., and Oshima, T. 1993, A physical map of the sulfur-dependent archaebacterium *Sulfolobus acidocaldarius* 7 chromosome, *J. Bac.*, **175**, 1532–1536.
8. Ewing, B., Hillier, L., Wendl, M. C., and Green, P. 1998, Base-calling of automated sequencer traces using Phred. I. Accuracy assessment, *Genome Res.*, **8**, 175–185.
9. Ewing, B. and Green, P. 1998, Base-calling of automated sequencer traces using Phred. II. Error probabilities, *Genome Res.*, **8**, 186–194.
10. Gordon, D., Abajian, C., and Green, P. 1998, Consed: A graphical tool for sequence finishing, *Genome Res.*, **8**, 195–202.
11. Brock, T. D., Brock, K. M., Belly, R. T., and Weiss, R. L. 1972, *Sulfolobus*: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature, *Arch. Mikrobiol.*, **84**, 54–68.
12. Yamagishi, A. and Oshima, T. 1990, Circular chromosomal DNA in the sulfur-dependent archaebacterium *Sulfolobus acidocaldarius*, *Nucleic Acids Res.*, **18**, 1133–1136.
13. Smith, T. F. and Waterman, M. S. 1981, Identification of common molecular subsequences, *J. Mol. Biol.*, **147**, 195–197.
14. Fleishmann, R. D., Adams, M. D., White, O. et al. 1995, Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd, *Science*, **269**, 496–512.
15. Lowe, T. M. and Eddy, S. R. 1997, tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence, *Nucleic Acids Res.*, **25**, 955–964.
16. Nakamura, Y. and Tabata, S. 1997, Codon-anticodon assignment and detection of codon usage trends in seven microbial genomes, *Microbial & Comparative Genomics*, **2**, 299–312.
17. Li, F., Wang, J., and Steitz, T. A. 2000, *Sulfolobus shibatae* CCA-adding enzyme forms a tetramer upon binding two tRNA molecules: A scrunching-shuttling model of CCA specificity, *J. Mol. Biol.*, **304**, 483–492.
18. Cudny, H., Lupski, J. R., Godson, G. N., and Deutscher, M. P. 1986, Cloning, sequencing, and species relatedness of the *Escherichia coli cca* gene encoding the enzyme tRNA nucleotidyltransferase, *J. Biol. Chem.*, **261**, 6444–6449.
19. Deutscher, M. P., Foulds, J., and McClain, W. H. 1974, Transfer ribonucleic acid nucleotidyl-transferase plays an essential role in the normal growth of *Escherichia coli* and in the biosynthesis of some bacteriophage T4 transfer ribonucleic acids, *J. Biol. Chem.*, **249**, 6696–6699.
20. Zhu, L. and Deutscher, P. M. 1987, tRNA nucleotidyltransferase is not essential for *Escherichia coli* viability, *EMBO J.*, **6**, 2473–2477.
21. Aebi, M., Kirchner, G., Chen, J.-Y., Vijayraghavan, U., Jacobson, A., Martin, N. C., and Abelson, J. 1990, Isolation of a temperature-sensitive mutant with altered tRNA nucleotidyltransferase and cloning of the gene encoding tRNA nucleotidyltransferase in the yeast *Saccharomyces cerevisiae*, *J. Biol. Chem.*, **265**, 16216–16220.
22. Curnow, A. W., Hong, K. W., Yuan, R., kim, S. I., Martins, O., Winkler, W., Henkin, T. M., and Söll, D. 1997, Glu-tRNA^{Gln} amidotransferase: A novel heterotropic enzyme required for correct decoding of glutamine codons during translation, *Proc. Natl. Acad. Sci.*

USA, **94**, 11819–11826.
23. Curnow, A. W., Tumbula, D. L., Pelaschier, J. T., Kin, B., and Söll, D. 1998, Glutamyl-tRNA^{Gln} amidotransferase in *Deinococcus radiodurans* may be confined to asparagine biosynthesis, *Proc. Natl. Acad. Sci. USA*, **95**, 12838–12843.
24. Zhang, Q., Iwasaki, T., Wakagi, T., and Oshima, T. 1996, 2-Oxoacid:Ferredoxin oxidoreductase from the thermoacidophilic archaeon, *Sulfolobus* sp. Strain7, *J. Biochem.*, **120**, 587–599.
25. Schleper, C., Horz, I., Janekovic, D., Murphy, J., and Zillig, W. 1995, A multicopy plasmid of the extremely thermophilic archaeon *Sulfolobus* effects its transfer to recipients by mating, *J. Bacteriol.*, **177**, 4417–4426.
26. She, Q., Phan, H., Garrett, R. A., Albers, S. V., Stedman, K. M., and Zilling, W. 1998, Genetic profile of pNOB8 from *Sulfolobus*: the first conjugative plasmid from an archaeon, *Extremophiles*, **2**, 417–425.
27. Peng, X., Holz, I., Zilling, W., Garrett, R. A., and She, Q. 2000, Evolution of the family of pRN plasmids and their integrase-mediated insertion into chromosome of the crenarchaeon *Sulfolobus solfataricus*, *J. Mol. Biol.*, **303**, 449–454.
28. Schleper, C., Kubo, K., and Zilling, W. 1992, The particle SVV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: demonstration of infectivity and of transfection with viral DNA, *Proc. Natl. Acad. Sci. USA*, **89**, 7645–7649.
29. Makarova, K. S., Aravind, L., and Koonin, E. V. 1999, A superfamily of archaeal, bacterial, and eukaryotic proteins homologous to animal transglutaminases, *Protein Sci.*, **8**, 1714–1719.
30. DiRuggiero, J., Brown, J. R., Bogert, A. P., and Robb, F. T. 1999, DNA repair systems in archaea: mementos from the last universal common ancestor?, *J. Mol. Envol.*, **49**, 474–484.
31. Jain, R., Rivera, M. C., and Lake, J. A. 1999, Horizontal gene transfer among genomes: the complexity hypothesis, *Proc. Natl. Acad. Sci. USA*, **96**, 3801–3806.
32. Denda, K., Konishi, J., Oshima, T., Date, T., and Yoshida, M. 1988, The membrane-associated ATPase from *Sulfolobus acidocaldarius* is distantly related to F1-ATPase as assessed from the primary structure of its a-subunit, *J. Biol. Chem.*, **263**, 6012–6015.
33. Denda, K., Konishi, J., Oshima, T., Date, T., and Yoshida, M. 1988, Molecular cloning of the b-subunit of a possible non-F0F1 type ATP synthase from the acidothermophilic archaebacterium, *Sulfolobus acidocaldarius*, *J. Biol. Chem.*, **263**, 17251–17254.
34. Denda, K., Konishi, J., Oshima, T., Date, T., and Yoshida, M. 1989, A gene encoding the proteolipid subunit of *Sulfolobus acidocaldarius* ATPase complex, *J. Biol. Chem.*, **264**, 7119–7121.
35. Denda, K., Konishi, J., Hajiro, K., Oshima, T., Date, T., and Yoshida, M. 1990, Structure of an ATPase operon of an acidothermophilic archaebacterium, *Sulfolobus acidocaldarius*, *J. Biol. Chem.*, **265**, 21509–21513.
36. Walker, J. E., Saraste, M., and Gay, N. J. 1984, The unc operon. Nucleotide sequence, regulation and structure of ATP-synthase, *Biochem. Biophys. Acta*, **768**, 164–200.
37. Cozens, A. L. and Walker, J. E. 1987, The organization and sequence of the genes for ATP synthase subunits un the cyanobacterium *Synechococcus* 6301. Support for an endosymbiotic origin of chloroplasts, *J. Mol. Biol.*, **194**, 359–383.
38. Hirata, R., Ohsumi, Y., Nakano, A., Kawasaki, H., Suzuki, K., and Anraku, Y. 1990, Molecular structure of a gene, VMA1, encoding the catalytic subunit of H^+-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*, *J. Biol. Chem.*, **265**, 6726–6733.
39. Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebl, M., and Stevens, T. H. 1990, Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H^+-adenosine triphosphatase, *Science*, **250**, 651–657.
40. Yamano, S., Sako, Y., Nomura, N., and Maruyama, T. 1999, A cambialistic SOD in a strictly aerobic hyperthermophilic archaeon, *Aeropyrum pernix*, *J. Biochem. (Tokyo)*, **126**, 218–225.
41. Knapp, S., Kardinahl, S., Hellgren, N., Tibbelin, G., Schafer, G., and Ladenstein, R. 1999, Refined crystal structure of a superoxide dismutase from the hyperthermophilic archaeon *Sulfolobus acidocaldarius* at 2.2 A resolution, *J. Mol. Biol.*, **285**, 689–702.
42. Kardinahl, S., Anemuller, S., and Schafer, G. 2000, The hyper-thermostable Fe-superoxide dismutase from the Archaeon *Acidianus ambivalens*: characterization, recombinant expression, crystallization and effects of metal exchange, *Biol. Chem.*, **381**, 1089–1101.

Table 11. List of ORFs showing similarities to registered genes with known function.

ORF ID	Length (a.a.)	Predicted Product	Locus	Zscore	% Identity
Amino acid biosynthesis					
Aromatic amino acid family					
ST2272	352	3-dehydroquinate synthase	Swiss_Prot:P36919	63.75	40.249
ST2277	408	3-phosphoshikimate 1-carboxyvinyltransferase	Swiss_Prot:Q57925	49.00	31.542
ST1231	345	anthranilate phosphoribosyltransferase	Swiss_Prot:P50384	96.60	54.810
ST1229	433	anthranilate synthase component I	Swiss_Prot:Q06128	105.03	63.889
ST1228	193	anthranilate synthase component II	Swiss_Prot:Q06129	90.42	76.923
ST2271	330	aroA(G) protein	Swiss_Prot:P39912	114.69	49.275
ST2274	390	chorismate synthase	Swiss_Prot:P23353	79.64	49.157
STS222	50	dehydroquinate dehydratase	OWL:NGI390803	8.20	27.778
ST1226	246	indole-3-glycerol phosphate synthase	Swiss_Prot:Q06121	93.62	58.065
ST1230	145	N-(5'-phosphoribosyl)anthranilate isomerase	Swiss_Prot:P50386	48.74	44.366
ST0294	202	P-protein (chorismate mutase: prephenate dehydratase	Swiss_Prot:P27603	36.38	34.911
ST2273	260	shikimate 5-dehydrogenase	Swiss_Prot:Q58484	45.71	36.742
ST1232	245	tryptophan synthase alpha subunit	Swiss_Prot:P50382	131.41	63.525
ST1233	422	tryptophan synthase beta subunit	Swiss_Prot:P50383	195.87	80.529
ST1567	449	tryptophan synthase beta subunit	Swiss_Prot:P50383	149.72	51.202
Aspartate family					
ST0385	338	5-methyltetrahydropteroyltriglutamate--homocystein methyltransferase	Swiss_Prot:P80877	50.35	31.928
ST0342	415	adenosylhomocysteinase	Swiss_Prot:P50252	83.54	74.820
ST1225	399	aspartate aminotransferase	Swiss_Prot:P14909	149.47	63.706
ST1242	350	aspartate-semialdehyde dehydrogenase	Swiss_Prot:Q57658	109.78	44.737
ST2171	359	aspartate-semialdehyde dehydrogenase	Swiss_Prot:Q57658	93.31	39.429
ST1241	447	aspartokinase	Swiss_Prot:Q57991	70.61	33.120
ST0506	377	cystathionine gamma-synthase	Swiss_Prot:P55217	69.49	42.059
ST2235	285	dihydrodipicolinate synthase	Swiss_Prot:Q57695	47.33	32.990
ST2479	296	dihydrodipicolinate synthase	Swiss_Prot:P06995	30.20	23.958
ST1301	460	homocitrate synthase	Swiss_Prot:Q58787	66.84	41.207
ST1519	312	homoserine dehydrogenase	Swiss_Prot:P52986	39.60	45.517
ST0505	306	homoserine kinase	Swiss_Prot:Q58504	76.12	38.403
ST1529	382	succinyl-diaminopimelate desuccinylase	Swiss_Prot:Q57899	40.29	30.530
ST0960	337	threonine synthase	Swiss_Prot:Q58860	43.88	30.712
ST1240	395	threonine synthase	Swiss_Prot:Q58860	105.42	54.103
Glutamate family					
ST0194	261	acetylglutamate kinase	Swiss_Prot:Q60382	30.67	34.483
ST0191	387	acetylornithine aminotransferase	Swiss_Prot:O30156	63.75	44.142
ST2549	374	acetylornithine deacetylase	Swiss_Prot:P23908	22.63	35.714
ST1502	446	argininosuccinate lyase	Swiss_Prot:Q58201	58.37	35.696
ST1501	390	argininosuccinate synthase	Swiss_Prot:Q60174	79.83	47.656
ST0100	137	glutamate dehydrogenase	Swiss_Prot:P80053	33.62	35.514
ST2241	422	glutamate dehydrogenase	Swiss_Prot:P80053	101.19	76.960
ST2198	635	glutamate synthase large subunit	Swiss_Prot:P55038	50.47	40.613
ST0156	483	glutamine synthetase	Swiss_Prot:P21154	48.77	37.566
ST0568	429	glutamine synthetase	Swiss_Prot:Q60182	48.99	30.562
ST1387	473	glutamine synthetase	Swiss_Prot:P23794	109.50	78.981
ST0195	363	N-acetyl-gamma-glutamyl-phosphate reductase	Swiss_Prot:Q58496	89.26	44.160
ST1245	306	ornithine carbamoyltransferase	Swiss_Prot:Q51742	58.00	49.673
ST0646	259	pyrroline-5-carboxylate reductase	Swiss_Prot:P74572	53.61	38.073
Pyruvate family					
ST0538	386	2-isopropylmalate synthase	Swiss_Prot:Q58595	73.79	53.562
ST2203	522	2-isopropylmalate synthase	Swiss_Prot:Q58787	61.70	32.813
ST0599	415	3-isopropylmalate dehydratase large subunit	Swiss_Prot:O27439	91.07	45.146
ST0600	168	3-isopropylmalate dehydratase small subunit	Swiss_Prot:Q58673	73.95	53.333
ST0433	337	3-isopropylmalate dehydrogenase*	Owl:D86857	193.59	100.00
ST0202	582	acetolactate synthase large subunit	Swiss_Prot:Q02137	35.88	29.480
ST0752	530	acetolactate synthase large subunit	Swiss_Prot:O85293	40.79	27.553
ST1447	572	acetolactate synthase large subunit	Swiss_Prot:Q57725	73.53	46.749
ST2410	531	acetolactate synthase large subunit	Swiss_Prot:O85293	23.36	29.630
ST2172	560	dihydroxy-acid dehydratase	Swiss_Prot:P51785	102.82	45.586
ST1445	332	ketol-acid reductoisomerase	Swiss_Prot:P29107	97.60	49.844
ST0295	404	threonine dehydratase	Swiss_Prot:P05792	85.40	46.795
Serine family					
ST1380	300	cysteine synthase	Swiss_Prot:P16703	42.65	31.034
ST0694	297	D-3-phosphoglycerate dehydrogenase	Swiss_Prot:P45637	26.41	31.923
ST1218	313	D-3-phosphoglycerate dehydrogenase	Swiss_Prot:O29445	63.26	44.625
ST1354	433	serine hydroxymethyltransferase	Swiss_Prot:Q58992	89.29	47.794
Histidine family					
ST1459	283	ATP phosphoribosyltransferase	Swiss_Prot:O33771	144.00	73.239
ST1462	249	hisF protein (cyclase)	Swiss_Prot:O33774	87.50	75.697
ST1466	197	hisH protein (amidotransferase)	Swiss_Prot:O33777	115.68	65.829
ST1464	394	histidinol dehydrogenase	Swiss_Prot:O33775	124.39	61.404
ST1458	286	histidinol-phosphate aminotransferase	Swiss_Prot:O33770	132.74	58.246
ST1461	194	imidazoleglycerol-phosphate dehydratase	Swiss_Prot:O33773	97.76	68.421
ST1465	108	phosphoribosyl-AMP cyclohydrolase	Swiss_Prot:O33776	75.67	68.085
ST1467	119	phosphoribosyl-AMP cyclohydrolase	Swiss_Prot:O33778	73.83	72.840
ST1460	230	phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase	Swiss_Prot:O33772	107.91	52.155
General					
ST0165	323	aminotransferase	Swiss_Prot:Q10503	17.94	32.051
ST0735	419	aminotransferase	Swiss_Prot:P94427	50.88	37.736
ST1114	441	aminotransferase	Swiss_Prot:P30268	46.92	32.493
ST1011	433	D-hydantoinase	Swiss_Prot:Q45515	59.49	40.000
ST1020	413	D-hydantoinase	Swiss_Prot:Q44184	41.28	27.602
ST1118	464	D-hydantoinase	Swiss_Prot:Q45515	69.35	41.189
ST1686	641	hydantoin utilization protein	Swiss_Prot:Q01262	98.00	37.452
ST1685	503	N-methylhydantoinase (ATP-hydrolyzing)	Swiss_Prot:Q10094	79.16	37.620
Purines, pyrimidines, nucleosides, and nucleotides					
2'-Deoxyribonucleotide metabolism					
ST1259	831	ribonucleoside-diphosphate reductase large subunit	Swiss_Prot:P07742	41.79	27.856
Nucleotide and nucleoside interconversions					
ST0408	180	adenylate kinase	Swiss_Prot:O05982	86.26	59.880
ST0410	197	adenylate kinase	Swiss_Prot:P35028	128.23	78.238
ST0280	148	nucleoside diphosphate kinase	Swiss_Prot:Q58661	49.01	60.741
Purines					
ST0288	450	adenylosuccinate lyase	Swiss_Prot:O58582	98.06	46.939
ST0291	335	adenylosuccinate synthetase	Swiss_Prot:Q57981	110.86	49.701
ST1496	443	amidophosphoribosyltransferase	Swiss_Prot:Q57657	80.43	43.534
ST1497	403	amidophosphoribosyltransferase	Swiss_Prot:Q57657	65.41	41.290
ST0296	171	bis(5'-adenosyl)-triphosphatase	Swiss_Prot:Q58276	29.44	33.028
ST0878	188	GMP synthase	Swiss_Prot:P49057	35.57	32.787
ST2345	370	GMP synthase	Swiss_Prot:O25165	50.89	38.222
ST0770	254	inosine-5'-monophosphate dehydrogenase	Swiss_Prot:Q58069	40.49	43.697
ST1498	477	phosphoribosylamine--glycine ligase	Swiss_Prot:Q58347	80.80	38.626
ST0807	364	phosphoribosylaminoimidazole carboxylase ATPase subunit	Swiss_Prot:O06457	164.02	70.787
ST0806	157	phosphoribosylaminoimidazole carboxylase catalytic subunit	Swiss_Prot:O06456	85.13	82.166
ST1493	234	phosphoribosylaminoimidazole-succinocarboxamide synthase	Swiss_Prot:P12046	70.56	41.048
ST1499	319	phosphoribosylformylglycinamidine cyclo-ligase	Swiss_Prot:Q57656	74.60	45.421
ST1494	150	phosphoribosylformylglycinamidine synthase I	Swiss_Prot:P12041	46.13	44.000
ST1495	703	phosphoribosylformylglycinamidine synthase II	Swiss_Prot:P12042	115.15	39.165
ST2307	400	phosphoribosylglycinamide formyltransferase	Swiss_Prot:Q58881	105.23	50.627
ST1946	291	ribose-phosphate pyrophosphokinase	Swiss_Prot:Q42583	45.28	28.938
Pyrimidines					
ST1480	298	aspartate carbamoyltransferase catalytic subunit	Swiss_Prot:Q55338	105.05	70.234
ST1479	163	aspartate carbamoyltransferase regulatory subunit	Swiss_Prot:P74766	103.75	72.785
ST1504	1049	carbamoyl-phosphate synthase large subunit	Swiss_Prot:Q59969	130.12	75.760
ST1503	372	carbamoyl-phosphate synthase small subunit	Swiss_Prot:Q59968	142.23	76.944
ST0739	682	carbon-monoxide dehydrogenase large subunit	Swiss_Prot:Q46509	19.74	26.515
ST1781	744	carbon-monoxide dehydrogenase large subunit	Swiss_Prot:Q46509	37.77	25.174
ST2339	706	carbon-monoxide dehydrogenase large subunit	Swiss_Prot:P77489	32.66	23.236
ST0562	278	carbon-monoxide dehydrogenase middle subunit	Swiss_Prot:Q46800	32.61	25.735
ST1783	283	carbon-monoxide dehydrogenase middle subunit	Swiss_Prot:Q46800	43.65	28.058
ST1782	135	carbon-monoxide dehydrogenase small subunit	Swiss_Prot:P77165	51.89	46.622
ST0237	495	CTP synthase	Swiss_Prot:Q58574	82.12	55.600
ST1014	355	cytosine deaminase	Swiss_Prot:P25524	50.51	31.395
ST2403	412	cytosine deaminase	Swiss_Prot:P25524	135.32	46.556
ST1478	340	dihydroorotase	Swiss_Prot:O08357	185.16	69.912
ST1120	350	dihydroorotate dehydrogenase	Swiss_Prot:P25889	74.59	35.455
ST1477	261	dihydroorotate dehydrogenase	Swiss_Prot:O08358	134.01	64.504
ST1632	167	dUTPase	Swiss_Prot:Q02103	38.78	31.944
ST1482	213	orotidine-5'-monophosphate decarboxylase	Swiss_Prot:Q57700	45.82	35.821
ST0462	190	thymidylate kinase	Swiss_Prot:Q57741	46.63	43.605
ST1543	213	thymidylate kinase	Swiss_Prot:O59366	31.70	33.793
ST1481	195	uridine 5'-monophosphate synthase	Swiss_Prot:O08359	109.00	61.340
Salvage of nucleosides and nucleotides					
ST0485	271	5'-methylthioadenosine phosphorylase	Swiss_Prot:Q13126	87.53	48.387
ST0975	237	5'-methylthioadenosine phosphorylase	Swiss_Prot:P50389	132.83	67.094
ST0484	155	hypoxanthine-guanine phosphoribosyltransferase	Swiss_Prot:P27605	13.39	31.868
ST0803	308	inosine-uridine preferring nucleoside hydrolase	Swiss_Prot:P32986	86.82	52.047
ST2115	306	inosine-uridine preferring nucleoside hydrolase	Swiss_Prot:P32986	119.59	71.176
ST0551	176	purine phosphoribosyltransferase	Swiss_Prot:Q64531	17.36	31.081
ST0281	216	uracil phosphoribosyltransferase	Swiss_Prot:P47276	46.23	34.300
Other					
ST1018	264	beta-ureidopropionase	Swiss_Prot:Q03248	29.28	34.194
ST1121	281	beta-ureidopropionase	Swiss_Prot:P49954	34.89	27.972
ST1116	490	methylmalonate-semialdehyde dehydrogenase	Swiss_Prot:P42412	63.14	48.333
ST2490	685	nicotine dehydrogenase subunit C	Swiss_Prot:Q46509	18.49	25.959
ST0093	762	xanthine dehydrogenase	Swiss_Prot:P80457	30.26	28.497
ST0094	438	xanthine dehydrogenase	Swiss_Prot:P77165	38.54	41.525
Fatty acid and phospholipid metabolism					
Biosynthesis					
ST1352	411	3-hydroxy-3-methylglutaryl-CoA reductase	Swiss_Prot:O08424	106.93	75.184
ST0070	253	3-oxoacyl-[acyl-carrier protein] reductase	Swiss_Prot:P51831	36.12	38.776
ST1109	250	3-oxoacyl-[acyl-carrier protein] reductase	Swiss_Prot:O67610	46.19	49.799
ST1868	260	3-oxoacyl-[acyl-carrier protein] reductase	Swiss_Prot:P51831	33.77	37.959
ST2287	235	3-oxoacyl-[acyl-carrier protein] reductase	Swiss_Prot:P39333	20.50	26.840
ST0514	397	acetyl-CoA acetyltransferase	Swiss_Prot:P45359	60.78	44.898
ST1349	348	acyl carrier protein synthase	Swiss_Prot:P43711	26.36	30.282
ST0972	257	diphosphomevalonate decarboxylase	Swiss_Prot:P32377	63.00	40.367
ST0487	283	isoprenyl diphosphate synthase	Swiss_Prot:Q53479	27.13	30.769
ST2353	386	sulfolipid biosynthesis protein sqdB	Swiss_Prot:P45602	12.95	45.455
Degradation					
ST1516	269	3-hydroxybutyryl-CoA dehydratase	Swiss_Prot:P52046	74.06	45.276
ST0069	652	3-hydroxybutyryl-CoA dehydrogenase	Swiss_Prot:P52041	55.78	45.745
ST1507	318	3-hydroxybutyryl-CoA dehydrogenase	Swiss_Prot:P14755	51.27	32.013
ST2094	300	3-hydroxybutyryl-CoA dehydrogenase	Swiss_Prot:P52041	67.45	44.840
ST2413	278	3-hydroxybutyryl-CoA dehydrogenase	Swiss_Prot:P52041	57.29	40.357
ST0066	389	acyl-CoA dehydrogenase	Swiss_Prot:P71539	52.18	27.532
ST0076	400	acyl-CoA dehydrogenase	Swiss_Prot:P45857	64.22	45.503
ST0087	360	acyl-CoA dehydrogenase	Swiss_Prot:P12007	38.65	31.801
ST0916	358	acyl-CoA dehydrogenase	Swiss_Prot:P45867	47.87	30.636
ST1791	404	acyl-CoA dehydrogenase	Swiss_Prot:P45867	59.72	38.032
ST1802	374	acyl-CoA dehydrogenase	Swiss_Prot:P45857	53.84	34.247
ST2256	379	acyl-CoA dehydrogenase	Swiss_Prot:P45867	57.73	35.809
ST0067	372	alpha-methylacyl-CoA racemase	Swiss_Prot:P76518	79.80	32.484
ST1048	254	enoyl-CoA hydratase	Swiss_Prot:P76082	55.47	35.156
ST2416	258	enoyl-CoA hydratase	Swiss_Prot:P34559	75.71	47.541
ST2417	250	enoyl-CoA hydratase	Swiss_Prot:P77467	49.38	33.333
ST0095	243	fatty-acid--CoA dehydratase	Swiss_Prot:P52046	36.91	35.385
ST0062	510	fatty-acid--CoA ligase	Swiss_Prot:Q00594	95.34	38.867
ST0077	536	fatty-acid--CoA ligase	Swiss_Prot:Q00594	71.94	31.953
ST0669	226	glycerophosphoryl diester phosphodiesterase	Swiss_Prot:P54527	40.89	34.199
ST1112	504	long-chain-fatty-acid--CoA ligase	Swiss_Prot:P94547	62.74	34.893
ST1388	454	long-chain-fatty-acid--CoA ligase	Swiss_Prot:P31687	36.85	30.345
General					
STS137	90	acylphosphatase	Swiss_Prot:P75877	39.65	52.500
ST0592	169	biotin carboxyl carrier protein	Swiss_Prot:P05115	33.38	43.478
ST0593	513	biotin carboxylase	Swiss_Prot:Q58626	85.76	55.281
ST0071	303	esterase	Swiss_Prot:P23872	32.10	25.714
ST0079	397	nonspecific lipid-transfer protein	Swiss_Prot:P32020	53.47	30.769
ST0096	373	nonspecific lipid-transfer protein	Swiss_Prot:Q07598	49.73	31.282
ST1110	386	nonspecific lipid-transfer protein	Swiss_Prot:P32020	21.13	25.108
ST1799	391	nonspecific lipid-transfer protein	Swiss_Prot:Q07598	32.68	29.730
ST2418	387	nonspecific lipid-transfer protein	Swiss_Prot:Q07598	58.72	33.075
Biosynthesis of cofactors, prosthetic groups, and carriers					
Heme, porphyrin, and cobalamin					
ST1821	354	cbiD protein	Swiss_Prot:Q60342	77.85	36.994
ST1826	268	cbiG protein	Swiss_Prot:Q05631	17.54	30.108
ST2342	269	cobalamin biosynthesis protein cobD	Swiss_Prot:P21634	30.50	39.423
ST2344	505	cobyric acid synthase	Swiss_Prot:Q58860	64.78	37.466
ST0389	437	cobyrinic acid a,c-diamide synthase	Swiss_Prot:P29946	79.78	33.184
ST0214	336	delta-aminolevulinic acid dehydratase	Swiss_Prot:Q60178	63.73	57.055
ST0215	304	glutamate-1-semialdehyde 2,1-aminomutase	Swiss_Prot:Q58020	58.52	47.351
ST0212	412	glutamyl-tRNA reductase	Swiss_Prot:O28304	34.53	39.683
ST0217	294	porphobilinogen deaminase	Swiss_Prot:Q57989	65.84	44.304
ST1817	222	precorrin-2 c20-methyltransferase	Swiss_Prot:Q58181	38.23	32.843
ST1823	253	precorrin-3b c17-methyltransferase	Swiss_Prot:Q05590	78.70	42.678
ST1825	222	precorrin-3b c17-methyltransferase	Swiss_Prot:O29534	24.59	29.508
ST1816	278	precorrin-4 c11-methyltransferase	Swiss_Prot:Q05630	50.97	40.110
ST1819	192	precorrin-8W decarboxylase	Swiss_Prot:Q57836	52.14	39.011
ST1824	333	precorrin-8x methylmutase	Swiss_Prot:Q58340	46.08	35.676
ST0211	229	siroheme synthase	Swiss_Prot:Q57605	31.91	32.000
ST0563	239	uroporphyrin-III C-methyltransferase	Swiss_Prot:P29564	72.20	49.761
Lipoate					
ST1896	374	lipoate-protein ligase	Swiss_Prot:P54511	22.98	27.731
Menaquinone and ubiquinone					
ST0360	212	3-octaprenyl-4-hydroxybenzoate carboxy-lyase	Swiss_Prot:Q57566	72.18	44.086
ST2204	297	4-hydroxybenzoate octaprenyltransferase	Swiss_Prot:O52366	36.09	26.990
ST0126	311	coenzyme PQQ synthesis protein	Swiss_Prot:P50182	23.99	29.091
ST0127	347	coenzyme PQQ synthesis protein	Swiss_Prot:Q58234	21.49	26.901
ST2412	338	coenzyme PQQ synthesis protein	Swiss_Prot:P27507	24.04	31.548
Molybdopterin					
ST1383	313	molybdenum cofactor biosynthesis protein A	Swiss_Prot:Q58234	91.14	44.718
ST2174	323	molybdenum cofactor biosynthesis protein A	Swiss_Prot:Q57567	63.99	35.838
ST2315	178	molybdenum cofactor biosynthesis protein B	Swiss_Prot:P30746	40.94	32.824
ST0472	151	molybdenum cofactor biosynthesis protein C	Swiss_Prot:Q58535	80.38	59.574
ST0504	545	molybdopterin biosynthesis moeA protein	Swiss_Prot:Q58296	78.79	41.161
ST2045	388	molybdopterin biosynthesis moeA protein	Swiss_Prot:Q58296	51.33	33.010
ST0527	236	molybdopterin converting factor	Swiss_Prot:P45308	24.75	27.869
STS230	93	pterin-4-alpha-carbinolamine dehydratase	PIR:S74881	51.97	45.349
Pantothenate and coenzyme A					
ST0533	267	3-methyl-2-oxobutanoate hydroxymethyltransferase	Swiss_Prot:P52996	90.16	48.438
Riboflavin, FMN, and FAD					
ST0394	157	6,7-dimethyl-8-ribityllumazine synthase	Swiss_Prot:Q57751	82.74	59.420
ST1385	215	GTP cyclohydrolase	Swiss_Prot:O06273	50.47	45.977
Thioredoxin, glutaredoxin, glutathione					
ST2355	478	gamma-glutamyltranspeptidase	Swiss_Prot:P18956	32.92	28.620
ST2123	140	thioredoxin	Swiss_Prot:P37395	35.45	50.562
STS171	86	thioredoxin	Swiss_Prot:P52231	38.53	48.235
ST0438	335	thioredoxin reductase	Swiss_Prot:P23160	51.28	36.755
ST0546	324	thioredoxin reductase	Swiss_Prot:O32823	49.59	41.042
ST2133	327	thioredoxin reductase	Swiss_Prot:O32823	39.84	30.769
Thiamine					
ST0526	448	phosphomethylpyrimidine kinase	Swiss_Prot:P44697	63.04	37.308
ST2329	402	phosphomethylpyrimidine kinase	Swiss_Prot:P44697	80.31	44.186
ST1864	429	thiamine biosynthesis protein thiC	Swiss_Prot:Q58432	118.43	44.917
ST2225	368	thiamine biosynthesis protein thiI	Swiss_Prot:O27720	71.63	36.685
ST0361	266	thiamine biosynthetic enzyme	Swiss_Prot:O29556	109.02	62.548
Pyridine nucleotides					
ST2160	269	glutamine-dependent NAD(+) synthetase	Swiss_Prot:O67091	36.24	35.156
ST1196	472	L-aspartate oxidase (quinolinate synthetase B)	Swiss_Prot:P38032	76.40	39.516
ST2159	279	NH(3)-dependent NAD(+) synthetase	Swiss_Prot:O27554	73.01	42.742
ST1198	276	nicotinate-nucleotide pyrophosphorylase	Swiss_Prot:O27860	77.66	47.273
ST1197	321	quinolinate synthetase A	Swiss_Prot:Q57850	102.65	44.951
Chlorophyll					
ST1021	312	chlorophyll a oxygenase	Swiss_Prot:Q44256	20.16	35.484
ST0056	532	magnesium-protoporphyrin ix monomethyl ester oxidative cyclase 66 kd subunit	Swiss_Prot:P26168	34.67	27.157
General					
ST0582	200	isochorismatase	Swiss_Prot:P37532	39.17	31.034
Central intermediary metabolism					
Amino sugars					
ST2186	592	glucosamine--fructose-6-phosphate aminotransferase	Swiss_Prot:Q56213	85.57	38.789
ST2545	581	glucosamine--fructose-6-phosphate aminotransferase	Swiss_Prot:Q56213	76.74	33.162
ST2546	371	N-acetylglucosamine-6-phosphate deacetylase	Swiss_Prot:P96166	62.99	33.628
Phosphorus compounds					
ST1544	417	exopolyphosphatase	Swiss_Prot:P44828	28.59	29.286

Table 11. Continued.

ORF ID	Length (a.a.)	Predicted Product	Locus	Z-score	% Identity
ST0524	172	inorganic pyrophosphatase*	PIR:JE0392	122.67	100.00
ST1909	370	pholI-like protein	Swiss_Prot:P46343	27.56	30.178
Polyamine biosynthesis					
ST0352	295	agmatinase	Swiss_Prot:Q57757	58.87	36.059
ST0346	300	spermidine synthase	Swiss_Prot:P09158	56.30	35.185
Sulfur metabolism					
ST0723	156	NADH-dependent FMN oxydoreductase	Swiss_Prot:P54990	23.28	28.689
ST2567	239	phosphoadenosine phosphosulfate reductase	Swiss_Prot:O05927	77.55	44.872
ST2568	412	sulfate adenylyltransferase	Swiss_Prot:P08536	45.03	35.659
ST2566	628	sulfite reductase	Swiss_Prot:P52673	56.21	30.900
ST1127	311	sulfur oxygenase reductase	Swiss_Prot:P29082	188.56	68.730
ST1784	267	thiosulfate sulfurtransferase	Swiss_Prot:Q50036	77.84	41.481
ST2564	293	thiosulfate sulfurtransferase	Swiss_Prot:O05793	85.62	48.057
Nitrogen fixation					
ST2267	287	hesA protein	Swiss_Prot:P30138	42.41	30.508
ST2446	276	nodulation ATP-binding protein	Swiss_Prot:P32010	32.91	35.242
ST1956	504	nodulation protein	Swiss_Prot:Q58458	63.91	35.333
Nitrogen metabolism					
ST1023	148	cyanate lyase	Swiss_Prot:Q59948	59.35	42.958
ST0478	396	enantiomer-selective amidase	Swiss_Prot:O27540	49.57	45.783
ST2489	310	morphine 6-dehydrogenase	Swiss_Prot:P22045	35.77	31.933
ST1414	255	tropinesterase	Swiss_Prot:P07383	27.39	30.928
ST1028	493	urease alpha subunit	Swiss_Prot:Q07397	59.55	55.070
ST1029	224	urease gamma subunit	Swiss_Prot:P42822	56.22	49.107
Degradation of amino acids					
ST1010	208	sulfite oxidase	Swiss_Prot:Q07116	24.14	41.333
General					
ST2099	193	2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrola	Swiss_Prot:P47229	22.20	33.621
ST1705	281	2-hydroxyhepta-2,4-diene-1,7-dioate isomerase	Swiss_Prot:P34673	54.78	34.375
ST1659	502	4-hydroxyphenylacetate 3-monooxygenase large subunit	PIR:E69291	170.73	49.684
ST0344	350	glycerol dehydrogenase	Swiss_Prot:Q58122	84.87	45.745
ST0591	523	methylmalonyl-CoA decarboxylase alpha subunit	Swiss_Prot:P54541	99.44	51.984
ST2096	140	methylmalonyl-CoA mutase alpha subunit	Swiss_Prot:Q05065	40.80	42.424
ST2055	208	phosphatase	Swiss_Prot:O67359	19.90	28.824
ST1013	637	quinoline 2-oxidoreductase	Swiss_Prot:P08793	13.88	26.768
ST1571	693	quinoline 2-oxidoreductase	Swiss_Prot:P80457	20.99	28.242
Energy metabolism					
Aerobic respiration					
ST0109	529	cytochrome b	Swiss_Prot:P39480	30.48	28.610
ST0137	566	cytochrome b	Swiss_Prot:P39480	150.32	63.588
ST1667	494	cytochrome b	Swiss_Prot:P39480	87.05	49.038
ST5007	55	cytochrome b	OWL:AF072619	9.09	27.500
ST0697	184	NADH dehydrogenase	Swiss_Prot:O28017	39.76	32.450
ST2573	341	NADH dehydrogenase	Swiss_Prot:P00393	21.27	30.612
STS031	86	NADH dehydrogenase subunit I	OWL:AF004650	10.54	27.536
ST2219	353	NADH-plastoquinone oxidoreductase subunit 1	Swiss_Prot:P42032	32.45	31.804
ST2223	462	NADH-plastoquinone oxidoreductase subunit 2	Swiss_Prot:P06256	36.24	28.337
ST2216	128	NADH-plastoquinone oxidoreductase subunit 3	Swiss_Prot:P06259	16.22	25.610
ST2218	392	NADH-plastoquinone oxidoreductase subunit 4	Swiss_Prot:Q56220	64.00	36.387
ST2217	168	NADH-plastoquinone oxidoreductase subunit 5	Swiss_Prot:P19125	34.73	31.092
ST2222	1138	NADH-plastoquinone oxidoreductase subunit 5	Swiss_Prot:P26849	23.42	30.894
ST2221	165	NADH-plastoquinone oxidoreductase subunit 6	Swiss_Prot:P26523	12.06	23.977
ST2220	169	NADH-plastoquinone oxidoreductase subunit 9	Swiss_Prot:P77423	27.19	34.286
STS016	59	NADH-ubiquinone oxidoreductase subunit 4L	Swiss_Prot:P24874	9.82	26.415
ST1527	173	NADH-ubiquinone oxidoreductase subunit K	Swiss_Prot:O75251	50.20	42.958
ST0136	524	quinol oxidase polypeptide I	Swiss_Prot:P98004	149.44	58.824
ST2396	524	quinol oxidase polypeptide I	Swiss_Prot:P98004	112.83	50.547
ST0103	815	quinol oxidase polypeptide I/III	Swiss_Prot:P39481	85.43	58.564
ST0135	172	quinol oxidase polypeptide II	Swiss_Prot:P39479	109.79	64.072
ST2395	179	quinol oxidase polypeptide II	Swiss_Prot:P39479	32.05	33.714
ST0108	244	Rieske iron-sulfur protein	Swiss_Prot:P70758	13.17	55.556
ST0134	327	Rieske iron-sulfur protein	PIR:S71348	120.40	60.500
ST1666	326	Rieske iron-sulfur protein	PIR:S71348	192.70	77.912
ST0104	197	sulfocyanin (blue copper protein)	Swiss_Prot:Q53765	94.59	52.792
ST2393	225	sulfocyanin (blue copper protein)	Swiss_Prot:Q53765	56.46	49.080
ST2394	217	sulfocyanin (blue copper protein)	Swiss_Prot:Q53765	64.41	49.080
Amino acids and amines					
ST0619	291	3-hydroxyisobutyrate dehydrogenase	Swiss_Prot:Q55702	48.98	31.142
ST0151	318	acetamidase	Swiss_Prot:Q07838	28.01	63.830
ST1206	346	aminomethyltransferase (glycine cleavage system T protein)	Swiss_Prot:P54261	56.96	33.333
ST1031	378	arginine deiminase	Swiss_Prot:P13981	27.43	28.087
ST1205	142	glycine cleavage system H protein	Swiss_Prot:P23884	51.41	47.458
ST1894	136	glycine cleavage system H protein	Swiss_Prot:Q30733	33.53	38.542
ST1895	148	glycine cleavage system H protein	Swiss_Prot:P23884	39.01	40.000
ST1207	449	glycine dehydrogenase subunit 1 (glycine cleavage system P-protein subunit)	Swiss_Prot:P54376	113.23	44.851
ST1208	505	glycine dehydrogenase subunit 2 (glycine cleavage system P-protein subunit)	Swiss_Prot:P54377	122.33	49.263
ST1264	448	L-asparaginase	Swiss_Prot:Q60331	134.89	46.394
ST1827	271	L-asparaginase	Swiss_Prot:P37595	37.07	50.000
ST1108	373	mandelate racemase	Swiss_Prot:P31458	42.72	28.966
ST0726	494	phenol hydroxylase	PIR:H75574	70.49	31.670
Aerobic					
ST2279	100	3-phenylpropionate dioxygenase ferredoxin subunit	Swiss_Prot:P77266	25.77	28.889
ST2095	474	phenylacetyl-CoA ligase	PIR:D69501	41.16	29.293
ST2420	429	phenylacetyl-CoA ligase	PIR:B69115	102.16	43.287
Anaerobic					
ST1789	391	anaerobic dimethyl sulfoxide reductase	Swiss_Prot:P18776	40.04	51.456
ST2369	396	anaerobic glycerol-3-phosphate dehydrogenase subunit C	Swiss_Prot:P13034	37.47	28.936
ST2574	367	L-carnitine dehydratase	Swiss_Prot:P76518	53.42	32.166
ST1839	270	thiosulfate reductase electron transport protein phsB	Swiss_Prot:P45015	33.83	35.172
ATP-proton motive force interconversion					
STS175	55	ATP synthase protein 8	Swiss_Prot:Q34942	10.28	37.838
STS213	52	ATP synthase protein 8	Swiss_Prot:Q02653	9.17	34.091
Electron transport					
ST0064	468	aldehyde dehydrogenase	Swiss_Prot:Q59931	37.69	35.637
ST1664	466	cytochrome b558/566 subunit A	OWL:SAY10108	249.26	75.322
STf665	314	cytochrome b558/566 subunit B	OWL:SAY101081	102.55	44.333
ST0611	569	cytochrome c oxidase polypeptide I	Swiss_Prot:P33518	28.65	29.588
ST0676	522	cytochrome c oxidase polypeptide I	Swiss_Prot:P33518	29.26	29.020
ST2053	593	cytochrome c oxidase polypeptide I	Swiss_Prot:P33518	29.75	29.389
ST2595	590	cytochrome c oxidase polypeptide I	Swiss_Prot:Q06473	33.98	27.484
ST0105	147	cytochrome c oxidase polypeptide II	Swiss_Prot:Q04441	21.71	37.113
ST1148	367	cytochrome P450	Swiss_Prot:Q55080	148.59	63.514
ST1794	282	electron transfer flavoprotein alpha subunit	Swiss_Prot:P31574	48.76	44.828
ST1793	241	electron transfer flavoprotein beta subunit	Swiss_Prot:P97089	46.73	37.056
ST1175	104	ferredoxin	PIR:JC4907	108.56	90.385
ST1837	457	ferredoxin	Swiss_Prot:P00211	21.06	37.931
STS199	93	ferredoxin	Swiss_Prot:P53658	27.49	37.895
STS201	89	ferredoxin	Swiss_Prot:P53658	25.10	36.047
ST0163	104	ferredoxin*	PIR:JC4907	106.44	100.00
ST1776	602	fixB protein	Swiss_Prot:P71153	62.76	41.089
ST1775	384	fixC protein	Swiss_Prot:P10331	52.69	35.474
ST1795	406	fixC protein	Swiss_Prot:P31575	67.73	32.840
ST1873	365	heterodisulfide reductase	Swiss_Prot:P32009	12.81	50.000
STS260	85	NADH dehydrogenase subunit III	OWL:AF0763741	8.70	27.419
STS234	94	NADH-plastoquinone oxidoreductase subunit 4L	Swiss_Prot:Q00244	23.84	34.091
ST0090	546	oxidoreductase	Swiss_Prot:P32382	69.82	37.104
STS024	53	photosystem I reaction centre subunit VIII (PSI-I)	Swiss_Prot:P52768	9.85	37.209
ST0228	233	redoxin-like protein	Swiss_Prot:Q57755	29.17	38.824
ST0092	378	reductase	Swiss_Prot:P43494	57.30	32.479
ST0938	245	reductase	Swiss_Prot:P54554	32.99	36.646
ST1299	265	reductase	Swiss_Prot:O67610	31.29	31.034
ST2239	262	reductase	Swiss_Prot:P39640	31.20	30.830
ST2370	144	rubrerythrin	Swiss_Prot:P24931	46.86	36.552
ST0615	384	sulfide dehydrogenase [flavocytochrome c] flavoprotein subunit	Swiss_Prot:Q06530	23.91	27.823
ST0971	390	sulfide dehydrogenase [flavocytochrome c] flavoprotein subunit	Swiss_Prot:P81540	43.11	72.917
STS219	66	terminal oxidase small hydrophobic subunit	OWL:AADOXP84B3	62.16	71.429
Fermentation					
ST1532	386	2-ketovalerate ferredoxin oxidoreductase alpha-2	Swiss_Prot:O05651	123.64	43.850
ST1533	300	2-ketovalerate ferredoxin oxidoreductase beta-2	Swiss_Prot:P52647	17.25	29.050
ST1787	306	2-ketovalerate ferredoxin oxidoreductase beta-2	Swiss_Prot:Q53046	23.12	31.214
ST1884	332	acetoin utilization acuC protein	Swiss_Prot:P39067	33.32	34.906
ST2305	362	acetoin utilization acuC protein	Swiss_Prot:P39067	66.35	40.938
ST2324	329	acetoin utilization acuC protein	Swiss_Prot:Q57955	76.90	43.791
ST1350	363	acetyl-CoA synthase	Swiss_Prot:P22307	20.64	29.032
ST0050	562	acetyl-CoA synthetase	Swiss_Prot:P39062	103.15	41.682
ST0730	662	acetyl-CoA synthetase	Swiss_Prot:P27095	90.47	51.047
ST0783	659	acetyl-CoA synthetase	Swiss_Prot:P27095	85.44	44.984
ST1190	497	acetyl-CoA synthetase	Swiss_Prot:P39062	78.21	36.756
ST1796	609	acetyl-CoA synthetase	Swiss_Prot:P27095	67.44	32.131
ST1803	621	acetyl-CoA synthetase	Swiss_Prot:P27095	70.82	34.603
ST2019	661	acetyl-CoA synthetase	Swiss_Prot:P27095	90.53	50.161
ST2139	576	acetyl-CoA synthetase	Swiss_Prot:P27550	54.63	31.350
ST2575	544	acetyl-CoA synthetase	Swiss_Prot:P39062	78.23	35.727
ST0038	307	alcohol dehydrogenase	Swiss_Prot:P39462	47.81	37.847
ST0053	344	alcohol dehydrogenase	Swiss_Prot:P50381	93.74	55.172
ST0480	334	alcohol dehydrogenase	Swiss_Prot:P12311	40.25	32.647
ST0569	333	alcohol dehydrogenase	Swiss_Prot:P42327	43.15	32.544
ST2056	361	alcohol dehydrogenase	Swiss_Prot:P49645	35.07	41.209
ST2252	332	alcohol dehydrogenase	Swiss_Prot:P42328	30.48	29.878
ST2577	347	alcohol dehydrogenase	Swiss_Prot:P39462	135.99	89.337
ST2605	330	alcohol dehydrogenase	Swiss_Prot:P42328	38.82	33.945
STS015	54	alcohol dehydrogenase	Swiss_Prot:P12311	19.02	46.000
ST2484	730	aldehyde oxidoreductase	Swiss_Prot:P77489	33.83	24.957
ST1786	252	citrate lyase beta subunit	Swiss_Prot:O53078	46.73	31.955
ST0552	562	coenzyme B12-dependent isobutyryl-CoA mutase	Swiss_Prot:P22033	90.21	45.756
ST2295	458	D-lactate dehydrogenase	Swiss_Prot:P52075	106.82	37.558
ST0081	973	formate dehydrogenase alpha subunit	Swiss_Prot:P06131	115.46	40.816
ST0348	654	formate dehydrogenase alpha subunit	Swiss_Prot:P07658	90.60	34.548
ST0084	255	formate dehydrogenase subunit fdhD	Swiss_Prot:Q10820	65.14	40.551
ST0823	252	glutaconate CoA-transferase subunit A	Swiss_Prot:Q59111	26.55	29.688
ST0732	598	indolepyruvate oxidoreductase subunit A	Swiss_Prot:Q01700	24.39	38.596
ST0734	186	indolepyruvate oxidoreductase subunit B	Swiss_Prot:P80906	14.94	68.750
ST1811	311	L-lactate dehydrogenase	Swiss_Prot:P33572	32.17	28.205
ST0114	414	malate oxidoreductase	Swiss_Prot:P16468	114.66	46.073
ST0278	443	malate oxidoreductase	Swiss_Prot:P16468	117.25	47.656
ST0075	311	NAD-dependent alcohol dehydrogenase	Swiss_Prot:P77280	24.71	28.906
ST1531	181	pyruvate ferredoxin oxidoreductase	Swiss_Prot:O05650	43.56	33.516
Glycolysis/gluconeogenesis					
ST1212	416	enolase	Swiss_Prot:Q60173	49.17	48.434
ST1356	343	glyceraldehyde 3-phosphate dehydrogenase	Swiss_Prot:P39460	120.44	68.915
ST2477	506	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase	Swiss_Prot:Q59931	46.78	34.591
ST1058	602	phosphoenolpyruvate carboxykinase	Swiss_Prot:P05153	53.86	33.892
ST0698	309	phosphoenolpyruvate synthase	Swiss_Prot:O29548	47.52	31.429
ST1235	764	phosphoenolpyruvate synthase	Swiss_Prot:P46893	154.87	62.578
ST1357	415	phosphoglycerate kinase	Swiss_Prot:P50317	100.90	70.690
ST1617	438	pyruvate kinase	Swiss_Prot:Q57572	53.96	36.689
ST2030	229	triosephosphate isomerase	Swiss_Prot:Q58923	83.89	42.791
Pentose phosphate pathway					
ST1704	360	glucose 1-dehydrogenase	Swiss_Prot:P13203	82.52	44.345
ST2556	253	glucose 1-dehydrogenase	Swiss_Prot:Q56318	33.28	37.200
ST1302	219	ribose 5-phosphate isomerase	Swiss_Prot:P44725	30.21	30.374
ST2268	273	transketolase	Swiss_Prot:Q58094	84.26	43.902
ST2269	313	transketolase	Swiss_Prot:Q58092	76.07	42.071
Pyruvate dehydrogenase					
ST1487	452	dihydrolipoamide dehydrogenase	Swiss_Prot:P11959	46.78	32.599
ST1879	437	dihydrolipoamide dehydrogenase	Swiss_Prot:P75393	41.24	30.444
ST0687	455	lipoamide dehydrogenase	Swiss_Prot:P11959	41.52	31.087
ST1865	412	lipoamide dehydrogenase	Swiss_Prot:P35484	37.52	32.177
Sugars					
ST2525	641	alpha-glucosidase	Swiss_Prot:O59645	198.53	61.897
ST0773	384	beta-galactosidase	Swiss_Prot:P14288	57.40	36.842
ST2366	396	dgoA protein	Swiss_Prot:P31458	60.95	34.007
ST2561	370	dgoA protein	Swiss_Prot:P31458	24.22	26.794
ST2572	392	dgoA protein	Swiss_Prot:P31458	55.90	30.645
ST2478	311	fructokinase	Swiss_Prot:P50845	39.48	28.155
ST0238	223	hexulose-6-phosphate synthase	Swiss_Prot:P42405	54.11	36.058
ST0574	285	ribokinase	Swiss_Prot:P36945	19.63	26.132
ST2328	291	ribokinase	Swiss_Prot:P05054	58.81	38.947
ST2341	339	sorbitol dehydrogenase	Swiss_Prot:Q06004	47.51	33.718
TCA cycle					
ST2435	628	2-oxoacid--ferredoxin oxidoreductase alpha subunit	PIR:JC4919	196.66	59.076
ST2300	627	2-oxoacid--ferredoxin oxidoreductase alpha subunit*	PIR:JC4919	272.62	99.841
ST2433	304	2-oxoacid--ferredoxin oxidoreductase beta subunit	PIR:JC4920	154.09	71.906
ST2298	305	2-oxoacid--ferredoxin oxidoreductase beta subunit*	PIR:JC4920	186.58	99.672
ST0833	855	aconitate hydratase	Swiss_Prot:P37032	75.22	49.943
ST0587	373	citrate synthase	Swiss_Prot:Q53554	65.12	51.330
ST1805	378	citrate synthase	Swiss_Prot:P80148	96.83	66.845
ST2023	439	fumarate hydratase	Swiss_Prot:P39461	136.45	75.571
ST2166	409	NADP-dependent isocitrate dehydrogenase	Swiss_Prot:P08200	86.12	49.878
ST2090	318	malate dehydrogenase	Swiss_Prot:P16142	65.28	37.828
ST0497	566	succinate dehydrogenase subunit A	Swiss_Prot:P51054	70.21	45.009
ST0498	320	succinate dehydrogenase subunit B	Swiss_Prot:Q10761	54.33	38.767
ST0499	290	succinate dehydrogenase subunit C	PIR:S76574	74.93	33.452
ST1870	307	succinate dehydrogenase subunit C	PIR:B70336	107.08	43.871
ST0500	116	succinate dehydrogenase subunit D	OWL:SASDHABCD-	82.77	57.759
ST0962	259	succinyl-CoA synthetase alpha subunit	Swiss_Prot:Q58643	70.95	58.687
ST0963	272	succinyl-CoA synthetase beta subunit	Swiss_Prot:Q57663	68.62	46.468
Biosynthesis and degradation of polysaccharides					
ST0926	217	alpha-amylase	Swiss_Prot:Q10769	64.80	44.000
ST0927	359	alpha-amylase	Swiss_Prot:Q10769	71.15	42.169
ST1008	951	alpha-mannosidase	Swiss_Prot:P21139	92.45	32.179
ST2337	235	dolichol monophosphate mannose synthase	Swiss_Prot:Q58619	36.75	32.124
ST0928	716	glycogen debranching enzyme	Swiss_Prot:Q10767	160.72	50.602
ST0818	572	granule-bound starch synthase	Swiss_Prot:P54138	23.11	40.323
haloalkane					
ST1836	200	2-haloalkanoic acid dehalogenase	Swiss_Prot:P24070	20.27	31.707
ST2570	201	2-haloalkanoic acid dehalogenase	Swiss_Prot:P24070	34.07	30.168
Other					
ST0045	214	dehydrogenase	Swiss_Prot:P17201	14.98	27.523
ST0047	560	dehydrogenase	Swiss_Prot:Q46509	21.94	23.864
ST0130	267	dehydrogenase	Swiss_Prot:P30863	34.05	31.780
ST0561	171	dehydrogenase	Swiss_Prot:P77165	50.67	43.636
ST0649	440	glycolate oxidase subunit glcD	Swiss_Prot:P52075	55.07	27.376
ST2296	276	glycolate oxidase subunit glcE	Swiss_Prot:P52073	20.42	27.049
ST0724	316	metapyrocatechase	Swiss_Prot:P08127	39.01	28.839
Transport and binding proteins					
Amino acids, peptides and amines					
ST1757	507	amino acid transporter	Swiss_Prot:P46349	18.13	24.074
ST0700	313	oligopeptide transporter ATP-binding protein	Swiss_Prot:P24137	46.25	45.769
ST0701	321	oligopeptide transporter ATP-binding protein	Swiss_Prot:P76027	46.16	45.033
ST0702	482	oligopeptide transporter permease protein	Swiss_Prot:P42063	43.91	33.476
ST0704	353	oligopeptide transporter permease protein	Swiss_Prot:Q53191	46.57	26.606
ST1647	279	peptide transporter ATP-binding protein	Swiss_Prot:P45289	28.06	28.643
ST2536	322	peptide transporter ATP-binding protein	Swiss_Prot:P45095	37.69	35.669
ST2537	312	peptide transporter ATP-binding protein	Swiss_Prot:P45094	36.74	32.459
ST2542	294	peptide transporter ATP-binding protein	Swiss_Prot:P26905	45.45	45.734
ST2543	323	peptide transporter ATP-binding protein	Swiss_Prot:P45094	41.32	44.444
ST1649	409	peptide transporter permease protein	Swiss_Prot:P45054	25.47	24.091
ST2532	285	peptide transporter permease protein	Swiss_Prot:P26904	42.74	32.847
ST2535	333	peptide transporter permease protein	Swiss_Prot:Q07742	50.54	30.350
ST2540	350	peptide transporter permease protein	Swiss_Prot:P77308	57.80	32.013
ST2541	358	peptide transporter permease protein	Swiss_Prot:P77463	58.01	39.574
ST2576	429	proline/betaine transporter	Swiss_Prot:P37643	59.69	33.073
Anions					
ST1818	103	myoglobin	Swiss_Prot:P02163	8.41	28.571
ST2428	458	phosphate transporter	Swiss_Prot:P31679	45.43	26.471
ST2033	272	sulfate transporter permease protein	Swiss_Prot:P18795	39.38	31.250
Carbohydrates, organic alcohols, and acids					
ST2658	445	4-methyl-o-phthalate/phthalate permease	Swiss_Prot:P77228	41.53	33.684
ST0644	374	oxalate:formate antiporter	Swiss_Prot:P05715	127.14	44.693
ST1853	364	oxalate:formate antiporter	Swiss_Prot:P05715	114.08	42.017
ST0911	476	sugar transporter	Swiss_Prot:O52733	41.94	30.227

Table 11. Continued.

ORF ID	Length (a.a.)	Predicted Product	Locus	Zscore	% Identity
ST1788	398	2-ketovalerate ferredoxin oxidoreductase alpha-2	Swiss_Prot:O05651	101.45	43.363
ST2557	459	sugar-proton symport	Swiss_Prot:P46333	50.03	33.021
Cations					
ST0661	518	ammonium transporter	Swiss_Prot:O26759	84.91	41.058
ST1785	155	bacterioferritin comigratory protein	Swiss_Prot:P23480	37.46	36.641
ST1715	740	cation-transporter ATPase	Swiss_Prot:P32113	57.48	33.805
ST2208	260	erythrocyte band 7 integral membrane protein	Swiss_Prot:Q58237	65.68	47.594
ST1436	591	membrane-associated ATPase alpha subunit*	Swiss_Prot:P09639	101.01	99.662
ST1437	465	membrane-associated ATPase beta subunit*	Swiss_Prot:P13052	80.36	98.925
ST1439	101	membrane-associated ATPase c subunit*	Swiss_Prot:P23040	92.04	100.000
ST1435	194	membrane-associated ATPase delta subunit*	Swiss_Prot:P22722	112.67	100.000
ST5172	59	membrane-associated ATPase epsilon subunit*	Swiss_Prot:P23039	64.36	100.000
ST1438	216	membrane-associated ATPase gamma subunit*	Swiss_Prot:P22721	168.15	100.000
ST0638	399	Na(+)/H(+) antiporter	Swiss_Prot:P26235	42.23	27.301
ST0728	386	Na(+)/H(+) antiporter	Swiss_Prot:P26235	46.92	29.538
ST1861	500	Na(+)/H(+) antiporter	Swiss_Prot:P26235	25.07	25.946
ST2443	383	nickel transporter	Swiss_Prot:P23516	100.60	45.734
ST1846	271	senescence marker protein	Swiss_Prot:P46218	146.09	63.636
ST2555	285	senescence marker protein	Swiss_Prot:P46218	82.20	44.074
Nucleosides, purines and pyrimidines					
ST1564	523	allantoin permease	Swiss_Prot:P94575	54.66	28.875
Other					
ST1034	266	heme exporter protein A	Swiss_Prot:P52218	27.60	36.054
ST0666	478	multidrug resistance protein	Swiss_Prot:P39886	46.44	26.142
ST0707	560	multidrug resistance protein	Swiss_Prot:Q00538	35.64	26.905
ST1006	473	multidrug resistance protein	Swiss_Prot:P54585	47.42	25.743
ST1670	468	multidrug resistance protein	Swiss_Prot:P54585	44.18	27.586
ST1709	550	multidrug resistance protein	Swiss_Prot:P39886	35.60	23.126
ST1852	415	multidrug resistance protein	Swiss_Prot:P27173	16.42	25.773
ST2604	480	multidrug resistance protein	Swiss_Prot:P39886	31.62	22.798
Unknown substrate					
ST0265	233	ABC transporter ATP-binding protein	Swiss_Prot:P19844	30.81	30.594
ST0595	240	ABC transporter ATP-binding protein	Swiss_Prot:Q57855	43.80	42.601
ST0608	294	ABC transporter ATP-binding protein	Swiss_Prot:P36879	33.40	37.611
ST0756	224	ABC transporter ATP-binding protein	Swiss_Prot:P46903	31.69	44.203
ST0842	219	ABC transporter ATP-binding protein	Swiss_Prot:Q58429	23.18	33.803
ST0902	232	ABC transporter ATP-binding protein	Swiss_Prot:Q58206	43.60	50.442
ST1099	851	ABC transporter ATP-binding protein	Swiss_Prot:Q11047	38.52	32.400
ST1106	367	ABC transporter ATP-binding protein	Swiss_Prot:P94360	45.65	42.582
ST1136	471	ABC transporter ATP-binding protein	Swiss_Prot:P40755	33.74	38.265
ST1201	248	ABC transporter ATP-binding protein	Swiss_Prot:P80866	64.31	49.590
ST1577	237	ABC transporter ATP-binding protein	Swiss_Prot:P19844	28.18	34.862
ST1672	331	ABC transporter ATP-binding protein	Swiss_Prot:P42332	33.95	36.937
ST1730	268	ABC transporter ATP-binding protein	Swiss_Prot:Q58429	27.12	31.905
ST1735	298	ABC transporter ATP-binding protein	Swiss_Prot:P54592	37.59	39.269
ST1772	334	ABC transporter ATP-binding protein	Swiss_Prot:P32010	41.24	39.677
ST1917	259	ABC transporter ATP-binding protein	Swiss_Prot:P37009	33.51	44.693
ST2035	203	ABC transporter ATP-binding protein	Swiss_Prot:Q58762	28.78	34.426
ST2437	301	ABC transporter ATP-binding protein	Swiss_Prot:P19844	36.96	40.278
ST2473	287	ABC transporter ATP-binding protein	Swiss_Prot:P42332	30.46	35.610
ST2509	234	ABC transporter ATP-binding protein	Swiss_Prot:P56344	28.47	32.569
ST0841	325	ABC transporter permease protein	Swiss_Prot:Q57130	46.32	27.152
ST1104	214	ABC transporter permease protein	Swiss_Prot:P18812	41.98	39.752
ST1105	297	ABC transporter permease protein	Swiss_Prot:O58760	35.14	35.135
ST1918	547	ABC transporter permease protein	Swiss_Prot:P40979	28.62	31.200
ST0535	244	ATP-binding transporter	Swiss_Prot:P54592	32.44	38.967
ST1582	274	nodulation ATP-binding protein	Swiss_Prot:P72335	32.11	38.914
ST0054	442	transporter	Swiss_Prot:P30848	45.10	24.096
ST0065	465	transporter	Swiss_Prot:P37514	38.31	24.714
ST0091	418	transporter	Swiss_Prot:P31679	57.73	29.454
ST0339	365	transporter	Swiss_Prot:P76350	24.74	27.795
ST0691	446	transporter	Swiss_Prot:P31679	27.14	29.834
ST0713	478	transporter	Swiss_Prot:P37514	40.26	26.580
ST0852	425	transporter	Swiss_Prot:Q10858	39.33	28.289
ST0861	379	transporter	Swiss_Prot:P39352	22.26	27.044
ST0934	321	transporter	Swiss_Prot:P39352	23.73	21.854
ST1117	533	transporter	Swiss_Prot:P94575	60.80	26.971
ST1134	103	transporter	Swiss_Prot:Q43932	17.75	30.137
ST1170	448	transporter	Swiss_Prot:P37514	60.45	31.313
ST1828	401	transporter	Swiss_Prot:P44699	29.22	28.646
ST2363	494	transporter	Swiss_Prot:P25297	38.62	31.027
ST2422	388	transporter	Swiss_Prot:P39352	27.72	24.781
ST2476	648	transporter	Swiss_Prot:P27837	19.63	31.034
DNA metabolism					
DNA replication, recombination, and repair					
ST2024	151	7,8-dihydro-8-oxoguanine triphosphatase	Swiss_Prot:P36639	41.98	43.590
ST0401	911	ATP-dependent helicase	Swiss_Prot:P95940	66.67	33.448
STS044	51	copy number control protein copG	PIR:T00008	17.30	50.000
STS109	72	copy number control protein copG	PIR:T00008	21.57	54.839
ST0573	353	dth protein	Swiss_Prot:Q47155	58.51	35.484
ST0889	432	deoxyribodipyrimidine photolyase	Swiss_Prot:Q04449	61.65	41.262
ST0223	600	DNA ligase	Swiss_Prot:Q02093	182.63	71.765
ST0297	324	DNA repair protein radA	Swiss_Prot:Q55075	113.20	85.443
ST0467	548	DNA replication licensing factor mcm	Swiss_Prot:P41389	49.49	46.387
ST1216	664	DNA topoisomerase	Swiss_Prot:P46799	42.93	28.032
STS226	64	DNA-binding protein 7e	Swiss_Prot:P13125	64.48	85.246
ST1426	876	DNA-directed DNA polymerase	Swiss_Prot:O50607	182.51	99.533
ST1680	540	DNA-directed DNA polymerase II	Swiss_Prot:Q07635	208.79	67.754
ST2076	781	DNA-directed DNA polymerase II	Swiss_Prot:O05706	108.89	44.586
STS189	88	DNA-directed DNA polymerase II	Swiss_Prot:Q07635	49.02	72.727
ST2238	220	DNA-directed DNA polymerase, bacteriophage-type	Swiss_Prot:P30314	22.12	31.765
ST1401	224	endonuclease III	Swiss_Prot:P44319	35.14	35.519
ST1910	241	exodeoxyribonuclease	Swiss_Prot:P37454	65.01	37.302
ST0210	304	flap endonuclease	Swiss_Prot:P39748	63.21	37.457
ST0590	704	helicase	Swiss_Prot:P53327	30.60	28.293
STS236	71	integrase	Swiss_Prot:P20214	20.87	50.000
ST1393	284	integrase/recombinase	Swiss_Prot:P21891	26.21	36.196
ST0967	156	methylated-DNA--protein-cysteine methyltransferase	Swiss_Prot:P19220	32.08	36.607
ST2243	146	mutT protein	Swiss_Prot:P32056	19.87	42.105
STS110	59	plasmid copy-number control protein cop-6	OWL:SIU930824	48.26	74.510
STS248	55	plasmid copy-number control protein cop-6	Swiss_Prot:P25921	11.69	29.545
ST0387	245	proliferating cell nuclear antigen	Swiss_Prot:Q57697	46.16	24.691
ST0397	248	proliferating cell nuclear antigen	Swiss_Prot:Q57697	31.30	25.703
ST0944	246	proliferating cell nuclear antigen	Swiss_Prot:Q57697	36.37	26.556
ST0002	579	prolyl endopeptidase	Swiss_Prot:P55577	47.68	37.004
ST0473	413	replication factor C large subunit	Swiss_Prot:P35600	21.12	29.717
ST0475	327	replication factor C small subunit	Swiss_Prot:P80937	72.70	42.628
ST0374	1085	reverse gyrase	Swiss_Prot:Q08582	243.93	61.213
ST1290	1156	reverse gyrase	Swiss_Prot:Q08582	129.79	32.478
STS141	99	reverse gyrase	OWL:SSHTOPR1	108.78	93.750
ST0299	353	sun protein	Swiss_Prot:Q60343	71.45	50.673
ST0683	370	sun protein	Swiss_Prot:P94464	39.41	30.370
ST1390	349	sun protein	Swiss_Prot:P36929	31.90	30.189
ST2423	342	sun protein	Swiss_Prot:P94464	30.74	26.818
ST1522	448	tldD protein	Swiss_Prot:O26944	73.50	35.692
ST1295	387	type II DNA topoisomerase VI subunit A	Swiss_Prot:O05208	203.19	83.247
ST1294	532	type II DNA topoisomerase VI subunit B	Swiss_Prot:O05207	195.40	71.012
Restriction/modification					
ST1151	551	5-methylcytosine-specific restriction enzyme	Swiss_Prot:P15005	20.42	33.088
ST1153	683	5-methylcytosine-specific restriction enzyme	Swiss_Prot:P15005	15.72	27.778
ST0335	322	modification methylase	Swiss_Prot:Q57983	99.58	48.710
Transcription					
Degradation of RNA					
ST0519	208	ribonuclease HII	Swiss_Prot:Q57599	50.03	40.367
ST0443	247	ribonuclease PH	Swiss_Prot:Q17533	57.41	35.160
DNA-dependent RNA polymerase					
ST0276	884	DNA-directed RNA polymerase subunit A'	Swiss_Prot:P11512	127.95	83.068
ST0275	398	DNA-directed RNA polymerase subunit A"	Swiss_Prot:P11514	135.13	75.573
ST0277	1126	DNA-directed RNA polymerase subunit B	Swiss_Prot:P11513	141.21	84.889
ST2067	264	DNA-directed RNA polymerase subunit D	Swiss_Prot:P39471	141.11	75.573
ST0379	183	DNA-directed RNA polymerase subunit E	Swiss_Prot:P39466	126.25	74.863
STS052	67	DNA-directed RNA polymerase subunit E	Swiss_Prot:P39466	29.90	55.769
STS038	99	DNA-directed RNA polymerase subunit H	Swiss_Prot:P11521	82.43	79.762
STS136	86	DNA-directed RNA polymerase subunit K	Swiss_Prot:P39463	56.06	60.000
STS238	93	DNA-directed RNA polymerase subunit L	Swiss_Prot:P46217	79.08	73.333

ORF ID	Length (a.a.)	Predicted Product	Locus	Zscore	% Identity
ST2454	473	sugar transporter	Swiss_Prot:P37514	27.67	23.739
ST2261	112	DNA-directed RNA polymerase subunit M	Swiss_Prot:Q07271	78.07	68.750
STS221	66	DNA-directed RNA polymerase subunit N	Swiss_Prot:P39472	74.45	77.273
Transcription factors					
ST2114	141	ferric uptake regulator	Swiss_Prot:P44561	25.42	32.308
ST1054	225	sugar fermentation stimulation protein	Swiss_Prot:P18273	26.62	31.000
ST1275	166	TATA-box binding protein	Swiss_Prot:Q55031	95.63	88.957
STS247	54	transcription factor	OWL:ATZ95790	9.90	27.500
ST0349	308	transcription initiation factor IIB	Swiss_Prot:P50387	158.42	86.885
ST1271	295	transcription initiation factor IIB	Swiss_Prot:Q51731	75.90	36.735
ST2236	137	transcription repressor	Swiss_Prot:Q57988	15.00	24.272
ST1369	152	transcription termination-antitermination factor	Swiss_Prot:P27341	107.80	73.026
ST0684	205	transcriptional activator	Swiss_Prot:P25052	24.94	25.000
ST1007	216	transcriptional activator	Swiss_Prot:P25052	72.22	37.559
ST0193	140	transcriptional regulator	Swiss_Prot:P42180	30.77	38.318
ST0489	151	transcriptional regulator	Swiss_Prot:P44337	47.78	40.580
ST0749	123	transcriptional regulator	Swiss_Prot:Q58482	18.34	33.333
ST0980	107	transcriptional regulator	Swiss_Prot:Q58482	14.98	29.474
ST1022	150	transcriptional regulator	Swiss_Prot:Q57615	34.20	30.872
ST1076	128	transcriptional regulator	Swiss_Prot:P37374	31.50	39.506
ST1473	158	transcriptional regulator	Swiss_Prot:Q57615	33.43	29.060
ST1710	146	transcriptional regulator	Swiss_Prot:P40676	20.24	26.119
ST1867	127	transcriptional regulator	Swiss_Prot:Q58898	12.70	29.167
ST1889	109	transcriptional regulator	Swiss_Prot:Q58958	12.48	40.476
ST1890	135	transcriptional regulator	Swiss_Prot:P27245	13.41	30.556
ST2050	128	transcriptional regulator	Swiss_Prot:Q58898	15.18	22.000
ST2084	515	transcriptional regulator	Swiss_Prot:P06153	22.93	34.722
ST2562	213	transcriptional regulator	Swiss_Prot:O05494	49.77	33.668
STS086	76	transcriptional regulator	PIR:E71028	12.55	25.862
STS042	83	transcriptional regulator asnC	PIR:E71108	45.01	49.333
ST0173	116	transcriptional regulator Lrs14	Swiss_Prot:Q58898	13.19	18.269
RNA processing					
ST0343	637	cleavage and polyadenylation specificity factor large subunit	Swiss_Prot:Q58633	176.41	47.961
ST2188	422	cleavage and polyadenylation specificity factor small subunit	Swiss_Prot:Q60355	97.63	36.941
ST0856	123	cryptogene protein G4	PIR:S51910	11.27	26.549
ST0570	339	RNA 3'-terminal phosphate cyclase	Swiss_Prot:O59198	90.01	46.061
STS030	90	small nuclear ribonucleoprotein	PIR:C69186	26.95	41.333
RNA modification					
ST0456	220	rRNA adenine N-6-methyltransferase	Swiss_Prot:P02979	18.83	27.059
Protein synthesis					
tRNA aminoacylation					
ST0167	207	alanyl-tRNA synthetase	Swiss_Prot:Q57984	32.02	34.884
ST1364	904	alanyl-tRNA synthetase	Swiss_Prot:Q57984	122.24	40.547
ST1258	623	arginyl-tRNA synthetase	Swiss_Prot:Q57689	40.55	29.828
ST0205	429	aspartyl-tRNA synthetase	Swiss_Prot:Q58950	87.06	49.420
ST2244	471	cysteinyl-tRNA synthetase	Swiss_Prot:O29836	82.93	42.918
ST1283	476	glutamyl-tRNA (gln) amidotransferase subunit A	Swiss_Prot:P56114	90.54	47.901
ST0282	479	glutamyl-tRNA (gln) amidotransferase subunit B	Swiss_Prot:Q57624	97.74	42.029
STS140	97	glutamyl-tRNA (gln) amidotransferase subunit C	PIR:D70484	34.18	37.500
ST1423	566	glutamyl-tRNA synthetase	Swiss_Prot:P95968	152.28	63.874
ST0353	571	glycyl-tRNA synthetase	Swiss_Prot:Q57681	106.24	42.483
ST0323	426	histidyl-tRNA synthetase	Swiss_Prot:Q58406	73.65	35.071
ST0431	907	isoleucyl-tRNA synthetase	Swiss_Prot:P46215	104.54	76.923
ST0625	945	leucyl-tRNA synthetase	Swiss_Prot:O30250	117.86	33.116
ST1455	894	leucyl-tRNA synthetase	Swiss_Prot:O30250	129.14	37.862
ST2083	444	lysyl-tRNA synthetase	Swiss_Prot:P95970	127.19	82.022
ST1432	571	methionyl-tRNA synthetase	Swiss_Prot:O26687	86.00	40.194
ST1492	113	methionyl-tRNA synthetase	Swiss_Prot:O33925	42.26	57.273
ST1416	469	phenylalanyl-tRNA synthetase alpha subunit	Swiss_Prot:P95961	168.95	58.925
ST1415	540	phenylalanyl-tRNA synthetase beta subunit	Swiss_Prot:O13432	56.91	29.979
ST1440	483	prolyl-tRNA synthetase	Swiss_Prot:Q58635	125.27	57.447
ST1469	453	seryl-tRNA synthetase	Swiss_Prot:O28244	104.51	55.531
ST0966	540	threonyl-tRNA synthetase	Swiss_Prot:P18256	67.54	44.548
ST2187	267	threonyl-tRNA synthetase	Swiss_Prot:Q58597	30.96	36.975
ST0169	406	tryptophanyl-tRNA synthetase	Swiss_Prot:P23381	90.28	50.843
ST2074	294	tyrosyl-tRNA synthetase	Swiss_Prot:P95982	152.01	68.487
ST1224	742	valyl-tRNA synthetase	Swiss_Prot:Q58413	99.10	41.444
Nucleoproteins					
ST2383	182	histone macroH2A.1	Swiss_Prot:Q02874	36.13	32.317
Ribosomal proteins					
ST1424	126	30S ribosomal protein HS6	Swiss_Prot:P55858	90.19	86.508
ST0268	102	30S ribosomal protein S10	Swiss_Prot:P17199	84.86	88.235
ST2068	135	30S ribosomal protein S11	Swiss_Prot:P39469	74.78	92.424
ST1272	147	30S ribosomal protein S12	Swiss_Prot:P11524	83.91	93.662
ST0384	153	30S ribosomal protein S13	Swiss_Prot:P54012	68.52	60.544
ST2070	172	30S ribosomal protein S13	Swiss_Prot:P95986	98.77	75.316
STS059	57	30S ribosomal protein S14	OWL:SSSPCOPER3	72.08	85.185
STS121	63	30S ribosomal protein S15	Swiss_Prot:P21473	9.42	31.250
ST0423	116	30S ribosomal protein S17	Swiss_Prot:O26120	52.88	51.579
STS046	81	30S ribosomal protein S17	PIR:C69207	29.85	37.097
ST0426	140	30S ribosomal protein S19	Swiss_Prot:O26114	64.36	60.902
ST1375	153	30S ribosomal protein S19	Swiss_Prot:O27653	55.25	46.099
ST2063	225	30S ribosomal protein S2	Swiss_Prot:P95993	97.91	72.646
ST0362	120	30S ribosomal protein S24	Swiss_Prot:P14249	22.41	25.773
ST0372	109	30S ribosomal protein S25	Swiss_Prot:P07282	16.24	27.059
STS173	55	30S ribosomal protein S25e	Swiss_Prot:P39938	13.62	45.833
STS051	66	30S ribosomal protein S27	PIR:H69388	24.91	51.163
STS105	66	30S ribosomal protein S27	Swiss_Prot:P54028	45.01	52.542
STS040	83	30S ribosomal protein S28	PIR:C71025	63.04	71.014
ST0424	225	30S ribosomal protein S3	Swiss_Prot:P54034	66.81	45.109
ST0453	193	30S ribosomal protein S3	Swiss_Prot:P54059	77.03	41.530
STS139	55	30S ribosomal protein S30	Swiss_Prot:Q05472	21.71	48.780
ST0421	178	30S ribosomal protein S4	Swiss_Prot:O05634	104.27	64.045
ST2069	177	30S ribosomal protein S4	Swiss_Prot:P39467	98.54	86.093
ST0414	214	30S ribosomal protein S5	Swiss_Prot:O28374	70.50	56.545
ST0382	141	30S ribosomal protein S6	Swiss_Prot:O26360	52.53	48.515
ST0270	194	30S ribosomal protein S7	Swiss_Prot:P17198	91.54	87.629
ST0419	133	30S ribosomal protein S8	Swiss_Prot:O05636	91.30	79.699
ST0188	128	30S ribosomal protein S8e	Swiss_Prot:P54055	66.15	53.488
ST2064	137	30S ribosomal protein S9	Swiss_Prot:P39468	94.65	82.308
ST1367	222	50S ribosomal protein L1	Swiss_Prot:P35024	119.19	77.626
ST2264	179	50S ribosomal protein L10	Swiss_Prot:Q57963	63.32	50.000
ST1368	170	50S ribosomal protein L11	Swiss_Prot:P96037	90.90	78.916
ST1365	108	50S ribosomal protein L12	Swiss_Prot:P08055	69.63	81.905
ST2065	146	50S ribosomal protein L13	Swiss_Prot:P39473	90.87	68.750
ST5087	77	50S ribosomal protein L13e	OWL:RL:613198	12.41	35.088
ST0422	141	50S ribosomal protein L14	Swiss_Prot:P54037	55.34	62.500
STS054	96	50S ribosomal protein L14	Swiss_Prot:P54054	58.34	66.265
ST0412	144	50S ribosomal protein L15	Swiss_Prot:O05643	112.16	75.694
ST0450	215	50S ribosomal protein L15	Swiss_Prot:O58706	68.92	62.941
ST0415	196	50S ribosomal protein L18	Swiss_Prot:O05640	108.23	68.205
ST2066	118	50S ribosomal protein L18	Swiss_Prot:P39474	80.54	74.528
ST0416	151	50S ribosomal protein L19	Swiss_Prot:O05639	82.03	64.667
ST0427	241	50S ribosomal protein L2	Swiss_Prot:O28357	64.39	54.274
ST0459	103	50S ribosomal protein L21	Swiss_Prot:P54013	58.29	55.319
ST0425	156	50S ribosomal protein L22	Swiss_Prot:O28359	57.24	45.395
STS062	82	50S ribosomal protein L23	PIR:E71187	51.01	68.000
STS039	61	50S ribosomal protein L24	PIR:F69345	50.64	64.706
STS060	85	50S ribosomal protein L24	OWL:SSSPCOPER	61.66	73.239
STS061	88	50S ribosomal protein L29	PIR:A71671	13.72	40.000
ST0429	343	50S ribosomal protein L3	Swiss_Prot:P54014	82.36	47.181
ST0274	106	50S ribosomal protein L30	Swiss_Prot:P11522	75.79	57.000
ST0413	158	50S ribosomal protein L30	Swiss_Prot:O05642	95.57	58.861
STS168	88	50S ribosomal protein L31	Swiss_Prot:P38618	78.32	77.907
ST0417	131	50S ribosomal protein L32	Swiss_Prot:O05638	80.11	63.115
STS058	87	50S ribosomal protein L34	Swiss_Prot:O05981	83.84	72.414
STS049	61	50S ribosomal protein L37	PIR:B69186	59.31	73.333
STS063	70	50S ribosomal protein L37	PIR:B71204	35.22	50.746
STS169	51	50S ribosomal protein L39	OWL:RL39_SULSO	53.91	68.000
ST0428	269	50S ribosomal protein L4	Swiss_Prot:O59420	80.56	52.778
STS025	56	50S ribosomal protein L40	Swiss_Prot:O26653	19.98	39.535
STS106	95	50S ribosomal protein L44	OWL:PFMAL3P28	39.27	49.451
ST0420	178	50S ribosomal protein L5	Swiss_Prot:P41202	77.58	75.706
ST0418	186	50S ribosomal protein L6	Swiss_Prot:O05637	95.08	68.280

Table 11. Continued.

ORF ID	Length (a.a.)	Predicted Product	Locus	Zscore	% Identity
ST1472	101	DNA-directed RNA polymerase subunit M	Swiss_Prot:Q58548	18.91	49.091
ST1366	337	acidic ribosomal protein p0	Swiss_Prot:P35023	129.43	73.414
ribosome assemble					
ST2459	337	ATP-dependent RNA helicase deaD	Swiss_Prot:Q58083	44.86	41.390
Ribosomal proteins modification					
ST1505	282	ribosomal protein S6 modification protein	Swiss_Prot:P17116	27.46	26.761
rRNA modification					
ST1266	236	fibrillarin-like pre-rRNA processing protein	Swiss_Prot:P35553	72.79	49.778
tRNA modification					
ST1269	374	N2,N2-dimethylguanosine tRNA methyltransferase	Swiss_Prot:P20300	36.31	55.263
ST0952	413	tRNA nucleotidyltransferase	Swiss_Prot:P77978	172.28	57.282
Translation factors					
ST0269	435	elongation factor 1-alpha	Swiss_Prot:P17196	50.70	88.276
STS027	91	elongation factor 1-beta	Swiss_Prot:Q64214	92.08	79.121
ST0437	724	elongation factor 2	Swiss_Prot:P23112	114.41	87.155
ST0483	340	eukaryotic peptide chain release factor subunit 1	Swiss_Prot:Q58239	86.36	42.905
ST0941	263	translation initiation factor 2 alpha subunit (eIF-2-alp	Swiss_Prot:Q57581	74.40	40.000
ST0516	165	translation initiation factor 2 beta subunit (eIF-2-beta	Swiss_Prot:O58312	44.91	36.296
ST0381	422	translation initiation factor 2 gamma subunit (eIF-2-g	Swiss_Prot:O26361	94.50	55.718
ST0922	362	translation initiation factor 2b alpha subunit (eIF-2b-	Swiss_Prot:O29877	90.70	47.661
ST1296	131	translation initiation factor 5a (eIF-5a)	Swiss_Prot:P28461	104.12	85.496
ST1372	223	translation initiation factor 6 (eIF-6)	Swiss_Prot:P38619	121.36	72.906
ST0811	125	translational inhibitor protein	Swiss_Prot:O58584	72.27	69.106
Protein fate					
Protein and peptide secretion and trafficking					
ST0411	463	preprotein translocase secY subunit	Swiss_Prot:P49978	186.89	75.594
ST1297	356	signal recognition particle protein	Swiss_Prot:O27376	58.51	53.097
ST1370	103	signal recognition particle protein	Swiss_Prot:P27414	68.30	77.670
Protein modification and repair					
ST1293	311	deoxyhypusine synthase	Swiss_Prot:P49366	66.89	35.827
ST1278	254	diphthine synthase	Swiss_Prot:P32469	33.90	34.524
ST1419	297	methionine aminopeptidase	Swiss_Prot:P55963	121.55	63.176
ST0258	167	N-terminal acetyltransferase	Swiss_Prot:P41227	38.83	39.103
ST0940	767	oligosaccharyl transferase	Swiss_Prot:P46977	25.39	19.820
ST0182	211	protein-L-isoaspartate O-methyltransferase	Swiss_Prot:Q57636	44.22	41.358
ST1123	231	protein-L-isoaspartate O-methyltransferase	Swiss_Prot:Q57636	50.07	40.609
Protein folding and stabilization					
ST0836	311	heat shock protein	Swiss_Prot:O30004	43.88	34.831
ST0345	233	peptidyl prolyl cis-trans isomerase	Swiss_Prot:Q58235	59.32	33.810
ST1253	559	thermosome, alpha subunit*	Swiss_Prot:O24734	86.53	100.000
ST0321	552	thermosome, beta subunit*	Swiss_Prot:O24735	81.64	100.000
ST0820	545	thermosome, unidentified subunit	Swiss_Prot:O24734	57.91	58.519
Degradation of proteins, peptides, and glycopeptides					
ST0330	205	26S protease regulatory subunit	Swiss_Prot:O28303	40.38	63.351
ST0779	583	acylamino-acid-releasing enzyme	Swiss_Prot:P13676	32.74	30.288
ST1737	536	acylamino-acid-releasing enzyme	Swiss_Prot:P34422	32.35	30.876
ST1745	565	acylamino-acid-releasing enzyme	Swiss_Prot:P13676	29.26	28.821
ST0623	781	aminopeptidase	Swiss_Prot:P37808	65.86	36.743
ST1746	786	aminopeptidase	Swiss_Prot:P32454	64.30	37.013
ST1650	490	carboxypeptidase	Swiss_Prot:P50848	111.63	35.569
ST1063	292	microsomal dipeptidase	Swiss_Prot:P43477	25.09	31.343
ST0920	310	mucD	Swiss_Prot:P45129	50.10	38.202
ST0363	336	O-sialoglycoprotein endopeptidase	Swiss_Prot:O27476	98.00	47.892
STS197	56	pG1 protein	PIR:A44803	8.99	27.500
ST0890	173	protease I	Swiss_Prot:Q51732	80.70	52.096
ST0446	235	proteasome alpha subunit	Swiss_Prot:O29760	57.69	51.542
ST0324	197	proteasome beta subunit	Swiss_Prot:O29769	52.77	38.462
ST0477	207	proteasome beta subunit	Swiss_Prot:O29769	58.77	45.109
ST2364	1301	pseudomonapepsin	Swiss_Prot:P42790	24.38	34.259
ST2492	1288	pseudomonapepsin	Swiss_Prot:P42790	20.41	32.287
ST2082	360	S2P metalloprotease	Swiss_Prot:O54862	21.81	31.731
ST0894	598	thermopsin	Swiss_Prot:P17118	50.85	35.922
ST2475	988	thermopsin	Swiss_Prot:P17118	44.84	35.565
ST2615	1062	thermopsin	Swiss_Prot:P17118	49.57	32.967
ST2626	788	thermopsin	Swiss_Prot:P17118	59.12	36.162
ST0908	195	tricorn protease	PIR:S73091	106.06	65.730
ST1168	854	tricorn protease	PIR:S73091	321.91	74.070
ST1384	359	X-Pro dipeptidase	Swiss_Prot:P46545	65.89	33.238
ST2323	353	X-Pro dipeptidase	Swiss_Prot:P54518	67.65	38.496
STS190	58	xylX protein	OWL:COU51165	24.26	46.000
Regulatory functions					
Protein interactions					
ST1518	301	cytochrome o ubiquinol oxidase assembly factor	Swiss_Prot:Q04444	53.42	29.286
STS263	58	degradation enzyme regulation protein degQ	Swiss_Prot:P06532	9.71	26.087
ST0575	452	TATA-binding protein-interacting protein	PIR:E72723	120.87	55.479
ST1027	159	urease accessory protein ureE	Swiss_Prot:Q55055	13.25	28.155
ST1026	169	urease accessory protein ureF	Swiss_Prot:P44395	12.46	25.234
ST1025	209	urease accessory protein ureG	Swiss_Prot:Q59066	63.52	53.171
Salvage of nucleosides and nucleotides					
ST2318	460	succinoglycan biosynthesis regulator (exsB)	Swiss_Prot:Q58873	13.03	45.000
RNA stability					
ST0317	603	RNase L inhibitor	Swiss_Prot:Q58129	152.61	52.437
regulation					
ST0872	199	Trp repressor binding protein	Swiss_Prot:P30849	53.14	40.104
Other					
ST2039	211	diphtheria toxin repressor	Swiss_Prot:P33120	31.76	34.848
ST1115	151	leucine-responsive regulatory protein	Swiss_Prot:Q57615	30.42	31.544
ST1587	256	leucine-responsive regulatory protein	Swiss_Prot:Q58133	18.84	38.333
Cell envelope					
Surface structures					
STS112	52	structural protein VP1	Swiss_Prot:P20223	42.96	74.074
Biosynthesis and degradation of surface polysaccharides and lipopolysaccharides					
ST1969	176	dTDP-4-dehydrorhamnose 3,5-epimerase	Swiss_Prot:P37780	68.43	45.882
ST1970	273	dTDP-4-dehydrorhamnose reductase	Swiss_Prot:P26392	33.75	41.104
ST1972	320	dTDP-glucose 4,6-dehydratase	Swiss_Prot:P29782	60.08	41.509
ST0968	228	GDP-mannose pyrophosphorylase	Swiss_Prot:P41940	38.76	35.776
ST0816	419	glucose-1-phosphate nucleotidyltransferase	Swiss_Prot:P39122	26.81	36.424
ST0452	401	glucose-1-phosphate thymidylyltransferase	Swiss_Prot:P08075	41.36	30.030
ST1971	344	glucose-1-phosphate thymidylyltransferase	Swiss_Prot:P08075	95.04	47.222
ST2352	258	glucose-1-phosphate thymidylyltransferase	Swiss_Prot:P08075	30.68	31.944
ST0929	704	glycosyltransferase	Swiss_Prot:Q10768	64.46	32.160
ST2021	415	hyaluronan synthase	Swiss_Prot:Q07755	19.65	30.120
ST2284	357	mannose-1-phosphate guanylyltransferase	Swiss_Prot:P37820	87.41	64.286
ST0242	455	phospho-sugar mutase	Swiss_Prot:Q58500	83.70	34.136
ST0187	306	UDP-glucose 4-epimerase	Swiss_Prot:Q57664	55.98	33.117
ST2057	328	UDP-N-acetylglucosamine--dolichyl-phosphate N-acetylglucosaminephosphotransferase	Swiss_Prot:P39465	176.51	73.171
ST0189	262	undecaprenyl pyrophosphate synthetase	Swiss_Prot:O59258	73.63	44.141
Biosynthesis of membrane ether-linked lipids					
ST2058	329	geranylgeranyl pyrophosphate synthetase	Swiss_Prot:P39464	131.48	65.138
Membranes, lipoproteins, and porins					
ST1164	532	outer membrane protein	OWL:AF005208	94.34	37.129
ST2402	459	outer membrane protein	Swiss_Prot:P77774	13.10	33.333
Other					
ST0366	329	ioiS protein	Swiss_Prot:P54569	31.54	37.821
ST1771	294	transmembrane protein	Swiss_Prot:P48278	34.91	33.133
STS167	89	50S ribosomal protein LX	Swiss_Prot:P38613	65.55	62.353
Cellular processes					
Cell division					
ST1215	369	cell division control protein	Swiss_Prot:Q09803	35.78	42.678
ST1711	587	cell division control protein	Swiss_Prot:Q58556	42.90	38.661
ST2584	700	cell division control protein	Swiss_Prot:Q58556	57.57	49.273
ST0305	398	cell division control protein 6	Swiss_Prot:P29570	61.79	32.486
ST0402	237	centromere/microtubule binding protein cbf5	Swiss_Prot:O30001	73.11	46.383
STS055	84	centromere/microtubule binding protein cbf5	PIR:D71044	33.15	55.738
ST2078	341	DOM34 protein	Swiss_Prot:Q57638	69.54	36.158
ST2367	297	serine/threonine protein phosphatase	Swiss_Prot:P48726	30.84	39.796
Chemotaxis and motility					
ST2518	306	flagellin	Swiss_Prot:P17603	21.86	48.387
ST0155	353	flaG protein	Swiss_Prot:P38094	28.95	29.310
Detoxification					
ST0610	252	cyclohex-1-ene-1-carboxyl-CoA hydratase	Swiss_Prot:P31551	27.25	28.111
ST1638	327	maleylacetate reductase	Swiss_Prot:O30847	35.74	33.654
ST2247	440	NADH oxidase	Swiss_Prot:Q58065	83.40	33.333
ST1285	111	periplasmic divalent cation tolerance protein	Swiss_Prot:P36654	36.26	31.683
ST0584	215	peroxiredoxin	Swiss_Prot:O33665	101.26	78.037
ST2442	212	peroxiredoxin	Swiss_Prot:P95805	100.35	89.151
ST0602	389	serine--pyruvate aminotransferase	Swiss_Prot:P43567	59.76	36.769
ST1217	372	serine--pyruvate aminotransferase	Swiss_Prot:P14776	55.10	30.899
ST2283	211	superoxide dismutase	Swiss_Prot:P80857	86.86	87.560
ST2102	150	thioredoxin peroxidase	Swiss_Prot:P23480	30.62	33.566
Toxin production and resistance					
ST0377	413	3-phosphonopyruvate decarboxylase	Swiss_Prot:Q59007	132.70	46.394
ST1813	264	bacitracin resistance protein	Swiss_Prot:P31054	19.59	26.108
ST1808	248	carboxyvinyl-carboxyphosphonate phosphorylmutase	Swiss_Prot:P54528	99.55	52.846
ST2140	371	isopenicillin N epimerase	Swiss_Prot:P42253	44.77	26.374
ST1075	456	mercuric reductase	Swiss_Prot:P08663	66.15	40.305
STS211	84	rRNA adenine N-6-methyltransferase	Swiss_Prot:P06699	13.16	30.159
DNA transfer					
ST1397	474	virB11 protein	Swiss_Prot:Q58310	52.09	36.493
Adaptations to atypical conditions					
ST1441	336	stress-inducible protein	Swiss_Prot:O26762	78.53	60.204
Cell killing					
ST1374	118	apoptosis protein	Swiss_Prot:Q58103	41.64	45.714
Inhibition of sporulation					
ST2293	221	soj protein	Swiss_Prot:P37522	26.34	30.539
Registant for mutation					
ST0365	189	ham1 protein	Swiss_Prot:Q57679	82.03	47.849
Host cell attaching factor					
STS254	65	sepZ protein	OWL:ECSEPZGEN	13.44	35.593
General					
ST2108	882	purine NTPase	Swiss_Prot:P25386	12.21	19.129
STS205	61	T-cell receptor alpha subunit	OWL:HUMT1414TR	7.71	30.769
Other categories					
Plasmid functions					
ST0849	202	pepK protein	PIR:T00006	44.90	53.247
ST0913	166	pepK protein	Swiss_Prot:O58354	12.65	28.125
Transposon functions					
ST0547	264	extragenic suppressor protein suhB	Swiss_Prot:Q57573	29.88	27.273
Insertion element-related functions					
ST1908	316	insertion element DNA-binding protein	Swiss_Prot:P19780	19.25	26.596
ST0142	242	insertion element protein	Swiss_Prot:P03829	22.33	33.750
ST0671	265	insertion element protein	Swiss_Prot:P03829	21.75	32.500
ST1082	244	insertion element protein	Swiss_Prot:P03829	22.04	32.500
ST1092	205	insertion element protein	Swiss_Prot:Q60329	73.32	41.748
ST1139	244	insertion element protein	Swiss_Prot:P03829	22.22	33.750
ST1156	100	insertion element protein	Swiss_Prot:P03829	20.79	32.500
ST1907	244	insertion element protein	Swiss_Prot:P03829	22.30	33.750
ST2015	100	insertion element protein	Swiss_Prot:P03829	23.03	32.500
ST0306	102	integrase	Swiss_Prot:P20214	24.01	52.941
ST0043	136	transposase	PIR:D72205	79.29	58.065
ST0680	136	transposase	PIR:D72205	77.16	57.258
ST0854	386	transposase	Swiss_Prot:P19780	21.76	35.897
ST1048	136	transposase	PIR:D72205	80.00	59.677
ST1096	136	transposase	PIR:D72205	79.02	58.871
ST1158	136	transposase	PIR:D72205	76.48	58.065
ST1904	136	transposase	PIR:D72205	78.96	58.871
ST2430	331	transposase	Swiss_Prot:Q56897	17.87	36.842
ST2553	136	transposase	PIR:D72205	79.29	58.065
Neurotransmission					
ST1411	400	kynurenine/alpha-aminoadipate aminotransferase	Swiss_Prot:P39389	53.55	25.191
Unknown function					
ST1651	1136	antibiotic transport-associated protein	Swiss_Prot:Q10773	26.52	26.144
ST1678	842	antibiotic transport-associated protein	Swiss_Prot:Q10773	23.11	20.809
ST1914	837	antibiotic transport-associated protein	Swiss_Prot:Q10773	32.19	32.620
ST0656	223	arylmalonate decarboxylase	Swiss_Prot:Q05115	14.00	23.853
ST0721	154	bacterioferritin comigratory protein	Swiss_Prot:P55979	37.54	44.762
ST2117	584	bpa2 protein	Swiss_Prot:P32985	153.63	51.568
ST0404	401	clock-controlled gene-9	Swiss_Prot:Q11152	13.95	35.211
ST0303	307	cytosolic acyl-CoA thioester hydrolase	Swiss_Prot:P49851	33.59	27.660
STS077	64	DNA-binding protein 7e	Swiss_Prot:P13125	64.48	85.246
ST0948	293	elaC protein	Swiss_Prot:Q58897	63.98	39.350
ST0451	400	GTP-binding protein	Swiss_Prot:P42942	69.59	36.413
ST0520	359	GTP-binding protein	Swiss_Prot:P32235	28.28	44.444
ST0557	525	GTP-binding protein	Swiss_Prot:Q17045	94.43	41.148
ST1449	265	GTP-binding protein	Swiss_Prot:P53742	38.36	33.205
ST1491	334	GTP-binding protein	Swiss_Prot:O25505	16.69	32.370
ST0332	350	GTP-binding protein hflX	Swiss_Prot:P25519	39.63	30.769
ST1471	141	histidine triad nucleotide-binding (HINT) protein	Swiss_Prot:Q58276	61.98	54.206
ST2152	150	histidine triad nucleotide-binding (HINT) protein	Swiss_Prot:P47378	44.44	36.923
ST0976	399	interferon-gamma inducible protein	Swiss_Prot:P39651	38.66	31.298
ST1421	395	interferon-gamma inducible protein	Swiss_Prot:P39651	44.15	30.417
ST1983	422	interferon-gamma inducible protein	Swiss_Prot:P39651	39.59	34.137
ST2164	112	M protein	Swiss_Prot:P50468	7.05	29.670
ST1550	230	methyltransferase	Swiss_Prot:P49016	20.00	25.131
ST1810	431	mmge protein	Swiss_Prot:P45859	30.13	29.032
ST2179	348	mu-crystallin	Swiss_Prot:O54983	39.46	29.286
ST2334	311	mu-crystallin	Swiss_Prot:Q28488	33.46	29.907
STS185	85	nitrogen fixation protein nifW	OWL:AP1470557	10.92	21.795
ST1265	409	nucleolar protein	Swiss_Prot:Q58105	97.46	48.910
ST0174	298	nucleotide-binding protein	Swiss_Prot:Q57731	83.93	52.256
ST2404	294	Pitrh	Swiss_Prot:P46852	23.74	50.980
ST0440	125	protein KE2	Swiss_Prot:Q58304	45.02	45.370
ST1752	317	repeat motif-containing gene protein	Swiss_Prot:Q28758	14.88	17.416
ST0209	689	SAV protein	Swiss_Prot:Q07590	72.35	65.103
ST0376	747	SAV protein	Swiss_Prot:Q07590	104.29	82.329
ST0548	605	SAV protein	Swiss_Prot:Q07590	48.49	40.182
ST0059	462	selenium-binding protein	Swiss_Prot:Q63836	103.36	40.678
ST2554	674	sip1 protein	PIR:S45033	26.70	32.850
STS093	79	small nuclear ribonucleoprotein	PIR:C69186	46.29	54.688
ST0630	318	veg136 protein	Swiss_Prot:Q58873	58.14	35.621
ST1453	322	veg136 protein	Swiss_Prot:Q58558	64.00	36.601
ST1943	133	virulence-associated protein	Swiss_Prot:P71363	26.41	28.846
STS174	75	virus coat protein	OWL:D50472	6.97	30.909

The columns from left to right represent the ORF ID, length of amino acid residues of ORFs, probable gene products, locus of genes which showed the highest similarity to corresponding ORFs, Zscore taken from Smith-Waterman similarity search, and percent identity between the reported genes and assigned ORFs.

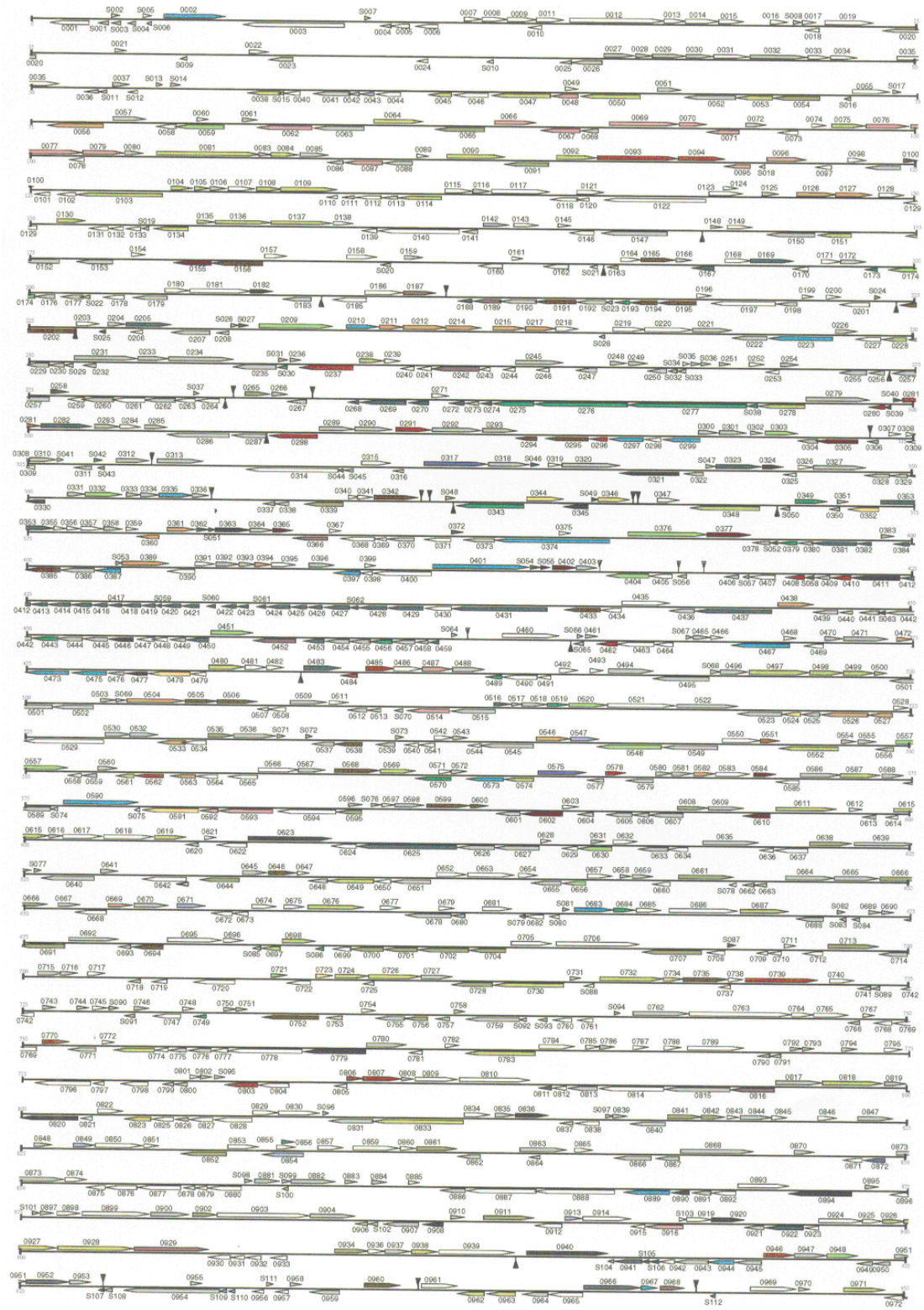

Figure 1. Gene map of the *Sulfolobus tokodaii* strain7 genome. A total of 2826 putatively identified potential protein-coding regions are shown, and the direction of transcription is indicated by arrows. Each line in the figure represents 100,000 bp of sequence in the *S. tokodaii* strain7 genome. Positions are given by numbers above or below the tic marks in each row. The ORFs are color coded by role category as described in the key. ORF ID numbers placed above or below the ORFs correspond to those in Table 11. The rRNA operon and tRNA genes, are labeled.

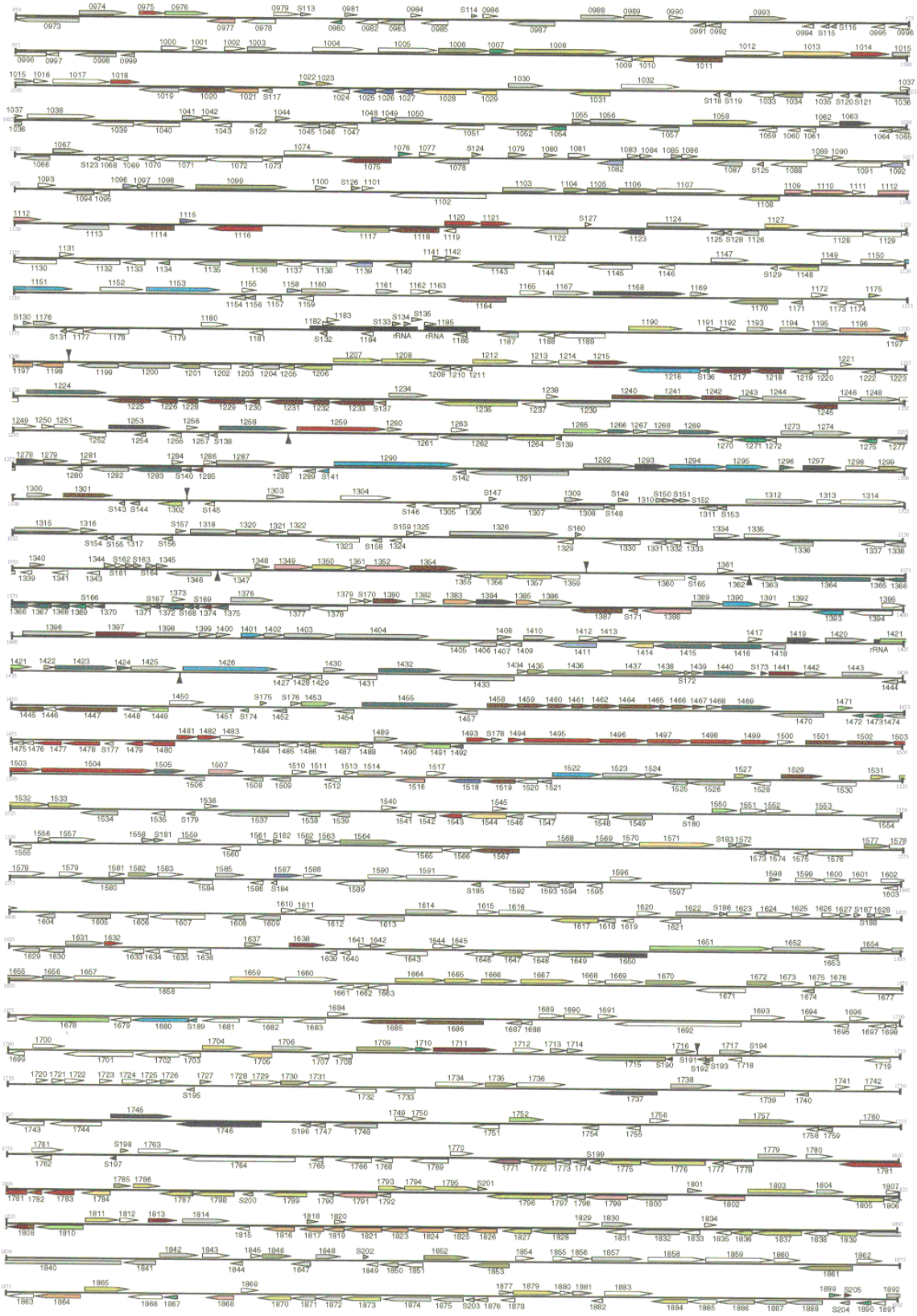

Figure 1. Continued.

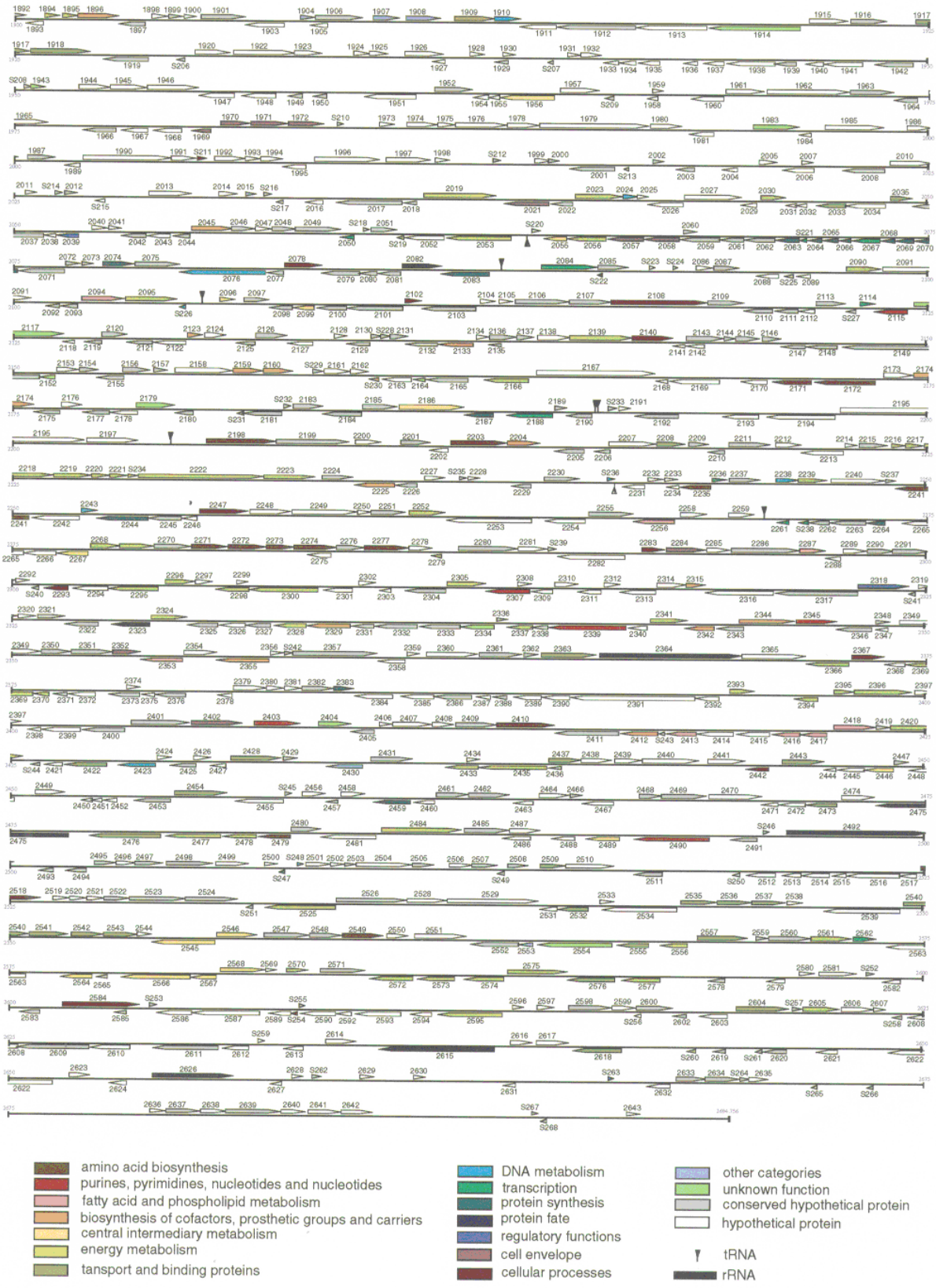

Figure 1. Continued.