

## Complete Sequence and Gene Organization of the Genome of a Hyper-thermophilic Archaeobacterium, *Pyrococcus horikoshii* OT3

Yutaka KAWARABAYASI,<sup>1,2,\*</sup> Mitsuhiro SAWADA,<sup>1</sup> Hiroshi HORIKAWA,<sup>1</sup> Yuji HAIKAWA,<sup>1</sup> Yumi HINO,<sup>1</sup> Saori YAMAMOTO,<sup>1</sup> Mitsuo SEKINE,<sup>1</sup> Sin-ichi BABA,<sup>1</sup> Hiroki KOSUGI,<sup>1</sup> Akira HOSOYAMA,<sup>1</sup> Yoshimi NAGAI,<sup>1</sup> Mari SAKAI,<sup>1</sup> Keiko OGURA,<sup>1</sup> Rie OTSUKA,<sup>1</sup> Hidekazu NAKAZAWA,<sup>1</sup> Minako TAKAMIYA,<sup>1</sup> Yuhko OHFUKU,<sup>1</sup> Tomomichi FUNAHASHI,<sup>1</sup> Toshihiro TANAKA,<sup>1</sup> Yutaka KUDOH,<sup>1</sup> Jun YAMAZAKI,<sup>1</sup> Norihiro KUSHIDA,<sup>1</sup> Akio OGUCHI,<sup>1</sup> Ken-ichi AOKI,<sup>1</sup> Takio YOSHIZAWA,<sup>1</sup> Yoshinobu NAKAMURA,<sup>1</sup> Frank T. ROBB,<sup>3</sup> Koki HORIKOSHI,<sup>4</sup> Yaeko MASUCHI,<sup>5</sup> Hiroaki SHIZUYA,<sup>6</sup> and Hisasi KIKUCHI<sup>1</sup>

National Institute of Technology and Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan,<sup>1</sup> National Institute of Bioscience and Human-Technology, 1-1 Higashi, Tsukuba, Ibaraki 305-0046, Japan,<sup>2</sup> Center of Marine Biotechnology, Columbus Center, University of Maryland, Biotechnology Institute, Baltimore, MD, USA,<sup>3</sup> Japan Marine Science and Technology Center, 2-15 Natsushima-cho, Yokosuka, Kanagawa 237-0061, Japan,<sup>4</sup> University of Tokyo, Meguro, Tokyo 153-0041, Japan,<sup>5</sup> and California Institute of Technology, Pasadena, CA, USA<sup>6</sup>

(Received 2 March 1998)

### Abstract

The complete sequence of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3, has been determined by assembling the sequences of the physical map-based contigs of fosmid clones and of long polymerase chain reaction (PCR) products which were used for gap-filling. The entire length of the genome was 1,738,505 bp. The authenticity of the entire genome sequence was supported by restriction analysis of long PCR products, which were directly amplified from the genomic DNA. As the potential protein-coding regions, a total of 2061 open reading frames (ORFs) were assigned, and by similarity search against public databases, 406 (19.7%) were related to genes with putative function and 453 (22.0%) to the sequences registered but with unknown function. The remaining 1202 ORFs (58.3%) did not show any significant similarity to the sequences in the databases. Sequence comparison among the assigned ORFs in the genome provided evidence that a considerable number of ORFs were generated by sequence duplication. By similarity search, 11 ORFs were assumed to contain the intein elements. The RNA genes identified were a single 16S-23S rRNA operon, two 5S rRNA genes and 46 tRNA genes including two with the intron structure. All the assigned ORFs and RNA coding regions occupied 91.25% of the whole genome. The data presented in this paper are available on the internet at <http://www.nite.go.jp>.

**Key words:** hyper-thermophilic archaeobacterium; *Pyrococcus horikoshii* OT3; genome sequencing; fosmid; open reading frames; shotgun sequencing

### 1. Introduction

*Pyrococcus horikoshii* OT3, which was isolated in 1992 from a hydrothermal vent at a depth of 1395 m in the Okinawa Trough in the Pacific Ocean, is a hyper-thermophilic, anaerobic archaeobacterium that grows at temperatures ranging from 85 °C to 105 °C and op-

timally at 98 °C.<sup>1</sup> The entire genome sequences have been reported for three thermophilic archaeobacteria, *Methanococcus jannaschii* (1.66 Mb),<sup>2</sup> *Methanobacterium thermoautotrophicum* (1.75 Mb)<sup>3</sup> and *Archaeoglobus fulgidus* (2.18 Mb),<sup>4</sup> but the optimal growth temperatures of these bacteria are 85 °C, 65 °C and 83 °C, respectively. It is surprising that *P. horikoshii* OT3 optimally grows at a temperature of nearly 100 °C. Thus, the sequence data of the entire genome of this microorganism should provide valuable information not only for understanding of the living mechanisms at such a high temperature, but also for the industrial application of the highly thermostable gene products.

Communicated by Mituru Takanami

\* To whom correspondence should be addressed. Tel. +81-3-3481-8951, Fax. +81-3-3481-8424, E-mail: kyutaka@kazusa.or.jp

† The entire genome sequence has been deposited in the DDBJ/Genbank/EMBL databases under the accession numbers AP000001-AP000007.

In this study, we constructed a fine restriction map of the *P. horikoshii* OT3 genome, using three kinds of restriction enzymes, and the clones from a fosmid library were allocated to construct the contig map. The nucleotide sequences of these ordered fosmid clones were then determined by the shotgun strategy followed by gap filling. In this way, the entire nucleotide sequence of the genome was deduced.

## 2. Materials and Methods

### 2.1. Bacterial Strains and preparation of genomic DNA

The *P. horikoshii* OT3 strain deposited in the Japan Collection of Microorganisms (JCM number 9974) was used for genome sequencing in this study. The JCM 151 pyrococcus medium was prepared under nitrogen atmosphere in screw-capped 1000-ml bottles, and the cells were cultured at 98 °C by standing. The genomic DNA was isolated principally based on the method of Maede et al.<sup>5</sup> *Escherichia coli* DH10B was used for the preparation of fosmid and plasmid clones.

### 2.2. Construction of a genomic library by fosmid vector

For construction of the fosmid library, DNA from the whole cell was prepared in the agarose-plug as described by Smith et al.,<sup>6</sup> partially digested with *Hind* III within the plug and resolved by 0.4% agarose gel electrophoresis. The DNA fraction with sequences approximately 40 kb in length was separated and directly inserted into the *Hind* III site of the fosmid cloning vector pBAC108L.<sup>7</sup> After packaging into bacteriophage  $\lambda$  capsids by an *in vitro* packaging system (Stratagene, La Jolla, CA, USA), the products were transfected into *E. coli* strain DH10B. A total of 1248 transformants were isolated, grown in microtiter wells and stocked.

### 2.3. DNA sequencing

The DNA of fosmid clones, which were purified by CsCl-EtBr density-gradient ultracentrifugation, were sonicated to two different extents. This was followed by size-fractionation with agarose gel electrophoresis, and the fractions 0.8–1.2 kb and 2.0–2.5 kb were cloned into the *Hinc* II site of pUC118. The plasmid DNA was prepared by the Autogen 740 automatic DNA preparation system (Autogen, Framingham, MA, USA). For the sequence determination of smaller gaps remaining within a fosmid clone, the polymerase chain reaction (PCR) products which were directly amplified from the genomic DNA using the flanking sequences were sequenced by the dye-terminator method. For closure of larger gaps which were not covered by any fosmid clone, long PCR was performed with primer sets designed from the flanking sequences, and the amplified DNA was shotgun-sequenced. The sequencing reaction was performed using two kinds of cycle sequencing kits, dye-primer cycle sequencing kits

and dye-terminator cycle sequencing kits, and ABI-DNA sequencers; 373XL and 377XL (Perkin-Elmer ABI, Foster City, CA, USA).

### 2.4. Data assembly and computational analysis

The accumulated files of single-run sequence data (raw data) were first treated by software Phred<sup>8,9</sup> to refine the basecalling and to remove the sequences derived from any contaminated host and vector DNAs. The treated data were then assembled into contigs by software Phrap and edited by Sequencher (Gene Codes, Ann Arbor, MI, USA). The construction of the entire genomic sequence was also done by Sequencher. The criteria used for assignment of potential coding regions on the genomic sequence are as follows. All the ORFs larger than 100 sense codons starting with ATG or GTG were first assigned, and then ORFs with 50–99 sense codons were taken, if the sequences showed some similarity to the coding sequences in the databases or to the protein motif sequences. Similarity search of assigned ORFs was performed by the Smith-Waterman algorithm.<sup>10</sup> The databases used for similarity search were Genbank release 103, EMBL release 53, Swiss-Prot release 43, PIR release 43 and owl release 32.

## 3. Results and Discussion

### 3.1. Construction of physical and contig maps

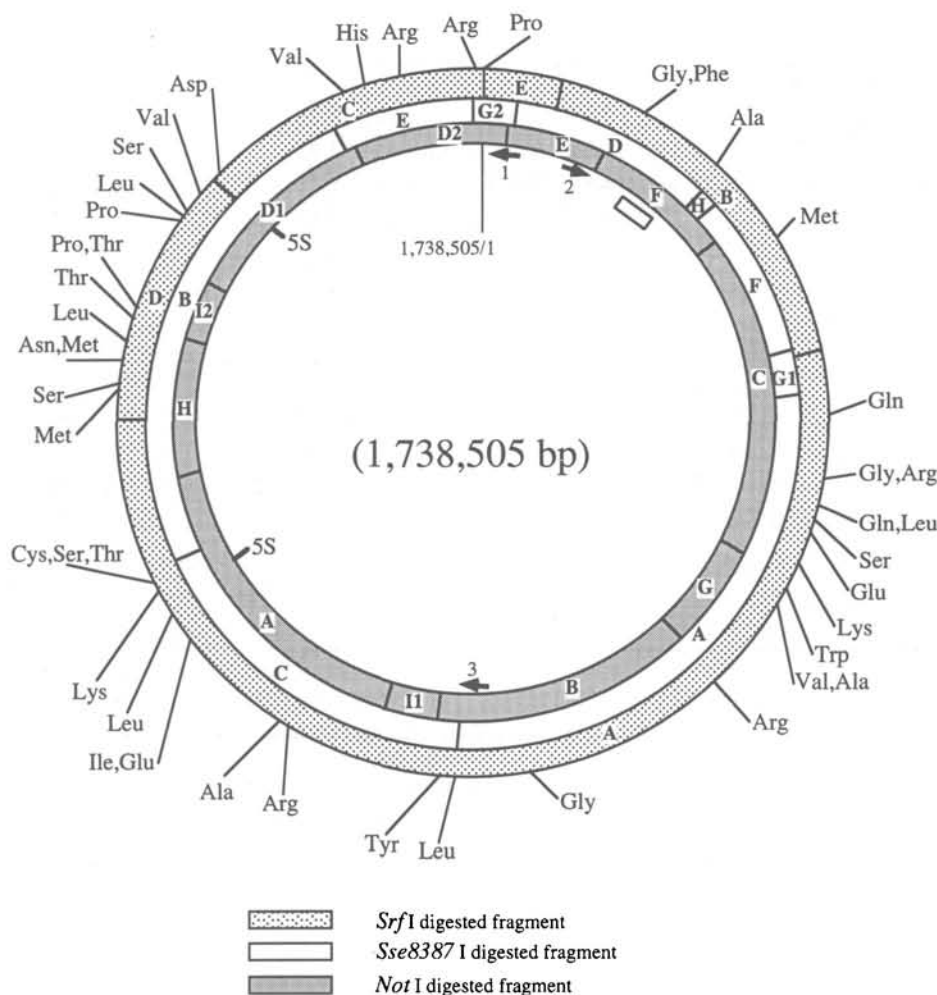
The genomic DNA of *P. horikoshii* OT3 was digested with restriction endonucleases *Not* I, *Srf* I and *Sse*8387 I, which generated 11, 3 and 9 fragments, respectively, and the resulting fragments were resolved and purified by the CHEF Mapper gel electrophoresis system (BioRad, Richmond, CA, USA). Then the relative order of the fragments was determined by hybridization analysis among the fragments and their double-digestion products.

Figure 1 shows the physical map which was finally confirmed by the determination of the entire genomic sequence. It was noted that, although *Srf* I generated three fragments, the sequence data indicated the presence of five *Srf* I sites. Since the fosmid DNAs containing the two undigested sites were easily cleaved by *Srf* I, it is likely that the nucleotides within these sites are modified in the living cells.

Based on the physical map of the genome constructed, fosmid clones were assigned on the restriction fragments by hybridization with the clone-specific tag sequences. Over 90% of the whole genome could be covered by the fosmid clones leaving eight gaps.

### 3.2. Determination of entire genome sequence

The entire genome sequence was determined by assembling the sequences of overlapped fosmid clones and long PCR products which were used for analysis of the gaps. For the sequence deduction of the respective clones,



**Figure 1.** The physical map of the *P. horikoshii* OT3 genome and location of RNA genes and characteristic repeating units. Restriction sites for it *Srf* I, *Sse8387* I and *Not* I are indicated in the circular map of the genome. The nucleotide positions were numbered from the junction of *Srf* I-C and -E fragments in the clockwise direction. The three clusters of repeating units [a consensus 29-bp sequence (GTTTCCGTAGAACTNAPTAGTGYSKAAAK) plus a variable 34–58 bp sequence] are indicated by wide arrows with numerals. Number of repeats in each site are: site 1, 18 times; site 2, 26 times; and site 3, 66 times. The position of the 16S-23S rRNA gene cluster is shown by a open box, and the arrows with 5S indicate the 5S rRNA genes. The positions and species of tRNA are indicated by bars with the names of chargeable amino acids.

a modified shotgun method was used.<sup>11</sup> Two shotgun plasmid libraries, one with approximately 900-bp inserts (S1 clones) and the other with about 2-kbp inserts (S2 clones), were directly prepared from a fosmid clone. The S1 clones were single-run sequenced from one end and the S2 clones from both ends. For most of the fosmid clones with 40-kb inserts, a single contiguous sequence was obtained by analysis of approximately 600 S1 and 400 S2 clones, which was approximately 10 times the size of the insert. The assembled sequences were confirmed by additional sequencing which excluded any ambiguity and also by sequencing of either the complementary strand or the same strand but using both the dye-primer and dye-terminator methods.

To confirm the authenticity of the genomic sequence

constructed, every 15–20 kb region of the genome was directly amplified from the genomic DNA by long PCR, and the occurrence of restriction sites were compared with those deduced from the sequence data.

The total length of the genome, which was constructed from 61 fosmid clones and eight long PCR fragments, was 1,738,505 bp. The nucleotide position was numbered from the one end of the *Srf* I cutting site on the *Sse8387* I G2 fragment, as shown in the physical map of Fig. 1. The distribution of ACGT along the strand of the entire genome was 29.0% A, 21.2% C, 20.7% G and 29.1% T, resulting in a GC content of 41.9%.

Through the processes of physical mapping and genomic sequencing, no evidence revealing the presence of extra-chromosomal units was obtained. Thus, we con-

**Table 1.** Condon-usages of putative protein coding regions (a) and summary of tRNA identified by open circles (b).

<b>(a)</b>											
UUU (Phe)	23.89	13632	UCU (Ser)	10.16	5795	UAU (Tyr)	19.69	11238	UGU (Cys)	3.11	1774
UUC (Phe)	21.98	12541	UCC (Ser)	9.69	5529	UAC (Tyr)	18.59	10607	UGC (Cys)	3.17	1808
UUA (Leu)	21.49	12262	UCA (Ser)	11.56	6596	UAA (End)	1.23	702	UGA (End)	1.62	924
UUG (Leu)	11.77	6717	UCG (Ser)	4.57	2607	UAG (End)	0.76	435	UGG (Trp)	11.7	6678
CUU (Leu)	25.62	14618	CCU (Pro)	12.14	6929	CAU (His)	8.01	4571	CGU (Arg)	1.21	693
CUC (Leu)	17.52	9996	CCC (Pro)	10.19	5814	CAC (His)	6.85	3910	CGC (Arg)	1.07	613
CUA (Leu)	18.17	10367	CCA (Pro)	17.48	9977	CAA (Gln)	7.91	4517	CGA (Arg)	1.00	566
CUG (Leu)	8.61	4913	CCG (Pro)	5.10	2910	CAG (Gln)	8.35	4765	CGG (Arg)	0.92	526
AUU (Ile)	25.93	14796	ACU (Thr)	13.58	7748	AAU (Asn)	18.44	10522	AGU (Ser)	10.33	5896
AUC (Ile)	17.09	9752	ACC (Thr)	10.70	6105	AAC (Asn)	16.79	9581	AGC (Ser)	12.07	6890
AUA (Ile)	44.52	25405	ACA (Thr)	10.56	6023	AAA (Lys)	31.17	17784	AGA (Arg)	20.11	11476
AUG (Met)	22.92	13079	ACG (Thr)	10.13	5782	AAG (Lys)	45.99	26242	AGG (Arg)	30.08	17164
GUU (Val)	35.48	20243	GCU (Ala)	21.55	12299	GAU (Asp)	30.89	17628	GGU (Gly)	15.13	8635
GUC (Val)	11.27	6430	GCC (Ala)	16.29	9298	GAC (Asp)	11.61	6622	GGC (Gly)	7.15	4077
GUA (Val)	17.51	9991	GCA (Ala)	18.41	10504	GAA (Glu)	40.74	23248	GGA (Gly)	31.53	17990
GUG (Val)	12.06	6879	GCG (Ala)	7.24	4130	GAG (Glu)	41.91	23915	GGG (Gly)	15.67	8943

<b>(b)</b>											
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)	○				

cluded that the *P. horikoshii* OT3 cell maintains a single chromosomal genetic unit.

### 3.3. Assignment and similarity search of potential protein-coding regions

The ORFs which are longer than 100 sense codons starting with ATG or GTG were first searched, and named with a two-letter code (PH) which represents the species name plus a number of four figures indicating the ORF position from the first nucleotide position. The total number of this class of ORFs assigned was 2002

including overlapped ORFs in the same or either strand.

Next, shorter ORFs 50–99 codons long were extracted from the regions where no ORFs longer than 100 codons were assigned. The amino acid sequences of these shorter ORFs were similarity searched against the public databases using the Smith-Watermann algorithm and motif sequences using the GCG software package (Genetic Computer Group, Madison, WI, USA). The shorter ORFs, which possessed significant similarity to registered sequences or motifs were taken as the probable protein-coding regions, and named with a three code (PHS) plus a number of three figures indicating the ORF position.

Table 2. ORFs identified intein elements.

ORF ID	putative products	entire ORF length (amino acids)	N-terminal positions*	C-terminal positions*	length of inteins (amino acids)	length of mature protein (amino acids)
PH0112	Activator 1 subunit	855	62	586	525	330
PH0263	DNA repair protein	529	153	324	172	357
PH0363	Ribonucleoside-diphosphate reductase	1291	468	852	385	906
PH0606	Cell division control protein	1107	335	502	168	679
			530	789	260	
PH0800	Reverse Gyrase	1624	954	1363	410	1214
PH1095	Translation initiation factor	1044	23	466	444	600
PH1110	Helicase	1351	347	821	475	876
PH1602	Hypothetical protein	871	98	487	390	481
PH1947	DNA-directed DNA polymerase	1235	493	952	460	775
PH1975	ATP synthase	964	241	616	376	588
PH0650	Hypothetical protein **	1136				

\*: Amino acid positions from the assigned initiation codons.

\*\* : Both terminal have not been assigned.

There were a total of 2061 ORFs thus assigned along the entire genomic sequence. Of these, 2002 were PH-class ORFs and 59 were PHS-class ORFs. The average size of the ORFs was 272 amino acid residues, and the longest one was 4436 residues (PH0954). The assigned 2061 ORFs occupy 90.72% of the whole genome DNA. It should be noted that the ORFs mentioned above merely represent the protein-coding potentiality under the defined assumptions.

For the anticipation of gene function, the products with a ZScore of higher than 40 or with over 30% identity in the amino acid sequence along the entire coding region were considered PH-class ORFs, and only those with a ZScore of higher than 20 for the PHS-class ORFs.

Of the ORFs assigned, 406 (19.70%) were related to the genes with functions and 453 (21.98%) showed significant similarity to the registered sequences having unknown functions. The remaining 1202 ORFs showed no significant similarity to the sequences in public databases. In the last group of ORFs, 199 (9.66%) just contained some motifs. The proportion of the anotatable ORFs is lower than that shown in previously sequenced genome, supporting the uniqueness of this hyper-thermophilic organism.

### 3.4. RNA coding genes

For assignment of the rRNA genes, the entire genomic sequence was similarity searched against the rRNA sequences registered in the databases. A single 16S-23S rRNA cluster and two 5S rRNA genes were assigned, as shown in Fig. 1. Such an organization is similar to that of *Archaeoglobus fulgidus*, but differs from that of two other archaeobacteria, *Methanoc. jannaschii* and *Methanob. thermoautotrophium*, each possessing two rRNA operons.

The tRNA genes were searched by tRNAscan,<sup>12</sup> and 46 tRNA genes were identified. The two tRNA genes,

Met and Trp, contained 71-bp and 31-bp introns at the anticodon region. Twenty-nine tRNA genes were mapped discretely, while the remaining 17 tRNA genes were mapped as clusters of two or three tRNA genes, suggesting that the genes in these clusters are transcribed in the same transcriptional units.

The species of tRNA genes identified are shown in Table 1. As noted in other microorganisms,<sup>13</sup> no tRNA genes containing A at the first position of the anticodon were discovered. Interestingly, the Met-tRNA gene was present in triplicate in the genome, suggesting that this may enable the effective initiation of translation at high temperature.

It is noteworthy that the sequences of Pro-tRNAs with CCG- and TGG-anticodons, Leu-tRNAs with CAG-, TAG- and GAG-anticodons, Gly-tRNA with CCC- and TCC-anticodons, and Ala-tRNAs with TGC- and GGC-anticodons were respectively identical except for one or two nucleotides at the anticodon regions. In contrast, the sequences of other tRNA species transferring the same amino acids were considerably different from each other.

All the corresponding tRNA synthetase genes have been identified in the ORFs assigned to the genome, except for those of Gln-tRNA and Trp-tRNA.

### 3.5. Other features noted

#### 3.5.1. ORFs containing inteins

Self-splicing portions of a polypeptide sequence that generally encode the characteristic motifs of endonucleases are called inteins,<sup>14,15</sup> and the presence of 33 inteins has been reported in archaeobacterial genes.<sup>2,3,16-20</sup> To investigate the possible occurrence of ORFs containing inteins in the genome of *P. horikoshii* OT3, the sequences of all the ORFs were similarity searched against the intein-specific motif sequences described by Pietrokovski<sup>21</sup> and Perler et al.,<sup>22</sup> and also against the amino acid sequences in the reported inteins using the

**Table 3.** Homologous ORFs identified within the 2002 ORFs assigned on the genome.

number of homologous ORFs*	number of groups	total ORFs
9	1	9
6	2	12
5	2	10
4	5	20
3	7	21
2	72	144

\*; ORFs with ZScore higher than 10 in SW search and amino acid identity higher than 30% in 80% of the entire region were taken.

GCG package. As a result, ten ORFs presumably containing inteins were identified, in which nine ORFs contained a single intein but the one named PH0606 appeared to contain two inteins. One more ORF (PH0650) was assumed to contain the intein by motif sequence search but no homologous gene was detected. The N- and C-terminal positions, the lengths of inteins and mature products are shown in Table 2.

The intein elements have been identified in the DNA polymerase gene of various organisms. The element was also found in the polymerase gene of *P. horikoshii* OT3. In addition, the inteins were identified in four other genes involved in DNA replication (DNA-repair protein, ribonucleoside-diphosphate reductase, reverse gyrase and helicase). Other notable features found by sequence comparison are as follows: (1) The putative products of PH0263, PH0606 and PH1975 were functionally related to the gene families of DNA-repair protein, cell-division control protein and ATP synthase, respectively, the intein elements of which have not been reported yet, (2) the intein sequence in the DNA polymerase gene of *P. horikoshii* OT3 is highly conserved in those of two other *Pyrococcus* strains (*P. spp.* KOD, 65.9% identity; *P. spp.* GB-D, 55.1% identity), and (3) the inteins in ORFs PH0112, PH0800 and PH1602 have been conserved in the corresponding genes in *Methanoc. jannaschii*.

### 3.5.2. Duplication of ORFs

By sequence comparison among the ORFs assigned to the genome, 216 ORFs could be allocated into 89 groups, with ORFs in each group coding for proteins with considerable similarity (Table 3). In the largest group, nine ORFs showed similarity to a single ORF. These results suggest 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 as repeats of a single or a cluster of ORFs at different locations. This is probably due to the differences in size of the duplication units.

### 3.5.3. Unique repeating sequences

The presence of unique repetitive sequences was noted by similarity search, while no IS- or Tn-like elements were identified. The repeating unit is composed of a well-conserved 29-bp sequence followed by a variable 34- to 58-bp sequence. Sequences which were repeated 18, 26 and 66 times were found at the three positions indicated in Fig. 1. This type of high-frequency repeat has also been reported in *Methanoc. jannaschii*, but their biological significance is not known.

**Acknowledgments:** We thank Drs. S. Tabata and H. Kotani and their staff at Kazusa DNA Research Institute for their valuable suggestions and help. Thanks are due to Drs. Itaru Watanabe for encouragement to carry out this project, D. Söll for kind help of assignment of Lys-tRNA synthetase and P. Green for donation of the assembling software Phred-Phrap. This work was supported by the Ministry of International Trading and Industry.

### References

- González, J. M., Masuchi, Y., Robb, F. T. et al. 1998, *Pyrococcus horikoshii* sp. nov., a Hyperthermophilic Archaeon Isolated from a Hydrothermal Vent at the Okinawa Trough, *Extremophiles*, in press.
- Bult, C. J., White, O., Olsen, G. J. et al. 1996, Complete Genome Sequence of the Methanogenic Archaeon, *Methanococcus jannaschii*, *Science*, **273**, 1058–1073.
- 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.
- 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.
- Meade, H. M., Long, S. R., Ruvkun, C. B., Brown, S. E., and Ausubel, F. M. 1982, Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn5 mutagenesis, *J. Bacteriol.*, **149**, 114–122.
- Smith, C. L., Econome, J. G., Schutt, A., Klco, A., and Cantor, C. R., 1987, A Physical Map of the *Escherichia coli* K12 Genome, *Science*, **236**, 1448–1453.
- Shizuya, H., Birren, B., Kim, U. et al. 1992, Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector, *Proc. Natl. Acad. Sci. USA*, **89**, 8794–8797.
- Ewing, B., Hillier, L., Wendl, M. C., and Green, P. 1998, Base-calling of Automated Sequencer Traces Using phred I. Accuracy Assessment, *Genome Res.*, in press.
- Ewing, B. and Green, P. 1998, Base-calling of Automated Sequencer Traces Using phred II. Error Probabilities, *Genome Res.*, in press.
- Smith, T. F. and Waterman, M. S. 1981, Identification of Common Molecular Subsequences, *J. of Mol. Biol.*, **147**, 195–197.

11. Kaneko, T., Tanaka, A., Sato, S. et al. 1995, Sequence Analysis of the Genome of the Unicellular Cyanobacterium *Synechocystis* sp. strain PCC6803. I. Sequence Features in the 1 Mb region from Map Positions 64% to 92% of the Genome, *DNA Res.*, **2**, 153–166.
12. Lowe, T. M. and Eddy, S. R. 1997, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, *Nuc. Acids Res.*, **25**, 955–964.
13. 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.
14. 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<sup>+</sup>-Translocating Adenosine Triphosphatase from Vacuolar Membranes of *Saccharomyces cerevisiae*, *J. Biol. Chem.*, **265**, 6726–6733.
15. Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebel, M., and Stevens, T. H. 1990, Protein Splicing Converts the Yeast TFP1 Gene Product to the 69-kD Subunit of the Vacuolar H<sup>+</sup>-Adenosine Triphosphatase, *Science*, **250**, 651–657.
16. Riera, J., Robb, F. T., Weiss, R., and Fontecave, M. 1990, Ribonucleotide reductase in the archaeon *Pyrococcus furiosus*: a critical enzyme in the evolution of DNA genomes?, *Proc. Natl. Acad. Sci. USA*, **94**, 475–478.
17. Xu, M., Southworth, M. W., Mersha, F. B., Hornstra, L. J., and Peler, F. B. 1993, In vitro protein splicing of purified precursor and the identification of a branched intermediate, *Cell*, **75**, 1371–1377.
18. Takagi, M., Nishioka, M., Kakihara, H. et al. 1997, Characterization of DNA polymerase from *Pyrococcus* sp. strain KOD1 and its application to PCR, *Appl. Environ. Microbiol.*, **63**, 4504–4510.
19. Peler, F. B., Comb, D. G., Jack, W. E. et al. 1992, Intervening sequences in an Archaea DNA polymerase gene, *Proc. Natl. Acad. Sci. USA*, **89**, 5577–5581.
20. Niehaus, F., Frey, B., and Antranikian, G. 1997, Cloning and characterisation of a thermostable alpha-DNA polymerase from the hyperthermophilic archaeon *Thermococcus* sp. TY, *Gene*, **204**, 153–158.
21. Pietrovski, S. 1994, Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins, *Protein Science*, **3**, 2340–2350.
22. Perler, F. B., Olsen, G. J., and Adams, E. 1997, Compilation and analysis of intein sequences, *Nucl. Acids Res.*, **25**, 1087–1093.

Table 4. List of ORFs showing similarities to registered genes with known and unknown function.

ORF ID	Position		Direction	Length (aa)	Predicted Product	Locus	ZScore	% Identity
	Initiation	Termination						
<b>ORFs with known function</b>								
PH0001	620	126	<	164	50S ribosomal protein L11	Swiss_Prot:P54030	96.13	70.63
PH0003	2,451	1,333	<	372	cell division protein FtsZ	owl:PWU56247	128.34	88.46
PH0011	9,327	8,476	<	283	nicotinate-nucleotide pyrophosphorylase	PIR:E64361	90.19	49.26
PH0013	11,032	10,130	<	300	quinolinate synthetase	PIR:G64350	126.5	56.71
PH0015	13,692	12,298	<	464	L-aspartate oxidase	Swiss_Prot:P38032	56.57	37.74
PH0017	17,172	15,523	<	549	thermophilic factor	PIR:S61294	120.41	89.44
PH0022	21,944	20,823	<	373	sugar-binding transport ATP-binding protein	owl:D8302640	54.64	48.75
PH0037	36,487	35,249	<	412	phosphonopyruvate decarboxylase	PIR:C64501	148.07	50.25
PH0039	37,239	36,523	<	238	modification methylase	PIR:C64370	98.24	54.87
PH0051	45,539	44,478	<	353	dolichol-phosphate mannosyltransferase	PIR:S76171	96.08	41.5
PH0052	46,259	45,576	<	227	fibrillar-like pre-rRNA processing protein	Swiss_Prot:P35553	97.54	54.38
PH0054	48,542	47,556	<	328	1-aminocyclopropane-1-carboxylate deaminase	owl:D9083216	76.84	41.1
PH0056	49,774	49,082	>	230	HESA protein	Swiss_Prot:P46048	62.07	44.04
PH0058	50,393	50,869	>	158	40S ribosomal protein S13	Swiss_Prot:P54012	92.43	70.67
PH0060	52,559	53,158	>	199	30S ribosomal protein S3a	Swiss_Prot:P54059	72.3	41.58
PH0066	57,594	56,608	<	328	L-asparaginase	Swiss_Prot:P26900	102.74	40.37
PH0075	64,024	63,497	<	175	cob(I)alamin adenosyltransferase	Swiss_Prot:P13040	51.97	33.14
PH0078	64,949	66,034	>	361	D-aminopeptidase	owl:OADAMP	85.96	39.65
PH0092	75,247	77,712	>	821	phosphoenolpyruvate synthase	Swiss_Prot:P42850	308.83	93.12
PH0112	93,837	96,404	>	855	replication factor C subunit	PIR:E64477	195.56	60.86
PH0114	98,754	97,804	<	316	molybdenum cofactor biosynthesis protein	PIR:H64402	98.58	51.47
PH0125	112,344	111,562	>	260	5'-methylthioadenosine phosphorylase	owl:I38969	93.52	53.36
PH0138	122,734	124,134	>	466	4-aminobutyrate aminotransferase	Swiss_Prot:P22256	75.19	36.47
PH0144	127,050	126,079	<	323	magnesium and cobalt transport protein	owl:S77053	71.41	42.04
PH0147	128,777	129,484	>	235	biotin-[acetyl-CoA-carboxylase] ligase	Swiss_Prot:P42975	46.88	36.25
PH0149	130,375	129,470	<	301	2-phosphoglycerate kinase	PIR:S41582	89.22	44.21
PH0178	158,101	157,421	<	226	glutaredoxin-like protein	PIR:S54843	188.04	88.5
PH0182	161,046	160,273	<	257	NH(3)-dependent NAD(+) synthetase	PIR:G64468	89.22	52.28
PH0184	162,054	162,503	>	149	molybdopterinconverting factor, subunit	owl:MTCY16429	46.01	38.53
PH0191	167,600	168,169	>	189	L-fucose phosphate aldolase	PIR:A64477	58.15	39.33
PH0193	168,536	170,437	>	633	alpha-amylase	Swiss_Prot:P49067	269.64	65.23
PH0194	170,498	171,625	>	375	multiple sugar-binding transport ATP-binding protein	owl:D8302640	64.36	57.03
PH0201	177,248	176,049	<	399	26S protease regulatory subunit	PIR:G64446	71.09	61.58
PH0203	178,407	177,319	<	362	maltose/maltodextrin transport ATP-binding protein	owl:D8302640	44.84	50.7
PH0208	184,467	183,493	<	324	translation initiation factor eIF-2B delta	PIR:B64315	124.11	57.72
PH0211	187,327	186,485	<	280	spermidine synthase	owl:BSTRZ6	66.42	33.82
PH0224	198,383	199,954	>	523	lysyl-tRNA synthetase	personal communication		
PH0232	206,485	207,402	>	305	L-asparaginase	owl:ECAE0001852	60.72	39.72
PH0239	212,901	212,185	<	238	phosphoribosylaminoimidazole-succinocarboxamide synthase	PIR:G64498	102.44	52.81
PH0240	212,958	214,307	>	449	amidophosphoribosyltransferase	PIR:E64325	103.57	56.94
PH0241	215,619	214,315	<	434	asparaginyl-tRNA synthetase	Swiss_Prot:P39772	121.79	47.24
PH0243	220,654	218,849	<	601	glutamine--fructose-6-phosphate aminotransferase	PIR:S69793	110.66	48.4
PH0245	222,384	221,791	<	197	proteasome beta subunit	PIR:D64454	56.34	46.32
PH0246	222,295	223,809	>	504	tddd protein	PIR:D64424	122.37	44.16
PH0263	232,354	233,943	>	529	DNA repair protein	PIR:S71093	75.06	58.71
PH0265	234,006	235,241	>	411	desuccinylase	PIR:A64357	108.33	43.81
PH0290	258,463	257,168	<	431	histidyl-tRNA synthetase	PIR:D64424	110.19	45.85
PH0292	258,631	259,827	>	398	5-aminolevulinic acid synthase	owl:BI0F_BACSH	81.35	42.64
PH0297	263,038	265,785	>	915	alanyl-tRNA synthetase	PIR:D64370	238.88	52.31
PH0302	267,332	268,474	>	380	Na(+)/H(+) antiporter	owl:BMU17283	99.90	37.63
PH0307	271,723	273,183	>	486	inosine-5'-monophosphate dehydrogenase	Swiss_Prot:P42851	183.44	93.2
PH0312	278,716	280,581	>	621	phosphoenolpyruvate carboxykinase	owl:MTCY8D55	60.24	34.51
PH0314	280,715	283,405	>	896	valyl-tRNA synthetase	PIR:F64425	188.98	51.86
PH0316	284,406	283,402	<	334	phosphoribosylformylglycinamide cyclo-ligase	PIR:D64325	107.86	55.74
PH0318	288,363	289,664	>	433	phosphoribosylglycinamide formyl transferase	PIR:E64485	137.56	55.18
PH0320	290,186	289,653	<	177	phosphoribosylaminoimidazole carboxylase catalytic subunit	owl:SSOUREK	79.47	68
PH0323	291,856	293,172	>	438	phosphoribosylamide--glycine ligase	PIR:A64417	149.93	55.48
PH0359	320,076	321,407	>	443	glutamine synthetase	Swiss_Prot:P36687	142.60	84.31
PH0365	328,925	327,942	<	327	galactose-1-phosphate uridylyltransferase	owl:MTCY19H51	47.64	33.92
PH0366	328,980	330,251	>	423	beta-glucosidase	owl:TSAL6BGLT	185.06	64.49
PH0369	333,695	332,643	<	350	galactokinase	Swiss_Prot:P31767	69.66	36
PH0376	337,949	338,800	>	283	cobalamin biosynthesis protein D	PIR:A64464	75.18	47.34
PH0378	340,177	341,133	>	318	UDP-glucose 4-epimerase	owl:S44960	65.97	40.65
PH0413	376,565	377,635	>	356	glucose-1-phosphate thymidyltransferase	owl:SAMTMDE	111.17	53.89
PH0414	377,637	378,647	>	336	dTDP-glucose 4,6-dehydratase	owl:SERORF11	73.87	50.16
PH0416	378,735	379,301	>	188	dTDP-4-dehydrorhamnose 3,5-epimerase	PIR:S23343	84.21	63.42
PH0417	379,302	380,174	>	290	dTDP-4-dehydrorhamnose reductase	PIR:S74471	61.98	40.95
PH0421	382,263	383,561	>	432	lipopolysaccharide O-side chain biosynthesis protein	owl:YEU468597	40.52	21.45
PH0429	389,697	391,064	>	455	Vi polysaccharide biosynthesis protein	owl:PAU503961	117.78	47.24
PH0438	400,946	401,965	>	339	adenylosuccinate synthetase	owl:PSPURAGEN	216.54	91.42
PH0439	401,997	402,764	>	255	competence-damage protein	owl:SS56KBFR18	72.56	45.46
PH0440	402,767	403,597	>	276	translation initiation factor eIF-2B	PIR:B64315	56.43	35.34



Table 4. Continued.

ORF ID	Position		Direction	Length (aa)	Predicted Product	Locus	ZScore	% Identity
	Initiation	Termination						
PH0443	405,411	406,697	>	428	methyl-accepting chemotaxis protein	Swiss_Prot:P42258	45.16	27.42
PH0452	412,627	416,010	>	1127	ATP-dependent protease La	owl:H64476	66.29	43.68
PH0465	425,561	427,108	>	515	thermostable carboxypeptidase	Swiss_Prot:P50848	133.26	38.48
PH0470	432,512	433,408	>	298	membrane protein	PIR:C64403	82.75	55.08
PH0478	437,064	436,627	<	145	purine-binding chemotaxis protein cheW	PIR:S61835	43.73	38.06
PH0481	438,529	439,371	>	280	chemotaxis protein methyltransferase	owl:BACVARGNS6	58.33	39.35
PH0482	439,364	439,726	>	120	chemotaxis protein	owl:TMU305014	70.30	71.3
PH0483	439,744	440,850	>	368	protein-glutamate methyltransferase	owl:D9083010	88.91	42.48
PH0484	440,813	443,113	>	766	histidine kinase (cheA)	owl:TMU305012	107.17	47.06
PH0491	444,856	447,075	>	739	chemotaxis protein	Swiss_Prot:P42259	44.05	28.32
PH0501	454,358	452,907	<	483	beta-mannosidase	owl:PFU60214	191.57	60.59
PH0503	456,471	457,463	>	330	dipeptide transport system permease protein AppB	owl:MMABCTRAN1	57.71	34.78
PH0504	457,467	458,381	>	304	oligopeptide transport system permease protein AppC	owl:MMABCTRAN2	62.34	40
PH0505	458,378	459,352	>	324	oligopeptide transport ATP-binding protein AppF	Swiss_Prot:P42065	58.20	45.76
PH0507	459,336	460,301	>	321	oligopeptide transport ATP-binding protein AppF	Swiss_Prot:P42065	51.20	44.9
PH0511	461,279	463,615	>	778	beta-galactosidase precursor	Swiss_Prot:P23780	40.73	47.46
PH0520	469,786	468,785	<	333	dehydrogenase	PIR:A64427	60.55	38.64
PH0540	483,446	484,723	>	425	S-adenosyl-L-homocysteine hydrolase	PIR:C64473	126.61	64.25
PH0546	489,011	489,724	>	237	flagellin B precursor	PIR:D64411	73.10	55.05
PH0548	489,757	490,380	>	207	flagellin B precursor	PIR:D64411	81.27	52.94
PH0549	490,395	491,399	>	334	flagellin B precursor	PIR:D64411	51.73	62.14
PH0550	491,456	492,100	>	214	flagellin B precursor	owl:MVU76620	60.30	46.6
PH0551	492,111	492,878	>	255	flagellin B precursor	PIR:D64411	47.63	58.25
PH0570	510,221	508,785	<	478	pyruvate kinase	owl:A57418	73.75	76.5
PH0572	510,518	511,855	>	445	NADH oxidase	PIR:A64381	97.78	38.52
PH0577	515,046	513,697	<	449	damage-inducible protein	Swiss_Prot:P54181	63.03	25.29
PH0582	520,714	519,506	<	402	molybdopterin biosynthesis moea protein	PIR:F64410	98.43	49.19
PH0596	534,899	535,519	>	206	pyrrolidone-carboxylate peptidase	owl:TLPCP	104.60	57.64
PH0597	536,669	535,539	<	376	dehydrogenase	owl:BSZ9404317	85.71	49.82
PH0605	545,067	544,645	<	140	translation initiation factor eIF-2 beta	PIR:A64312	73.39	51.11
PH0606	548,436	545,110	<	1108	cell division control protein	owl:CC54_YEAST	54.59	41.54
PH0612	550,505	549,768	<	245	cell division inhibitor MinD	PIR:C64368	80.34	41.38
PH0615	552,663	553,040	>	125	30S ribosomal protein S6E	Swiss_Prot:P54067	57.71	64.55
PH0622	557,792	559,852	>	686	DNA topoisomerase I	PIR:B64506	182.28	52.52
PH0628	564,217	565,104	>	295	methionyl aminopeptidase	PIR:H64465	116.79	50.35
PH0632	567,990	568,799	>	269	Gufa protein	Swiss_Prot:Q06916	46.92	42.94
PH0633	568,877	569,422	>	181	ubiquinol-cytochrome c reductase complex subunit VI	PIR:G64367	77.80	54.72
PH0635	569,817	570,581	>	254	coenzyme PQQ synthesis protein III	PIR:C64505	44.01	41.7
PH0636	572,307	570,877	<	476	cysteinyl-tRNA synthetase	Swiss_Prot:Q06752	129.74	48.71
PH0640	574,562	573,342	<	406	phosphate permease	Swiss_Prot:P45268	99.07	41.67
PH0643	576,460	577,152	>	230	2-hydroxyhepta-2,4-diene-1,7-dioate isomerase	PIR:F64506	95.69	57.22
PH0655	587,615	588,661	>	348	dehydrogenase	Swiss_Prot:P07913	87.17	44.51
PH0657	590,501	588,831	<	556	phenylalanyl-tRNA synthetase alpha chain	owl:C64438	148.80	44.75
PH0658	592,012	590,513	<	499	phenylalanyl-tRNA synthetase beta chain	PIR:G64360	91.68	41.84
PH0664	594,104	594,436	>	110	transcription-associated protein	PIR:A55263	86.70	65.14
PH0670	598,071	597,385	<	228	aspartate racemase	owl:D84067	148.87	70.98
PH0678	603,707	604,264	>	185	ferredoxin oxidoreductase gamma subunit	owl:PFORVOR5	155.79	90.27
PH0679	604,298	604,615	>	105	ferredoxin oxidoreductase delta subunit	owl:PFORVOR6	103.01	86.67
PH0680	604,617	605,810	>	397	ferredoxin oxidoreductase alpha subunit	owl:PFORVOR7	234.65	90.61
PH0681	605,807	606,751	>	314	ferredoxin oxidoreductase beta subunit	owl:PFORVOR8	226.14	93.57
PH0682	606,773	607,099	>	108	ferredoxin oxidoreductase delta subunit	owl:PFORVOR9	116.88	95.24
PH0684	607,106	608,302	>	398	oxidoreductase	owl:PFORVOR10	227.29	88.1
PH0685	608,304	609,308	>	334	ferredoxin oxidoreductase beta subunit	owl:PFORVOR11	233.63	92.15
PH0687	610,111	612,633	>	840	cell division control protein	PIR:C64444	98.15	68.55
PH0698	622,928	623,410	>	160	nucleoside-diphosphate kinase	PIR:H64457	61.84	63.5
PH0699	623,445	625,322	>	625	threonyl-tRNA synthetase	PIR:D64449	185.06	51.07
PH0702	628,476	627,382	<	364	translation initiation factor eIF-2 beta chain	PIR:S76472	113.83	48.16
PH0710	633,585	632,203	<	460	seryl-tRNA synthetase	Swiss_Prot:P37464	97.28	44.55
PH0720	642,629	643,555	>	308	aspartate carbamoyltransferase catalytic chain	owl:PAU61765	141.90	92.53
PH0721	643,556	644,014	>	152	aspartate carbamoyltransferase regulatory chain	owl:PAU617651	140.53	91.45
PH0722	645,170	644,004	<	388	amino acid amidohydrolase	PIR:S76858	102.36	45.7
PH0725	646,028	646,825	>	265	diphthine synthase	owl:PFDPH5ARG	126.56	83.85
PH0726	646,955	647,908	>	317	ornithine carbamoyltransferase	owl:PFARGF	145.30	96.49
PH0739	656,198	655,188	<	336	hydrogenase expression/formation protein	PIR:D64384	131.53	59.64
PH0743	656,991	658,118	>	375	dehydrogenase	owl:ADH1_CLOAB	86.50	40.39
PH0754	665,452	666,393	>	313	sugar transport system permease protein	PIR:S77536	49.51	35.55
PH0756	667,229	668,302	>	357	sugar transport ATP-binding protein	owl:S76829	53.57	55.83
PH0765	675,061	673,115	<	648	indolepyruvate synthase alpha subunit	owl:D86221	308.31	80.37
PH0768	677,627	676,857	<	256	SOJ protein	PIR:G64512	106.88	52.17
PH0769	678,912	677,668	<	414	cell division protein FtsZ	owl:PWU56247	73.09	50
PH0771	679,800	680,975	>	391	aspartate aminotransferase	PIR:A64300	85.4	48.1
PH0781	686,317	687,897	>	526	alanine transport protein	Swiss_Prot:P44917	78.73	42.2
PH0782	687,897	689,321	>	474	4-aminobutyrate aminotransferase	Swiss_Prot:P30268	79.27	36.02

Table 4. Continued.

ORF ID	Position		Direction	Length (aa)	Predicted Product	Locus	ZScore	% Identity
	Initiation	Termination						
PH0790	698,750	699,745	>	331	hemin permease	owl:HPAE0005986	68.47	38.77
PH0800	713,261	708,387	<	1624	reverse gyrase	owl:PFU66557	307.36	85.5
PH0807	716,178	717,971	>	597	oligopeptide transport permease protein APPA	Swiss_Prot:P42061	63.16	33.55
PH0808	718,013	718,984	>	323	oligopeptide transport permease protein APPB	Swiss_Prot:P42062	67.95	35.87
PH0809	718,996	719,847	>	283	oligopeptide transport permease protein APPC	Swiss_Prot:P42063	67.24	38.85
PH0810	719,848	720,831	>	327	oligopeptide transport ATP-binding protein APPD	owl:RSPNGR2342	51.81	43.69
PH0811	720,828	721,799	>	323	oligopeptide transport ATP-binding protein APPF	Swiss_Prot:P42065	54.85	47.48
PH0819	729,956	728,377	<	459	oxidative cyclase	Swiss_Prot:P26168	59.06	29.22
PH0834	744,793	743,078	<	571	oxaloacetate decarboxylase alpha chain	PIR:F64453	150.49	56.15
PH0835	744,910	747,480	>	856	alpha-mannosidase	PIR:S76045	78.83	31.67
PH0847	755,505	754,642	<	287	dihydrodipicoline synthase	Swiss_Prot:P39359	77.19	34.39
PH0851	760,113	758,761	<	450	fmu protein	PIR:S74920	57.75	34.46
PH0852	761,536	760,184	<	450	adenylosuccinate lyase	PIR:A64416	161.97	58.86
PH0857	765,691	764,507	<	394	threonine synthase	PIR:H64482	154.58	62.94
PH0864	773,677	773,051	<	208	transcription initiation factor IIB	Swiss_Prot:P29095	99.90	55.56
PH0876	784,634	783,216	<	472	glutamate synthase small chain	owl:D86223	204.25	76.97
PH0879	785,959	787,062	>	367	hydrogenase expression/formation protein hypD	PIR:A64424	110.99	46.13
PH0884	790,846	789,854	<	330	malic acid transport protein	PIR:B64395	126.14	48.15
PH0896	800,923	801,810	>	295	cation efflux system protein czcD	owl:BSU9387614	79.83	38.83
PH0897	801,902	804,223	>	773	transcriptional regulatory protein hypF	PIR:A64389	161.30	48.03
PH0909	818,103	820,073	>	656	DNA-binding protein	PIR:H64312	164.12	48.15
PH0923	832,997	831,630	<	455	phospho-sugar mutase	PIR:C64437	93.95	41.84
PH0925	834,396	833,002	<	464	mannose-1-phosphate guanylyl transferase	owl:XANB_XANCP	88.85	49.78
PH0949	855,866	856,753	>	295	MRP protein	PIR:D64335	96.86	55.64
PH0951	858,381	857,530	<	283	3-methyl-2-oxobutanoate hydroxymethyltransferase	Swiss_Prot:P52996	89.55	44.71
PH0961	876,080	875,253	<	275	translation initiation factor eIF-2 alpha chain	PIR:E64314	114.57	52.69
PH0965	881,508	878,605	<	967	leucyl-tRNA synthetase	PIR:A64379	241.55	53.99
PH0974	886,997	885,927	<	356	dipeptidase	owl:YQHT_BACSU	83.53	37.22
PH0978	891,481	892,197	>	238	50S ribosomal protein L15	Swiss_Prot:P54060	114.35	70.1
PH0985	897,298	898,014	>	238	arylmalonate decarboxylase	Swiss_Prot:Q05115	40.61	30.81
PH0993	904,490	902,319	<	723	methionyl-tRNA synthetase	PIR:F64457	135.62	46.91
PH0999	907,656	908,198	>	180	pyrazinamidase/nicotinamidase	owl:AF0082061	48.80	39.88
PH1006	915,499	916,941	>	480	prolyl-tRNA synthetase	owl:E64454	169.20	52.06
PH1009	917,594	918,169	>	191	TATA-binding protein	PIR:A54275	130.87	92.67
PH1011	918,718	919,845	>	375	tyrosyl-tRNA synthetase	owl:SS56KBF43	103.84	45.2
PH1014	921,363	920,818	<	181	3-octaprenyl-4-hydroxybenzoate carboxylase	PIR:F64312	87.61	54.29
PH1017	923,651	922,530	<	373	cofactor modifying protein	owl:PF64312	203.54	65.17
PH1020	927,215	925,899	<	438	aspartyl-tRNA synthetase	PIR:JC4352	221.30	83.79
PH1022	928,531	927,281	<	416	sugar-phosphate nucleotidyl transferase	Swiss_Prot:P41940	59.46	33.43
PH1026	931,902	932,864	>	320	membrane dipeptidase	owl:LMPONGEN	51.63	37.02
PH1032	940,401	941,393	>	330	modification methylase (adenine-specific)	PIR:S34414	93.61	49.06
PH1065	972,211	969,011	<	1066	isoleucyl-tRNA synthetase	Swiss_Prot:P46214	195.98	93.96
PH1072	977,251	978,279	>	342	geranylgeranyl pyrophosphate synthetase	owl:S75695B	92.81	49.23
PH1085	986,546	985,311	<	411	flavoprotein	PIR:D64391	59.31	32.39
PH1088	988,661	989,665	>	334	aspartate-semialdehyde dehydrogenase	PIR:F64325	116.36	50
PH1093	993,750	992,635	<	371	cystathionine gamma-lyase	owl:HPAE0005325	71.06	40.77
PH1103	1,003,561	1,003,178	<	127	30S ribosomal protein S8	Swiss_Prot:P54055	77.30	62.99
PH1110	1,011,546	1,007,488	<	1352	ATP-dependent helicase LHR	owl:G64336	169.28	59.48
PH1117	1,017,043	1,016,381	<	220	adenylate kinase	Swiss_Prot:P16304	66.00	42.86
PH1128	1,024,298	1,024,858	>	186	uridine 5'-monophosphate synthase	PIR:D64438	86.47	57.56
PH1132	1,028,367	1,027,087	<	426	GTP-binding protein hflX	PIR:E64440	101.06	48.53
PH1138	1,033,293	1,031,437	<	618	indolepyruvate ferredoxin oxidoreductase alpha subunit	owl:D86221	159.67	48.41
PH1143	1,039,359	1,040,156	>	265	5'-methylthioadenosine phosphorylase	PIR:D64307	87.31	48.79
PH1146	1,042,727	1,041,522	<	401	aminomethyltransferase	Swiss_Prot:P54378	97.80	44.54
PH1149	1,044,427	1,043,372	<	351	X-Pro dipeptidase	Swiss_Prot:P54518	93.38	43.19
PH1171	1,061,321	1,059,945	<	458	endoglucanase	Swiss_Prot:P19487	107.68	47.75
PH1205	1,087,359	1,085,374	<	661	ferrous iron transport protein B	PIR:F64370	173.64	42.33
PH1210	1,089,286	1,090,647	>	453	phospho-sugar mutase	PIR:C64437	115.35	44.67
PH1217	1,096,033	1,096,683	>	216	alkyl hydroperoxide reductase	PIR:H64391	132.30	76.39
PH1218	1,096,766	1,097,998	>	410	phosphoglycerate kinase	Swiss_Prot:P50316	161.47	78
PH1236	1,112,100	1,111,069	<	343	transport permease protein	PIR:G64310	136.06	58.94
PH1253	1,129,120	1,127,516	<	534	modulation protein	PIR:A64432	54.28	45
PH1262	1,134,266	1,136,119	>	617	prolyl endopeptidase	PIR:JC4084	315.32	85.55
PH1274	1,144,459	1,146,336	>	625	ferredoxin oxidoreductase	owl:TLFORPKPF2	123.51	89.73
PH1275	1,147,642	1,146,323	<	439	malate oxidoreductase (NAD)	owl:AF008220168	119.24	51.62
PH1277	1,148,721	1,147,639	<	360	malate dehydrogenase	Swiss_Prot:P16142	97.46	45.23
PH1282	1,152,623	1,153,567	>	314	carbamate kinase	owl:PF64310	193.73	89.1
PH1283	1,155,087	1,153,885	<	400	methylmalonyl-CoA decarboxylase beta chain	PIR:E49094	144.86	67.01
PH1284	1,155,543	1,155,094	<	149	methylmalonyl-CoA decarboxylase gamma chain	PIR:D49094	40.66	43.55
PH1287	1,157,444	1,155,876	<	522	methylmalonyl-CoA decarboxylase alpha chain	PIR:A49094	123.41	60.32
PH1290	1,159,144	1,160,262	>	372	cytochrome-c3 hydrogenase alpha chain	PIR:S48833	264.28	88.77
PH1291	1,160,259	1,161,137	>	292	cytochrome-c3 hydrogenase gamma chain	PIR:S48834	236.52	94.48
PH1292	1,161,139	1,161,939	>	266	cytochrome-c3 hydrogenase delta chain	PIR:S48835	211.98	89.45

Table 4. Continued.

ORF ID	Position		Direction	Length (aa)	Predicted Product	Locus	ZScore	% Identity
	Initiation	Termination						
PH1294	1,161,936	1,163,225	>	429	cytochrome-c3 hydrogenase alpha chain	PIR:S48836	265.00	86.22
PH1295	1,164,066	1,165,343	>	425	proton glutamate symport protein	owl:BSY1408315	68.84	37.47
PH1306	1,175,580	1,177,271	>	563	methylmalonyl-CoA mutase precursor	owl:MUTA_HUMAN	117.32	53.46
PH1308	1,178,546	1,179,706	>	386	serine aminotransferase	PIR:G64419	79.89	34.64
PH1317	1,185,950	1,185,534	<	138	glycine cleavage system H protein	Swiss_Prot:P23884	59.32	55.46
PH1322	1,189,236	1,190,441	>	401	aspartate aminotransferase	owl:HIU3271411	83.18	39.52
PH1325	1,191,458	1,191,910	>	150	30S ribosomal protein S19	Swiss_Prot:P54057	90.58	64.87
PH1346	1,213,446	1,212,877	<	189	GMP synthase	PIR:F64496	95.39	60.11
PH1347	1,214,380	1,213,454	<	308	GMP synthase	owl:B64441	130.32	67.33
PH1349	1,215,514	1,216,557	>	347	thiamine-binding periplasmic protein precursor	Swiss_Prot:P31550	66.54	37.37
PH1350	1,217,533	1,216,541	<	330	ATP-binding transport protein	Swiss_Prot:P37009	44.89	51.65
PH1353	1,221,139	1,219,100	<	679	formate dehydrogenase	owl:MTU526812	124.80	48.55
PH1355	1,222,832	1,221,825	<	335	ethylene-responsive protein	PIR:E64384	162.71	74.3
PH1357	1,223,981	1,223,214	<	255	thiamine biosynthetic enzyme	PIR:A64375	102.19	50.6
PH1361	1,227,738	1,229,351	>	537	proline permease	owl:BSU924662	76.39	35.58
PH1364	1,230,916	1,232,064	>	382	sarcosine oxidase	Swiss_Prot:P40875	53.78	29.75
PH1366	1,233,097	1,234,023	>	308	ATP-binding transport protein	owl:D785084	40.48	40.37
PH1371	1,238,357	1,237,188	<	389	aspartate aminotransferase	PIR:C33496	80.18	40.9
PH1374	1,241,951	1,241,004	<	315	proliferating-cell nucleolar protein p120	owl:A48998	66.30	42.91
PH1375	1,242,654	1,241,965	<	229	ribose 5-phosphate isomerase	PIR:S76567	99.59	50.47
PH1384	1,248,480	1,249,223	>	247	ATP-dependent transporter YCF16	Swiss_Prot:P48255	65.48	48.51
PH1387	1,253,202	1,252,279	<	307	phosphoglycerate dehydrogenase	PIR:A64427	85.61	55.96
PH1397	1,260,438	1,259,410	<	342	deoxyhypusine synthase	PIR:F64401	192.75	73.41
PH1402	1,262,896	1,263,519	>	207	proteasome beta subunit precursor	PIR:D64454	68.71	49.24
PH1406	1,265,487	1,266,809	>	440	threonine synthase	PIR:H64482	78.41	41.53
PH1410	1,270,034	1,271,032	>	332	dipeptide transport system permease protein dppB	owl:HPAE0005485	81.05	40.12
PH1411	1,271,044	1,272,201	>	385	dipeptide transport system permease protein dppC	Swiss_Prot:P37315	53.00	44.09
PH1412	1,272,185	1,273,174	>	329	dipeptide transport ATP-binding protein dppD	owl:RSPNGR2342	67.32	52.08
PH1413	1,273,176	1,274,165	>	329	dipeptide transport ATP-binding protein dppF	Swiss_Prot:P42065	63.44	53.61
PH1415	1,274,675	1,275,706	>	343	endonuclease	PIR:C64480	121.49	55.11
PH1423	1,281,315	1,279,951	<	454	4-aminobutyrate aminotransferase	owl:D5045389	77.24	38.54
PH1426	1,282,422	1,281,412	<	336	thioredoxin reductase	owl:BSZ940436	78.15	51.66
PH1434	1,286,972	1,287,493	>	173	NADH-ubiquinone oxidoreductase subunit	owl:RRU655104	69.93	53.03
PH1437	1,288,011	1,289,294	>	427	NADH-ubiquinone oxidoreductase subunit	owl:RRU655107	76.72	38.59
PH1444	1,294,013	1,295,218	>	401	protein DFP	PIR:A64414	91.78	42.16
PH1447	1,297,016	1,295,832	<	394	NADH-ubiquinone oxidoreductase subunit	Swiss_Prot:P22142	62.71	36.64
PH1449	1,298,126	1,297,539	<	195	NADH-ubiquinone oxidoreductase subunit	owl:NUKC_LUPLU	51.10	48.49
PH1459	1,308,012	1,307,080	<	310	fructokinase	Swiss_Prot:P37829	49.73	34.12
PH1463	1,309,541	1,310,857	>	438	L-asparaginase	PIR:D64302	160.79	52.84
PH1469	1,316,032	1,316,478	>	148	virulence factor regulator	owl:PAU898921	44.73	39
PH1478	1,323,462	1,321,573	<	629	arginyl-tRNA synthetase	PIR:F64329	94.37	40.98
PH1482	1,325,074	1,325,976	>	300	transcription initiation factor IIB	PIR:S34116	192.63	94.65
PH1483	1,326,447	1,326,139	<	102	30S ribosomal protein S10	Swiss_Prot:P26753	94.46	95.1
PH1484	1,327,770	1,326,484	<	428	elongation factor I-alpha	Swiss_Prot:P26751	71.56	94.39
PH1496	1,335,043	1,334,669	<	124	50S ribosomal protein L7	owl:SS56KBFR30	90.15	71.07
PH1498	1,336,393	1,335,725	<	222	endonuclease III	Swiss_Prot:P39788	68.16	42.22
PH1501	1,336,905	1,338,221	>	438	aminotransferase	owl:D5045389	75.32	38.29
PH1509	1,342,517	1,343,839	>	440	NADH oxidase	PIR:A64381	117.39	45.91
PH1510	1,343,931	1,345,256	>	441	NfeD protein	owl:REU41754	64.32	35.02
PH1516	1,351,376	1,352,287	>	303	dihydroorotate dehydrogenase	PIR:F64381	117.17	51.84
PH1527	1,359,523	1,358,462	<	353	Frv operon protein FrvX	PIR:C64369	114.69	47.63
PH1537	1,366,484	1,367,422	>	312	nucleolar protein	owl:D908275	55.12	44.3
PH1541	1,370,829	1,370,173	<	218	30S ribosomal protein S7	Swiss_Prot:P29159	133.97	77.21
PH1542	1,371,278	1,370,826	<	150	30S ribosomal protein S12	Swiss_Prot:P29161	105.49	95.92
PH1543	1,371,717	1,371,280	<	145	Nusa protein	Swiss_Prot:P29157	132.61	78.62
PH1544	1,373,222	1,372,029	<	397	DNA-directed RNA polymerase subunit A"	Swiss_Prot:P29158	172.58	80.82
PH1545	1,375,957	1,373,234	<	907	DNA-directed RNA polymerase subunit A'	Swiss_Prot:P31813	172.46	86.96
PH1546	1,379,314	1,375,961	<	1117	DNA-directed RNA polymerase subunit B	Swiss_Prot:P31814	205.97	86.2
PH1553	1,384,587	1,383,805	<	260	proteasome, alpha subunit	PIR:G64373	84.72	60.24
PH1583	1,404,608	1,405,987	>	459	tryptophan synthase beta chain	Swiss_Prot:P50383	172.28	56.84
PH1584	1,407,192	1,405,939	<	417	eukaryotic peptide chain release factor subunit 1	PIR:E64403	137.08	50.98
PH1589	1,409,727	1,410,275	>	182	deoxycytidylate deaminase	Swiss_Prot:P32393	60.62	48.67
PH1591	1,411,584	1,411,063	<	173	ferrityochelin binding protein	PIR:A64338	81.80	54.8
PH1592	1,412,149	1,411,724	<	141	transcriptional regulator	Swiss_Prot:P42180	99.70	94.23
PH1593	1,413,523	1,412,255	<	422	glutamate dehydrogenase	owl:A47410	149.19	96.43
PH1598	1,416,497	1,418,008	>	503	pyrimidine-nucleoside phosphorylase	PIR:C64383	179.65	57.26
PH1614	1,431,928	1,430,216	<	570	glycyl-tRNA synthetase	PIR:E64328	173.75	54.11
PH1617	1,435,066	1,434,056	<	336	UDP-N-acetylglucosamine 2-epimerase	PIR:G64487	61.66	57.14
PH1618	1,436,445	1,435,189	<	418	UDP-N-acetyl-D-mannosaminuronic acid dehydrogenase	PIR:D64353	104.41	50.65
PH1622	1,439,153	1,440,832	>	559	DNA ligase	Swiss_Prot:P54875	140.78	44.59
PH1629	1,445,216	1,444,599	<	205	30S ribosomal protein S2	Swiss_Prot:P54109	104.53	66.67
PH1632	1,446,906	1,446,547	<	119	DNA-directed RNA polymerase subunit N	Swiss_Prot:P39472	56.91	66.13
PH1633	1,447,210	1,446,803	<	135	30S ribosomal protein S9	Swiss_Prot:P54024	92.63	64.93

Table 4. Continued.

ORF ID	Position		Direction	Length (aa)	Predicted Product	Locus	ZScore	% Identity
	Initiation	Termination						
PH1634	1,447,649	1,447,221	<	142	50S ribosomal protein L13	Swiss_Prot:P54023	67.51	54.41
PH1636	1,448,018	1,447,656	<	120	50S ribosomal protein L18	Swiss_Prot:P54022	81.35	66.95
PH1637	1,448,959	1,448,174	<	261	DNA-directed RNA polymerase subunit D	Swiss_Prot:Q00813	50.93	34.41
PH1638	1,449,399	1,448,986	<	137	30S ribosomal protein S11	owl:SS56KBFR49	75.74	72.87
PH1640	1,449,938	1,449,396	<	180	30S ribosomal protein S4	Swiss_Prot:P54020	88.06	63.16
PH1641	1,450,395	1,449,949	<	148	30S ribosomal protein S13	owl:RS13_METJA	82.89	60.96
PH1644	1,452,697	1,451,693	<	334	centromere/microtubule binding protein	PIR:E64318	143.50	57.99
PH1647	1,454,598	1,455,388	>	396	molybdopterin biosynthesis moeA protein	PIR:B64383	83.54	42.25
PH1654	1,459,310	1,460,593	>	427	serine hydroxymethyltransferase	PIR:D64499	136.12	58.31
PH1660	1,464,590	1,464,033	<	185	ferredoxin oxidoreductase gamma subunit	PIR:H64366	103.84	58.33
PH1661	1,465,423	1,464,578	<	281	ferredoxin oxidoreductase beta subunit	PIR:A64367	136.74	54.37
PH1662	1,466,577	1,465,420	<	385	ferredoxin oxidoreductase alpha subunit	PIR:E64334	159.92	57.94
PH1663	1,467,086	1,466,574	<	170	ferredoxin oxidoreductase gamma subunit	PIR:H64366	66.47	48.82
PH1665	1,467,944	1,467,090	<	284	ferredoxin oxidoreductase beta subunit	PIR:A64367	144.62	57.3
PH1666	1,469,167	1,467,941	<	408	ferredoxin oxidoreductase alpha subunit	PIR:E64334	147.07	56.63
PH1669	1,469,586	1,470,566	>	326	signal recognition particle protein	owl:TSU95207	137.05	80.31
PH1672	1,471,595	1,472,758	>	387	developmentally regulated GTP-binding protein	PIR:E64465	83.39	40.72
PH1683	1,480,876	1,481,694	>	272	L(+)-tartrate dehydratase	PIR:E64461	86.51	50.19
PH1684	1,481,685	1,482,185	>	166	L(+)-tartrate dehydratase	owl:ECAE0003881	46.12	42.04
PH1686	1,484,738	1,483,026	<	570	glutamyl-tRNA synthetase	PIR:H64471	152.79	53.44
PH1694	1,489,735	1,488,398	<	445	signal recognition particle protein	owl:AAFFH1		51.49
PH1697	1,492,200	1,491,115	<	361	mannose-1-phosphate guanylyltransferase	owl:S76785	75.13	38.07
PH1704	1,497,990	1,497,490	<	166	intracellular proteinase	PIR:JC6003	158.49	89.76
PH1706	1,499,560	1,500,795	>	411	translation initiation factor eIF-2 gamma	PIR:D64457	172.89	67.07
PH1713	1,505,861	1,504,365	<	498	sugar transport ATP-binding protein	owl:BSZ939374	80.94	43.59
PH1716	1,509,530	1,508,430	<	366	acetylmethionine aminotransferase	PIR:A64390	77.81	42.21
PH1720	1,511,274	1,510,282	<	330	N-acetyl-gamma-glutamyl-phosphate reductase	PIR:G64436	76.72	43.13
PH1722	1,513,279	1,512,308	<	323	3-isopropylmalate dehydrogenase	PIR:C64499	61.07	45.58
PH1727	1,516,045	1,514,960	<	361	homocitrate synthase	PIR:G64473	81.76	42.2
PH1735	1,522,198	1,520,450	<	582	ABC transporter ATP-binding protein	Swiss_Prot:P54719	49.87	36.45
PH1742	1,528,222	1,527,302	<	306	Vi polysaccharide biosynthesis protein	PIR:D64326	81.92	44.67
PH1753	1,536,704	1,536,114	<	196	adenylate kinase	Swiss_Prot:P43409	79.66	52.17
PH1754	1,538,188	1,536,782	<	468	preprotein translocase secY subunit	Swiss_Prot:P28541	115.23	49.05
PH1755	1,538,674	1,538,222	<	150	50S ribosomal protein L15	Swiss_Prot:P54047	55.08	49.29
PH1756	1,539,143	1,538,676	<	155	50S ribosomal protein L30	Swiss_Prot:P54046	99.67	66.67
PH1757	1,539,865	1,539,155	<	236	30S ribosomal protein S5	Swiss_Prot:P54045	94.02	55.4
PH1758	1,540,482	1,539,862	<	206	50S ribosomal protein L18	Swiss_Prot:P54044	123.05	67.18
PH1759	1,540,936	1,540,484	<	150	50S ribosomal protein L19	Swiss_Prot:P54043	71.69	55.48
PH1761	1,541,338	1,540,946	<	130	50S ribosomal protein L32	Swiss_Prot:P54010	69.74	61.86
PH1763	1,541,912	1,541,349	<	187	50S ribosomal protein L6	Swiss_Prot:P54042	100.29	62.09
PH1764	1,542,306	1,541,914	<	130	30S ribosomal protein S8	Swiss_Prot:P54041	94.83	74.62
PH1765	1,543,058	1,542,492	<	188	50S ribosomal protein L5	Swiss_Prot:P41202	80.12	60.24
PH1766	1,543,800	1,543,069	<	243	30S ribosomal protein S4	Swiss_Prot:P54039	95.00	52.72
PH1767	1,544,174	1,543,800	<	124	50S ribosomal protein L24	Swiss_Prot:P54038	79.10	69.03
PH1768	1,544,610	1,544,176	<	144	50S ribosomal protein L14	Swiss_Prot:P54037	73.10	61.11
PH1770	1,544,957	1,544,607	<	116	30S ribosomal protein S17	Swiss_Prot:P54036	65.73	54.13
PH1772	1,546,345	1,545,713	<	210	30S ribosomal protein S3	Swiss_Prot:P54034	98.08	56.31
PH1773	1,546,817	1,546,350	<	155	50S ribosomal protein L22	Swiss_Prot:P54033	69.07	51.43
PH1774	1,547,226	1,546,828	<	132	30S ribosomal protein S19	Swiss_Prot:P54018	78.48	72.66
PH1775	1,547,956	1,547,237	<	239	50S ribosomal protein L2	Swiss_Prot:P21479	90.04	57.45
PH1776	1,549,002	1,548,235	<	255	50S ribosomal protein L4	Swiss_Prot:P54015	107.47	56.23
PH1777	1,550,101	1,549,013	<	362	50S ribosomal protein L3	Swiss_Prot:P54014	89.35	53.23
PH1792	1,559,303	1,560,976	>	557	CTP synthase	PIR:E64446	110.58	62.14
PH1798	1,566,598	1,563,059	<	1179	chromosome assembly protein	PIR:A64505	110.83	33.51
PH1805	1,570,669	1,569,644	<	341	3-hydroxy-3-methylglutaryl coenzyme A reductase	owl:SSU95360	78.13	56.93
PH1807	1,571,805	1,573,196	>	463	DNA damage-inducible protein	PIR:JC4561	206.85	76.94
PH1811	1,575,928	1,575,449	<	159	molybdenum cofactor biosynthesis protein	PIR:F64441	101.90	66.67
PH1813	1,577,321	1,578,811	>	496	TRK system potassium uptake protein	PIR:D64485	101.28	37.56
PH1821	1,584,208	1,583,144	<	354	Frv operon protein FrvX	PIR:C64369	88.00	39.39
PH1823	1,584,856	1,585,728	>	290	dimethyladenosine transferase	PIR:D64428	86.81	50.83
PH1826	1,587,912	1,588,769	>	285	integrase/recombinase	owl:STU925241	44.40	40.21
PH1830	1,591,416	1,592,420	>	334	glyceraldehyde-3-phosphate dehydrogenase	Swiss_Prot:P20286	200.79	90.72
PH1840	1,601,085	1,598,689	<	798	transitional endoplasmic reticulum ATPase	PIR:C64444	110.36	70.58
PH1852	1,609,753	1,608,230	<	507	methyl-accepting chemotaxis protein	PIR:A54078	44.02	26.59
PH1873	1,630,188	1,628,758	<	476	glutamate synthase small chain	owl:D86223	221.09	82.55
PH1884	1,638,869	1,638,174	<	231	triosephosphate isomerase	PIR:S66212	169.40	93.43
PH1886	1,640,335	1,639,517	<	272	L-isopartyl protein carboxyl methyltransferase	PIR:E64321	89.23	54.03
PH1899	1,653,999	1,651,792	<	735	elongation factor 2	Swiss_Prot:P29050	204.09	95.36
PH1907	1,659,954	1,660,490	>	178	inorganic pyrophosphatase	owl:TLU49440	147.21	89.77
PH1908	1,660,531	1,661,097	>	188	DNA-directed RNA polymerase subunit E	PIR:E64349	105.14	53.8
PH1923	1,670,714	1,671,577	>	287	ribose-phosphate pyrophosphokinase	PIR:E64470	77.80	41.3
PH1925	1,671,574	1,672,833	>	419	glucose-1-phosphate thymidyltransferase	PIR:D64437	118.15	45.81
PH1941	1,682,801	1,683,361	>	186	maf protein	Swiss_Prot:Q02169	59.87	44.97

Table 4. Continued.

ORF ID	Position		Direction	Length (aa)	Predicted Product	Locus	ZScore	% Identity
	Initiation	Termination						
PH1942	1,683,403	1,684,689	>	428	phosphoglycerate dehydratase	PIR:A64329	73.28	72.51
PH1947	1,686,361	1,690,068	>	1235	DNA-directed DNA polymerase	PIR:S68593	213.39	75.83
PH1953	1,696,800	1,694,683	<	705	phosphoribosylformylglycinamide synthase II	PIR:G64457	159.26	50.36
PH1955	1,698,110	1,697,427	<	227	phosphoribosylformylglycinamide synthase I	PIR:F64505	107.20	53.36
PH1958	1,700,495	1,699,488	<	335	oligopeptide transport ATP-binding protein appF	Swiss_Prot:P42065	58.77	46.56
PH1959	1,701,473	1,700,505	<	322	oligopeptide transport ATP-binding protein appD	Swiss_Prot:P42064	58.27	47.6
PH1960	1,702,899	1,701,475	<	474	oligopeptide transport system permease protein appC	owl:MMABCTRAN2	59.70	38.97
PH1961	1,703,960	1,702,911	<	349	oligopeptide transport system permease protein appB	Swiss_Prot:P42062	51.93	32.6
PH1963	1,706,990	1,708,243	>	417	dihydroorotase	PIR:A64486	65.78	40.27
PH1965	1,708,293	1,708,934	>	213	cytochrome-c3 hydrogenase	PIR:E64480	79.18	50
PH1972	1,714,142	1,713,498	<	214	H(+)-transporting ATP synthase subunit D	owl:DSU964876	133.39	73.46
PH1974	1,715,564	1,714,167	<	465	H(+)-transporting ATP synthase subunit B	owl:DSU964875	97.13	94.97
PH1975	1,718,455	1,715,561	<	964	H(+)-transporting ATP synthase subunit A	owl:DSU964874	107.28	90.52
PH1981	1,722,987	1,721,008	<	659	V-type sodium ATP synthase	PIR:G64327	99.98	33.78
PH1987	1,726,670	1,727,644	>	324	O-sialoglycoprotein endopeptidase	PIR:A64441	133.78	58.88
PH1994	1,733,615	1,732,107	<	502	glycine dehydrogenase subunit 2	Swiss_Prot:P54377	167.70	60.8
PH1995	1,734,965	1,733,616	<	449	glycine dehydrogenase subunit 1	Swiss_Prot:P54376	128.33	46.96
PH1998	1,736,839	1,736,504	<	111	50S ribosomal protein L12	Swiss_Prot:P54048	48.57	70.59
PH1999	1,737,901	1,736,873	<	342	acidic ribosomal protein P0 (L10E)	owl:RLA0_METJA	107.14	52.8
PH2000	61	1,737,907	<	219	50S ribosomal protein L1	Swiss_Prot:P54050	100.68	58.14
PHS019	424806	425003	>	65	nusA protein	PIR:D64430	15.43	35.29
PHS021	474972	474739	<	77	50S ribosomal protein LX	Swiss_Prot:P54052	36.93	48.57
PHS023	572436	572663	>	75	regulatory protein asnC	Swiss_Prot:P44337	21.78	34.78
PHS032	1024154	1023861	<	97	50S ribosomal protein L21	Swiss_Prot:P54013	73.93	67.37
PHS034	1130960	1131157	>	65	zinc finger protein	owl:NRL_1MEYG	13.31	43.48
PHS035	1137843	1138127	>	94	50S ribosomal protein L34	Swiss_Prot:P54053	66.41	63.10
PHS037	1177477	1177322	<	51	50S ribosomal protein L40	Swiss_Prot:P54058	48.31	71.11
PHS040	1334621	1334406	<	71	30S ribosomal protein S33	Swiss_Prot:P54065	53.78	64.18
PHS041	1353966	1353778	<	62	50S ribosomal protein L37	Swiss_Prot:P54011	44.02	60.00
PHS042	1354205	1353978	<	75	small nuclear ribonucleoprotein	owl:DRODEBB	23.68	37.33
PHS043	1372025	1371726	<	99	50S ribosomal protein L30	Swiss_Prot:P29160	75.89	70.83
PHS044	1379580	1379332	<	82	DNA-directed RNA polymerase subunit H	Swiss_Prot:P31815	71.81	71.95
PHS045	1488392	1488168	<	74	transcriptional regulator	Swiss_Prot:P42179	14.73	40.48
PHS046	1495611	1495399	<	70	archaeal histone	Swiss_Prot:P50486	66	96.97
PHS047	1542486	1542316	<	56	30S ribosomal protein S14	Swiss_Prot:P14041	38.51	65.91
PHS048	1545702	1545520	<	60	50S ribosomal protein L29	Swiss_Prot:P54035	25.93	45.46
PHS050	1548228	1547968	<	86	50S ribosomal protein L23	Swiss_Prot:P54016	64.03	69.77
PHS051	1553990	1554202	>	70	archaeal histone	Swiss_Prot:P50485	68.59	92.54
PHS052	1591315	1591052	<	87	50S ribosomal protein L35	Swiss_Prot:P20299	104.6	91.95
PHS055	1648193	1647921	<	90	DNA excision repair protein	PIR:H64487	32.96	53.49
PHS056	1655234	1655001	<	77	DNA-directed RNA polymerase	Swiss_Prot:P40422	12.69	52.94
PHS057	1655418	1655158	<	86	50S ribosomal protein L37	Swiss_Prot:P54051	55.1	60.5
<b>ORFs with unknown</b>								
PH0002	1,099	632	<	155	hypothetical protein	PIR:D64346	74.72	47.59
PH0006	4,123	3,299	<	274	hypothetical protein	PIR:G64320	47.62	45.19
PH0007	4,197	4,895	>	232	hypothetical protein	owl:SS56KBFR35	52.81	37.78
PH0010	7,878	8,498	>	206	hypothetical protein	PIR:B64401	101.74	60.31
PH0018	15,743	16,264	>	173	hypothetical protein	PIR:S59860	59.12	54.27
PH0023	22,756	21,923	<	277	hypothetical protein	PIR:S33182	41.82	32.65
PH0027	26,670	27,503	>	277	hypothetical protein	PIR:H64334	68.45	35.93
PH0028	29,753	27,888	<	621	hypothetical protein	owl:PFAOR1	102.16	36.57
PH0031	30,562	29,741	<	273	hypothetical protein	owl:D908222	54.29	39.6
PH0034	32,730	31,723	<	335	hypothetical protein	PIR:E64499	71.89	38.84
PH0044	41,476	38,909	<	855	hypothetical protein	PIR:H64474	183.01	44.31
PH0053	47,479	46,265	<	404	hypothetical protein	PIR:F64386	106.46	52.67
PH0057	50,258	49,752	<	168	hypothetical protein	owl:HPAE00058613	62.21	43.31
PH0059	50,891	52,327	>	478	hypothetical protein	PIR:A64422	53.2	31.66
PH0071	61,784	61,203	<	193	hypothetical protein	PIR:E64443	58.86	40.46
PH0073	62,507	63,094	>	195	hypothetical protein	PIR:A64360	75.49	56.03
PH0074	63,009	63,500	>	163	hypothetical protein	PIR:A64360	75.98	54.25
PH0080	66,031	67,098	>	355	hypothetical protein	PIR:E64352	181.57	62.28
PH0082	68,134	68,979	>	281	hypothetical protein	PIR:H64349	60.64	36.78
PH0083	69,041	69,892	>	283	hypothetical protein	Swiss_Prot:P19268	68.39	40.08
PH0085	71,181	70,372	<	269	hypothetical protein	PIR:B64468	67.53	41.05
PH0094	78,515	80,257	>	580	hypothetical protein	PIR:C64334	168.85	54.18
PH0099	83,374	83,928	>	184	hypothetical protein	PIR:G64495	68.58	48.82
PH0101	84,058	85,503	>	481	hypothetical protein	PIR:F64438	63.38	40.78
PH0113	96,411	97,817	>	468	hypothetical protein	PIR:D64410	89.09	38.7
PH0119	102,646	101,960	<	228	hypothetical protein	owl:D83176	118.09	66
PH0121	107,617	103,313	<	1434	hypothetical protein	PIR:D64503	239.22	56.2
PH0123	109,476	107,608	<	622	hypothetical protein	PIR:F64387	109.56	41.22
PH0124	110,789	109,476	<	437	hypothetical protein	Swiss_Prot:P29570	62.63	39.04

Table 4. Continued.

ORF ID	Position		Direction	Length (aa)	Predicted Product	Locus	ZScore	% Identity
	Initiation	Termination						
PH0126	112,436	113,161	>	241	hypothetical protein	PIR:B64313	66.84	39.82
PH0128	114,299	113,538	<	253	hypothetical protein	PIR:B64328	119.26	58.37
PH0132	117,356	116,574	<	260	hypothetical protein	PIR:G64435	47.13	49.1
PH0136	120,186	121,313	>	375	hypothetical protein	Swiss_Prot:P26235	85.09	34.42
PH0140	124,611	124,135	<	158	hypothetical protein	PIR:C64390	53.06	41.43
PH0151	131,762	132,784	>	340	hypothetical protein	PIR:A64448	129.16	55.15
PH0154	134,750	135,496	>	248	hypothetical protein	PIR:G64470	85.98	54.87
PH0155	135,493	136,446	>	317	hypothetical protein	PIR:B64511	60.23	41.03
PH0156	136,446	137,171	>	241	hypothetical protein	PIR:D64511	48.07	33.64
PH0157	137,173	138,210	>	345	hypothetical protein	PIR:F64470	49.07	53.25
PH0161	139,375	140,103	>	242	hypothetical protein	PIR:A64454	106.94	54.59
PH0168	145,694	146,947	>	417	hypothetical protein	PIR:H64507	59.47	37.66
PH0173	152,194	153,147	>	317	hypothetical protein	PIR:B64347	104.6	42.72
PH0175	153,401	153,889	>	162	hypothetical protein	PIR:A64347	63.2	36.25
PH0181	159,108	160,289	>	393	hypothetical protein	PIR:D64366	85.45	39.13
PH0186	163,610	164,365	>	251	hypothetical protein	PIR:F64469	71.39	41.03
PH0197	173,822	174,907	>	361	hypothetical protein	PIR:B64464	158.08	54.44
PH0200	175,990	175,418	<	190	hypothetical protein	PIR:A64340	72.94	48.33
PH0207	183,286	182,000	<	428	hypothetical protein	owl:SS56KBFR17	119.73	50.89
PH0209	184,850	184,455	<	131	hypothetical protein	PIR:H64385	47.72	40.48
PH0223	197,340	197,912	>	190	hypothetical protein	PIR:E64400	53.16	39.89
PH0224	198,383	199,954	>	523	hypothetical protein	PIR:C64367	130.37	40.97
PH0229	203,680	204,270	>	196	hypothetical protein	owl:D862211	60.81	49.17
PH0249	223,806	225,086	>	426	hypothetical protein	PIR:H64328	58.18	33.58
PH0256	228,748	227,615	<	377	hypothetical protein	owl:SS56KBFR33	43.94	33.13
PH0259	229,331	230,062	>	243	hypothetical protein	PIR:E64416	65.02	36.49
PH0266	235,270	236,013	>	247	hypothetical protein	PIR:F64305	71.83	45.06
PH0268	236,679	238,571	>	630	hypothetical protein	PIR:B64308	116.29	40.96
PH0272	240,980	240,570	<	136	hypothetical protein	Swiss_Prot:P54540	53.27	44.19
PH0274	241,930	240,977	<	317	hypothetical protein	owl:MTCY27717	74.75	41.15
PH0275	242,424	241,981	<	147	hypothetical protein	owl:AF008569	40.34	60
PH0276	242,532	243,152	>	206	hypothetical protein	PIR:B64386	97.58	52.28
PH0279	244,973	246,439	>	488	hypothetical protein	PIR:G64503	104.47	39.36
PH0280	246,441	247,967	>	508	hypothetical protein	PIR:C64313	75.15	38.93
PH0282	249,500	248,658	<	280	hypothetical protein	PIR:A64317	200.28	83.39
PH0283	250,344	249,505	<	279	hypothetical protein	PIR:D64431	47.72	56.17
PH0284	251,102	250,344	<	252	hypothetical protein	PIR:F64469	165.89	74.37
PH0286	254,283	251,584	<	899	hypothetical protein	PIR:H64494	57.62	33.88
PH0287	256,643	254,280	<	787	hypothetical protein	PIR:A64495	99.52	36.84
PH0295	262,069	261,557	<	170	hypothetical protein	PIR:G64477	64.88	45.83
PH0296	262,584	262,066	<	172	hypothetical protein	PIR:A64491	47.08	42.03
PH0300	267,256	266,324	<	310	hypothetical protein	PIR:E64484	147.96	53.93
PH0305	270,085	270,726	>	213	hypothetical protein	PIR:S41581	80.36	48.86
PH0309	273,279	274,112	>	277	hypothetical protein	PIR:F64349	113.12	52.67
PH0310	274,497	277,964	>	1155	hypothetical protein	owl:PFU55835	48.57	77.45
PH0311	278,516	277,986	<	176	hypothetical protein	owl:BS168NPRB6	49.91	44.79
PH0324	293,174	294,325	>	383	hypothetical protein	PIR:H64316	108.35	44.85
PH0329	299,431	300,384	>	317	hypothetical protein	owl:BSU511154	110.89	51.34
PH0336	304,789	305,796	>	335	hypothetical protein	PIR:S76512	61.60	38.19
PH0340	307,500	309,722	>	740	hypothetical protein	PIR:E64449	231.44	50.96
PH0343	309,719	310,231	>	170	hypothetical protein	PIR:G64499	71.99	49.09
PH0344	310,801	310,241	<	186	hypothetical protein	PIR:H64414	64.36	44.89
PH0348	311,752	312,573	>	273	hypothetical protein	owl:SS56KBFR61	48.54	34.07
PH0350	313,276	312,557	<	239	hypothetical protein	PIR:G64346	53.58	35.29
PH0351	314,468	313,347	<	373	hypothetical protein	PIR:S76517	61.75	38.43
PH0353	314,955	314,551	<	134	hypothetical protein	PIR:F64475	68.29	55.12
PH0354	316,121	314,955	<	388	hypothetical protein	PIR:H64463	156.18	57.65
PH0356	317,577	316,465	<	370	hypothetical protein	PIR:G64489	40.00	28.74
PH0361	323,632	321,752	<	626	hypothetical protein	owl:MTCY25325	112.77	38.68
PH0363	327,756	323,881	<	1291	hypothetical protein	owl:PFU78098	344.62	85.31
PH0373	336,578	335,886	<	230	hypothetical protein	owl:PYWCS	158.30	70.18
PH0374	337,254	336,811	<	147	hypothetical protein	PIR:A64466	44.63	51.75
PH0385	348,177	349,115	>	312	hypothetical protein	owl:BSU511154	106.72	45.15
PH0430	391,067	392,071	>	334	hypothetical protein	Swiss_Prot:P22639	40.18	42.99
PH0433	395,935	395,021	<	304	hypothetical protein	PIR:S77381	44.77	29.41
PH0434	396,121	397,131	>	336	hypothetical protein	PIR:H64446	49.82	32.53
PH0435	397,128	398,150	>	340	hypothetical protein	owl:YWLC_BACSU	124.95	52.19
PH0436	398,162	399,520	>	452	hypothetical protein	PIR:D64510	116.43	43.46
PH0445	407,349	406,690	<	219	hypothetical protein	PIR:A64441	46.63	42.46
PH0457	417,190	419,151	>	653	hypothetical protein	PIR:H64447	179.84	48.74
PH0461	421,816	421,196	<	206	hypothetical protein	PIR:H64472	95.44	60
PH0463	423,478	424,248	>	256	hypothetical protein	owl:HPAE0005842	45.51	42.38
PH0464	424,245	424,805	>	186	hypothetical protein	PIR:E64367	104.77	59.76

Table 4. Continued.

ORF ID	Position		Direction	Length (aa)	Predicted Product	Locus	ZScore	% Identity
	Initiation	Termination						
PH0469	429,868	430,308	>	146	hypothetical protein	PIR:F64509	48.75	39.71
PH0473	435,476	434,499	<	325	hypothetical protein	PIR:C64311	69.62	36.79
PH0490	444,354	444,842	>	162	hypothetical protein	Swiss_Prot:P40404	53.57	40.52
PH0495	450,186	448,864	<	440	hypothetical protein	owl:AVU256343	89.38	41.98
PH0498	451,327	450,188	<	379	hypothetical protein	PIR:F64351	73.66	34.69
PH0512	464,553	463,612	<	313	hypothetical protein	PIR:H64433	89.81	43.4
PH0519	468,782	467,784	<	332	hypothetical protein	PIR:C64369	74.16	40.12
PH0523	471,493	472,014	>	173	hypothetical protein	owl:AF008220145	48.31	47.85
PH0525	473,077	474,270	>	397	hypothetical protein	PIR:C64466	163.43	55.39
PH0528	475,668	474,985	<	227	hypothetical protein	PIR:H64305	82.85	44.64
PH0532	478,012	478,365	>	117	hypothetical protein	PIR:C64423	47.71	51.4
PH0534	478,695	479,684	>	329	hypothetical protein	owl:D64423	62.84	39.74
PH0541	485,685	484,720	<	321	hypothetical protein	PIR:A64507	81.83	44.34
PH0544	487,494	487,901	>	135	hypothetical protein	PIR:A64413	66.11	57.9
PH0545	487,903	488,874	>	323	hypothetical protein	PIR:B64413	68.21	34.48
PH0555	495,037	495,531	>	164	hypothetical protein	PIR:B64412	40.05	40.69
PH0556	495,546	496,244	>	232	hypothetical protein	PIR:C64412	98.86	43.04
PH0557	496,246	497,859	>	537	hypothetical protein	PIR:D64412	219.19	67.86
PH0559	497,861	499,582	>	573	hypothetical protein	PIR:E64412	147.59	41.67
PH0567	507,383	506,022	<	453	hypothetical protein	PIR:C64504	104.42	40
PH0575	513,739	512,321	<	472	hypothetical protein	owl:YPPNP_BACSU	56.48	24.13
PH0578	515,244	516,617	>	457	hypothetical protein	owl:S76942	199.37	59.03
PH0581	518,837	519,238	>	133	hypothetical protein	PIR:H64434	61.24	46.09
PH0589	530,022	528,649	<	457	hypothetical protein	PIR:C64500	64.56	34.53
PH0591	531,071	530,514	<	185	hypothetical protein	PIR:A64328	64.27	52.56
PH0600	538,763	537,585	<	392	hypothetical protein	PIR:C64475	43.48	30.89
PH0601	539,277	538,861	<	138	hypothetical protein	PIR:E64368	74.01	50.74
PH0619	555,216	554,623	<	197	hypothetical protein	PIR:A64397	58.25	40.67
PH0624	560,892	560,359	<	177	hypothetical protein	PIR:E64428	42.67	43.84
PH0625	560,891	562,213	>	440	hypothetical protein	PIR:F64373	115.53	44.11
PH0627	563,422	564,174	>	250	hypothetical protein	PIR:G64142	62.78	38.53
PH0637	573,345	572,665	<	226	hypothetical protein	Swiss_Prot:P44271	41.78	28.24
PH0644	578,243	577,149	<	364	hypothetical protein	PIR:F64354	56.2	42.75
PH0645	579,347	578,280	<	355	hypothetical protein	PIR:G64482	89.73	43.83
PH0650	582,120	585,530	>	1136	hypothetical protein	PIR:E64397	103.82	49.58
PH0663	594,021	593,371	<	216	hypothetical protein	PIR:H64321	44.28	38.92
PH0665	594,587	595,336	>	249	hypothetical protein	PIR:H64330	87.80	43.03
PH0673	600,137	599,538	<	199	hypothetical protein	owl:ECAE0001892	40.20	31.05
PH0674	600,816	600,361	<	151	hypothetical protein	owl:PFPPORVOR1	96.60	67.41
PH0675	601,244	600,840	<	134	hypothetical protein	owl:PFPPORVOR2	141.53	88.81
PH0676	602,415	601,246	<	389	hypothetical protein	owl:PFPPORVOR3	249.35	91.97
PH0677	603,481	602,429	<	350	hypothetical protein	owl:PFPPORVOR4	236.98	90.86
PH0686	609,339	610,130	>	263	hypothetical protein	owl:PFPPORVOR12	110.85	82.05
PH0691	614,757	616,076	>	439	hypothetical protein	PIR:G64492	83.16	40.92
PH0718	642,245	641,049	<	398	hypothetical protein	PIR:D64464	91.15	41.85
PH0728	649,788	648,739	<	349	hypothetical protein	PIR:H64458	196.50	64.18
PH0731	651,487	652,212	>	241	hypothetical protein	PIR:E64331	78.05	49.23
PH0734	652,205	652,732	>	175	hypothetical protein	PIR:G64478	61.27	49.37
PH0735	652,711	653,157	>	148	hypothetical protein	PIR:B64447	58.54	49.3
PH0737	654,681	653,671	<	336	hypothetical protein	owl:BSY1408313	128.72	48.95
PH0746	662,161	659,948	<	737	hypothetical protein	owl:D9076912	92.09	30.22
PH0755	666,390	667,229	>	279	hypothetical protein	PIR:A56641	70.62	36.23
PH0759	669,143	670,315	>	390	hypothetical protein	PIR:D64337	187.51	66.39
PH0760	670,921	670,316	<	201	hypothetical protein	PIR:C64509	73.89	45.36
PH0763	672,005	672,508	>	167	hypothetical protein	PIR:F64321	56.50	48
PH0764	673,118	672,510	<	202	hypothetical protein	owl:D862211	142.14	75.25
PH0766	676,435	675,062	<	457	hypothetical protein	PIR:F64373	136.77	47.36
PH0776	685,026	684,082	<	314	hypothetical protein	owl:BSU511154	120.95	51.68
PH0792	700,485	701,003	>	172	hypothetical protein	PIR:F64494	91.30	53.53
PH0793	701,000	701,836	>	278	hypothetical protein	PIR:D64494	89.70	48.77
PH0816	726,609	725,503	<	368	hypothetical protein	owl:ECAE0001842	71.58	36.22
PH0820	729,978	730,916	>	312	hypothetical protein	owl:SS100KBFR62	45.93	41.39
PH0823	732,171	732,683	>	170	hypothetical protein	PIR:A64372	41.01	41.33
PH0830	739,026	739,985	>	319	hypothetical protein	PIR:D64510	89.17	43.77
PH0833	742,734	741,283	<	483	hypothetical protein	owl:S76850	64.40	29.91
PH0841	750,410	751,381	>	323	hypothetical protein	PIR:D64510	91.83	41.41
PH0846	753,134	754,504	>	456	hypothetical protein	PIR:D64510	66.31	34.52
PH0848	755,606	757,051	>	481	hypothetical protein	PIR:D64424	85.56	35.54
PH0854	762,636	762,223	<	137	hypothetical protein	Swiss_Prot:P37552	64.41	55.74
PH0859	769,130	767,358	<	590	hypothetical protein	PIR:G64389	162.40	61.92
PH0870	776,339	776,836	>	165	hypothetical protein	PIR:A64399	51.62	48
PH0874	783,159	782,317	<	280	hypothetical protein	PIR:F64337	174.43	71.01
PH0882	788,182	788,895	>	237	hypothetical protein	PIR:C64304	63.59	50.23

Table 4. Continued.

ORF ID	Position		Direction	Length (aa)	Predicted Product	Locus	ZScore	% Identity
	Initiation	Termination						
PH0883	788,865	789,845	>	326	hypothetical protein	PIR:B64304	84.19	34.75
PH0890	795,307	794,114	<	397	hypothetical protein	PIR:A64381	48.20	33.75
PH0891	797,049	795,304	<	581	hypothetical protein	owl:PFAOR1	58.98	29.98
PH0892	798,924	797,056	<	622	hypothetical protein	owl:PFAOR1	65.39	30.35
PH0907	816,375	816,854	>	159	hypothetical protein	PIR:C64516	68.01	57.69
PH0910	821,345	820,929	<	138	hypothetical protein	PIR:E64513	81.29	59.4
PH0913	823,077	822,103	<	324	hypothetical protein	owl:SS100KBFR62	54.50	45.83
PH0929	840,733	838,094	<	879	hypothetical protein	owl:SAPNTPASE	65.34	26.53
PH0932	843,657	841,990	<	555	hypothetical protein	PIR:D64478	114.64	41.3
PH0934	844,850	843,678	<	390	hypothetical protein	Swiss_Prot:P26235	57.13	29.89
PH0937	846,434	847,585	>	383	hypothetical protein	PIR:B64306	101.67	42.59
PH0947	855,178	854,429	<	249	hypothetical protein	owl:MTC16517	80.09	41.03
PH0958	875,067	874,264	<	267	hypothetical protein	PIR:A64451	62.33	40.6
PH0962	876,098	877,321	>	407	hypothetical protein	owl:A64305	105.36	48.48
PH0963	877,290	878,564	>	424	hypothetical protein	PIR:D64441	131.06	49.75
PH0976	889,988	888,573	<	471	hypothetical protein	PIR:D64510	120.77	45.26
PH0977	891,541	890,051	<	496	hypothetical protein	PIR:D64510	135.26	46.12
PH0979	892,228	892,671	>	147	hypothetical protein	PIR:G64378	40.35	35.46
PH0981	893,295	894,224	>	309	hypothetical protein	PIR:G64463	162.49	58.9
PH0982	894,286	895,146	>	286	hypothetical protein	PIR:C64372	45.28	36.59
PH0983	895,143	896,045	>	300	hypothetical protein	PIR:B64372	49.31	38.81
PH0986	898,026	898,793	>	255	hypothetical protein	owl:AF0060004	112.77	48.98
PH0987	898,800	899,477	>	225	hypothetical protein	Swiss_Prot:P42967	73.76	50
PH0988	899,468	900,463	>	331	hypothetical protein	Swiss_Prot:P44298	70.96	34.77
PH0990	900,797	901,966	>	389	hypothetical protein	PIR:H64505	71.07	34.83
PH1001	909,607	909,155	<	150	hypothetical protein	PIR:B64408	41.07	39.25
PH1002	909,674	910,732	>	352	hypothetical protein	PIR:C64503	126.64	54.48
PH1003	910,729	913,350	>	873	hypothetical protein	PIR:E64496	237.67	54.75
PH1008	916,952	917,512	>	186	hypothetical protein	PIR:F64474	74.87	40.88
PH1015	921,424	922,242	>	272	hypothetical protein	PIR:A64425	67.68	54.8
PH1019	925,762	923,939	<	607	hypothetical protein	owl:PFAOR1	299.32	80.63
PH1028	932,861	933,706	>	281	hypothetical protein	PIR:H64453	161.57	66.91
PH1033	941,908	942,345	>	145	hypothetical protein	PIR:C64514	82.02	61.59
PH1036	945,451	944,621	<	276	hypothetical protein	PIR:S75972	68.44	39.38
PH1038	946,331	945,456	<	291	hypothetical protein	PIR:S77536	45.07	29.43
PH1043	950,299	949,136	<	387	hypothetical protein	Swiss_Prot:P80092	114.72	46.93
PH1044	951,788	950,313	<	491	hypothetical protein	owl:BSY140812	103.61	34.87
PH1056	960,282	960,692	>	136	hypothetical protein	PIR:F64497	80.26	57.5
PH1059	963,932	963,507	<	141	hypothetical protein	PIR:G64325	66.83	50.82
PH1060	964,833	964,486	<	115	hypothetical protein	PIR:C64386	41.12	53.85
PH1064	968,942	966,786	<	718	hypothetical protein	PIR:A64510	45.75	34.36
PH1069	974,050	973,448	<	200	hypothetical protein	PIR:E64488	88.33	45.79
PH1071	975,841	977,193	>	450	hypothetical protein	PIR:E64407	174.89	52.55
PH1074	978,541	979,374	>	277	hypothetical protein	PIR:E64414	65.67	40.55
PH1075	980,330	979,371	<	319	hypothetical protein	PIR:A64500	47.67	34.97
PH1078	981,360	982,712	>	450	hypothetical protein	PIR:B64303	47.41	36.15
PH1087	987,849	988,664	>	271	hypothetical protein	PIR:G64437	49.86	38.03
PH1095	997,964	994,830	<	1044	hypothetical protein	PIR:G64332	219.12	51.32
PH1102	1,003,056	1,001,614	<	480	hypothetical protein	PIR:C64439	54.68	45.3
PH1105	1,005,004	1,003,976	<	342	hypothetical protein	PIR:C64360	40.14	36.14
PH1109	1,007,011	1,007,445	>	144	hypothetical protein	owl:ECAE0001989	50.57	43.51
PH1113	1,013,034	1,012,021	<	337	hypothetical protein	owl:TLFORPKPF1	210.93	71.64
PH1116	1,016,338	1,014,590	<	582	hypothetical protein	PIR:D64354	141.08	55.96
PH1123	1,019,948	1,021,765	>	605	hypothetical protein	PIR:D64491	189.59	55.2
PH1124	1,022,526	1,021,762	<	254	hypothetical protein	PIR:A64456	80.09	46.46
PH1134	1,029,326	1,028,394	<	310	hypothetical protein	PIR:B64440	95.92	51.49
PH1137	1,031,396	1,030,173	<	407	hypothetical protein	PIR:S77558	53.58	31.94
PH1139	1,033,299	1,034,783	>	494	hypothetical protein	owl:AF008220136	52.78	29.55
PH1151	1,045,229	1,046,152	>	307	hypothetical protein	PIR:E64487	78.12	41.03
PH1155	1,050,006	1,048,666	<	446	hypothetical protein	Swiss_Prot:P44697	82.73	51.54
PH1156	1,050,488	1,050,003	<	161	hypothetical protein	PIR:S74943	50.90	44.37
PH1157	1,051,413	1,050,616	<	265	hypothetical protein	Swiss_Prot:P39593	80.51	43.85
PH1161	1,053,941	1,053,285	<	218	hypothetical protein	owl:HPAE0006332	44.12	31.75
PH1162	1,054,092	1,055,423	>	443	hypothetical protein	PIR:F64340	143.46	54.29
PH1164	1,057,004	1,055,889	<	371	hypothetical protein	PIR:E64302	103.05	45.19
PH1202	1,083,392	1,084,507	>	371	hypothetical protein	PIR:F64407	99.95	43.48
PH1208	1,088,464	1,087,712	<	250	hypothetical protein	PIR:E64497	69.81	47.34
PH1209	1,088,510	1,089,289	>	259	hypothetical protein	owl:B64439	149.46	58.82
PH1214	1,092,862	1,094,187	>	441	hypothetical protein	owl:BRUP39CYTO	51.18	29.06
PH1215	1,094,203	1,095,081	>	292	hypothetical protein	owl:SPU435263	60.17	30.63
PH1216	1,095,078	1,095,905	>	275	hypothetical protein	PIR:S75972	50.40	35.61
PH1223	1,101,588	1,102,259	>	223	hypothetical protein	owl:B64366	80.29	47.59
PH1234	1,109,486	1,110,325	>	279	hypothetical protein	PIR:F64440	113.95	51.87



Table 4. Continued.

ORF ID	Position		Direction	Length (aa)	Predicted Product	Locus	ZScore	% Identity
	Initiation	Termination						
PH1235	1,111,072	1,110,320	<	250	hypothetical protein	PIR:A64311	53.95	46.86
PH1237	1,113,207	1,112,107	<	366	hypothetical protein	PIR:E64310	179.87	61.72
PH1242	1,115,697	1,116,485	>	262	hypothetical protein	PIR:A64509	46.44	37.24
PH1244	1,116,984	1,117,784	>	266	hypothetical protein	PIR:H64503	55.54	44.7
PH1245	1,120,676	1,119,708	<	322	hypothetical protein	owl:B64347	204.97	71.12
PH1252	1,127,446	1,126,643	<	267	hypothetical protein	PIR:G64346	64.24	36.28
PH1255	1,129,502	1,130,044	>	180	hypothetical protein	PIR:A64431	53.24	42.07
PH1256	1,130,037	1,130,915	>	292	hypothetical protein	PIR:H64509	125.08	59.92
PH1257	1,131,205	1,131,888	>	227	hypothetical protein	owl:B64371	117.04	58.99
PH1263	1,137,064	1,136,102	<	320	hypothetical protein	PIR:D64510	163.54	61.64
PH1265	1,138,131	1,138,709	>	192	hypothetical protein	PIR:H64381	106.04	61.63
PH1267	1,139,004	1,140,011	>	335	hypothetical protein	PIR:G64366	124.12	49.67
PH1278	1,149,858	1,148,767	<	363	hypothetical protein	Swiss_Prot:P52917	48.04	45.78
PH1280	1,150,290	1,152,437	>	715	hypothetical protein	PIR:C64440	81.80	52.16
PH1297	1,166,966	1,166,538	<	142	hypothetical protein	PIR:E64375	56.74	54.65
PH1310	1,181,015	1,180,329	<	228	hypothetical protein	PIR:E64304	56.51	42.93
PH1312	1,181,947	1,182,261	>	104	hypothetical protein	PIR:D64308	50.33	44.55
PH1313	1,183,259	1,182,336	<	307	hypothetical protein	PIR:C64416	60.48	39.38
PH1314	1,183,332	1,184,009	>	225	hypothetical protein	PIR:F64482	61.78	38.03
PH1319	1,186,531	1,187,103	>	190	hypothetical protein	owl:B64505	55.16	42.69
PH1320	1,188,150	1,187,077	<	357	hypothetical protein	PIR:G64475	124.73	53.4
PH1323	1,191,035	1,191,454	>	139	hypothetical protein	PIR:D64381	43.35	43.22
PH1328	1,194,726	1,194,352	<	124	hypothetical protein	PIR:A64362	46.15	55.42
PH1329	1,194,780	1,197,518	>	912	hypothetical protein	owl:G64336	90.45	37.5
PH1348	1,215,382	1,214,381	<	333	hypothetical protein	PIR:H64316	126.61	49.85
PH1354	1,221,796	1,221,206	<	196	hypothetical protein	Swiss_Prot:P37528	71.33	50.54
PH1359	1,226,423	1,224,504	<	639	hypothetical protein	Swiss_Prot:P37519	58.52	30.41
PH1372	1,238,549	1,240,114	>	521	hypothetical protein	PIR:F64464	45.43	28.92
PH1379	1,244,552	1,245,241	>	229	hypothetical protein	PIR:E64369	61.41	50.91
PH1381	1,246,715	1,247,131	>	138	hypothetical protein	PIR:C64453	65.50	52.8
PH1382	1,247,966	1,247,142	<	274	hypothetical protein	PIR:F64337	54.80	34.59
PH1385	1,249,216	1,250,556	>	446	hypothetical protein	owl:MTU193641	76.81	36.92
PH1386	1,250,608	1,252,290	>	560	hypothetical protein	owl:S76831	84.76	36.27
PH1391	1,254,796	1,255,842	>	348	hypothetical protein	PIR:H64400	141.90	54.24
PH1399	1,260,500	1,261,273	>	257	hypothetical protein	PIR:A64403	56.22	40.66
PH1401	1,262,203	1,262,814	>	203	hypothetical protein	PIR:F64504	61.03	42.56
PH1404	1,263,527	1,265,482	>	651	hypothetical protein	PIR:C64454	251.29	59.84
PH1408	1,266,820	1,267,500	>	226	hypothetical protein	PIR:B64457	99.80	55.61
PH1416	1,276,682	1,275,690	<	330	hypothetical protein	PIR:C64410	91.35	45.63
PH1439	1,289,299	1,290,264	>	321	hypothetical protein	owl:RRU655103	53.52	35.74
PH1442	1,293,945	1,292,029	<	638	hypothetical protein	PIR:E64464	90.67	38.81
PH1458	1,305,044	1,306,807	>	587	hypothetical protein	PIR:C64377	151.29	47.83
PH1464	1,310,854	1,312,761	>	635	hypothetical protein	PIR:C64302	219.72	57.49
PH1465	1,312,758	1,314,710	>	650	hypothetical protein	PIR:H64502	131.16	40.69
PH1470	1,316,482	1,317,324	>	280	hypothetical protein	PIR:B64326	119.38	58.47
PH1471	1,318,055	1,317,321	<	244	hypothetical protein	PIR:H64492	105.06	52.75
PH1475	1,320,074	1,321,129	>	351	hypothetical protein	owl:D88555	147.11	54.84
PH1477	1,321,143	1,321,571	>	142	hypothetical protein	PIR:A64503	52.63	45.08
PH1479	1,324,544	1,323,474	<	356	hypothetical protein	PIR:G64321	113.29	41.79
PH1481	1,324,916	1,324,554	<	120	hypothetical protein	PIR:F64361	53.60	49.11
PH1487	1,328,052	1,328,801	>	249	hypothetical protein	owl:BSY1408318	48.67	38.84
PH1502	1,338,341	1,338,712	>	123	hypothetical protein	owl:B64490	51.51	46.09
PH1506	1,339,513	1,340,259	>	248	hypothetical protein	PIR:D64362	93.75	49.79
PH1511	1,345,253	1,346,053	>	266	hypothetical protein	Swiss_Prot:P27105	82.36	51.35
PH1512	1,348,585	1,346,072	<	837	hypothetical protein	Swiss_Prot:Q10639	163.17	39.94
PH1515	1,351,226	1,350,051	<	391	hypothetical protein	PIR:D64492	149.94	46.77
PH1517	1,352,381	1,353,763	>	460	hypothetical protein	PIR:F64373	127.72	47.2
PH1519	1,354,342	1,354,797	>	151	hypothetical protein	PIR:H64318	47.91	37.67
PH1529	1,360,699	1,361,808	>	369	hypothetical protein	PIR:A64303	127.35	51.1
PH1530	1,361,805	1,362,323	>	172	hypothetical protein	PIR:H64416	79.58	51.88
PH1535	1,365,986	1,364,730	<	418	hypothetical protein	PIR:C64305	100.65	56.02
PH1536	1,366,046	1,366,474	>	142	hypothetical protein	PIR:H64302	73.29	52.71
PH1538	1,368,901	1,367,672	<	409	hypothetical protein	PIR:D64373	57.24	41.7
PH1539	1,369,277	1,368,912	<	121	hypothetical protein	owl:C64306	62.52	50
PH1548	1,381,556	1,380,732	<	274	hypothetical protein	owl:HUMORFA10	57.31	35.83
PH1549	1,382,298	1,381,549	<	249	hypothetical protein	owl:CEB05641	57.65	36.59
PH1552	1,383,777	1,383,067	<	236	hypothetical protein	PIR:H64373	123.42	54.11
PH1557	1,386,469	1,387,845	>	458	hypothetical protein	PIR:D64424	92.50	37.26
PH1558	1,387,847	1,389,184	>	445	hypothetical protein	PIR:H64328	45.90	29.9
PH1563	1,392,326	1,391,421	<	301	hypothetical protein	owl:SSTOP6AB1	81.72	38.3
PH1564	1,394,238	1,392,544	<	564	hypothetical protein	owl:C64428	129.32	54.56
PH1566	1,394,907	1,394,248	<	219	hypothetical protein	PIR:C64355	76.51	50.58
PH1567	1,395,680	1,394,880	<	266	hypothetical protein	PIR:D64355	57.51	38.57

Table 4. Continued.

ORF ID	Position		Direction	Length (aa)	Predicted Product	Locus	ZScore	% Identity
	Initiation	Termination						
PH1568	1,396,037	1,395,687	<	116	hypothetical protein	PIR:E64355	66.34	51.02
PH1569	1,396,157	1,397,161	>	334	hypothetical protein	PIR:S76356	45.43	38.07
PH1573	1,399,581	1,398,601	>	326	hypothetical protein	owl:SS56KBFR68	54.53	34.75
PH1575	1,399,637	1,400,500	>	287	hypothetical protein	PIR:C64325	67.60	37.01
PH1579	1,401,978	1,402,499	>	173	hypothetical protein	PIR:D64443	63.29	53.62
PH1580	1,403,211	1,402,483	<	242	hypothetical protein	owl:PFU629861	54.28	43.1
PH1585	1,407,177	1,408,118	>	313	hypothetical protein	PIR:E64452	92.57	57.87
PH1590	1,411,066	1,410,272	>	264	hypothetical protein	PIR:C64471	114.32	54.69
PH1594	1,415,076	1,413,799	<	425	hypothetical protein	owl:BMU17283	49.27	31.33
PH1596	1,415,837	1,416,484	>	215	hypothetical protein	PIR:E64452	47.63	39.06
PH1600	1,418,691	1,419,344	>	217	hypothetical protein	PIR:C64368	99.24	52.56
PH1601	1,419,944	1,419,582	<	120	hypothetical protein	owl:B64420	65.74	59.81
PH1602	1,422,585	1,419,970	<	871	hypothetical protein	owl:B64385	233.42	57.91
PH1609	1,427,554	1,427,078	<	158	hypothetical protein	PIR:E64339	62.41	46.98
PH1615	1,431,990	1,433,252	>	420	hypothetical protein	PIR:G64312	106.27	46.72
PH1616	1,433,360	1,434,061	>	233	hypothetical protein	PIR:H64413	56.01	46.49
PH1619	1,436,692	1,437,804	>	370	hypothetical protein	PIR:A64383	77.65	38.8
PH1623	1,441,593	1,440,829	<	254	hypothetical protein	PIR:D64305	40.32	35.05
PH1625	1,442,597	1,441,590	<	335	hypothetical protein	owl:MTU47134	64.41	41.99
PH1626	1,443,512	1,442,628	<	294	hypothetical protein	owl:SS56KBFR55	62.24	37.87
PH1627	1,444,576	1,443,536	<	346	hypothetical protein	PIR:H64482	42.32	36.11
PH1642	1,451,093	1,450,497	<	198	hypothetical protein	PIR:B64410	83.67	44.68
PH1645	1,452,810	1,454,162	>	450	hypothetical protein	PIR:C64500	168.60	51.49
PH1650	1,457,514	1,456,852	>	220	hypothetical protein	PIR:G64316	60.93	40.47
PH1655	1,460,654	1,461,379	>	241	hypothetical protein	PIR:D64479	92.91	46.4
PH1671	1,471,536	1,470,826	>	236	hypothetical protein	owl:SS56KBFR22	67.46	45.21
PH1674	1,474,964	1,473,861	<	367	hypothetical protein	PIR:C64394	40.97	47.58
PH1680	1,478,587	1,479,549	>	320	hypothetical protein	PIR:E64444	82.83	40.2
PH1692	1,487,686	1,487,231	<	151	hypothetical protein	PIR:H64318	41.83	34.19
PH1695	1,490,384	1,489,767	<	205	hypothetical protein	PIR:F64336	48.74	42.78
PH1699	1,493,603	1,492,260	<	447	hypothetical protein	PIR:E64450	150.07	55.61
PH1701	1,495,358	1,494,072	<	428	hypothetical protein	PIR:F64485	101.08	41.64
PH1702	1,496,747	1,495,836	<	303	hypothetical protein	PIR:D64328	48.34	35.04
PH1705	1,498,455	1,499,483	>	342	hypothetical protein	PIR:B64332	100.84	42.62
PH1710	1,502,036	1,502,452	>	138	hypothetical protein	PIR:G64464	43.22	32.06
PH1711	1,503,362	1,502,454	<	302	hypothetical protein	owl:BSZ939376	55.64	35.64
PH1712	1,504,372	1,503,359	<	337	hypothetical protein	owl:BSZ939375	49.48	31.69
PH1724	1,513,824	1,513,333	<	163	hypothetical protein	PIR:D64459	77.11	50.00
PH1726	1,514,963	1,513,821	<	380	hypothetical protein	PIR:B64425	89.34	45.86
PH1730	1,517,174	1,517,833	>	219	hypothetical protein	PIR:G64416	50.98	40.4
PH1731	1,518,556	1,517,960	>	198	hypothetical protein	PIR:H64415	63.41	44.97
PH1733	1,519,377	1,520,057	>	226	hypothetical protein	Swiss_Prot:P03813	53.31	36.28
PH1736	1,523,921	1,522,179	<	580	hypothetical protein	owl:TTU509512	42.13	33.45
PH1739	1,525,960	1,525,322	>	212	hypothetical protein	PIR:F64373	52.56	46.02
PH1746	1,530,203	1,530,694	>	163	hypothetical protein	PIR:G64377	52.22	43.42
PH1771	1,545,325	1,544,942	<	127	hypothetical protein	PIR:H64357	42.79	51.72
PH1782	1,553,650	1,552,451	>	399	hypothetical protein	owl:BSUWFO	55.11	42.59
PH1786	1,556,183	1,556,788	>	201	hypothetical protein	PIR:C64416	73.08	50.52
PH1788	1,557,813	1,557,097	<	238	hypothetical protein	PIR:F64373	65.91	49.55
PH1789	1,559,228	1,557,810	>	472	hypothetical protein	PIR:F64373	133.80	45.37
PH1802	1,566,987	1,567,661	>	224	hypothetical protein	Swiss_Prot:P33373	65.35	38.71
PH1804	1,569,560	1,568,235	<	441	hypothetical protein	owl:AB002406	122.62	49.88
PH1808	1,574,328	1,573,186	>	380	hypothetical protein	PIR:C64445	77.67	37.14
PH1812	1,576,013	1,577,287	>	424	hypothetical protein	PIR:F64436	117.90	47.06
PH1815	1,578,811	1,579,665	>	284	hypothetical protein	owl:D64126B	46.65	44.05
PH1819	1,582,588	1,582,037	>	183	hypothetical protein	PIR:A64330	47.28	48.63
PH1820	1,583,147	1,582,581	<	188	hypothetical protein	owl:B64335	65.61	59.41
PH1822	1,584,303	1,584,920	>	205	hypothetical protein	PIR:F64304	79.71	55.37
PH1824	1,586,939	1,585,725	>	404	hypothetical protein	PIR:G64450	177.08	62.06
PH1829	1,591,033	1,589,888	<	381	hypothetical protein	owl:B64418	74.45	38.14
PH1831	1,592,485	1,592,811	>	108	hypothetical protein	Swiss_Prot:P20297	63.91	61.29
PH1832	1,593,411	1,592,803	>	202	hypothetical protein	Swiss_Prot:P20298	125.54	83.56
PH1833	1,593,464	1,594,393	>	309	hypothetical protein	PIR:D64303	47.94	42.19
PH1839	1,597,606	1,598,271	>	221	hypothetical protein	PIR:F64418	40.41	39.47
PH1844	1,602,774	1,603,919	>	381	hypothetical protein	PIR:F64500	45.71	31.99
PH1849	1,607,392	1,606,436	<	318	hypothetical protein	owl:SS100KBFR62	47.60	43.3
PH1858	1,614,454	1,615,215	>	253	hypothetical protein	PIR:F64316	63.81	47.06
PH1867	1,623,022	1,621,169	<	617	hypothetical protein	PIR:G64441	161.08	54.44
PH1875	1,631,166	1,632,464	>	432	hypothetical protein	PIR:C64408	123.31	43.86
PH1876	1,632,507	1,632,971	>	154	hypothetical protein	PIR:F64326	49.46	37.5
PH1887	1,640,542	1,641,162	>	206	hypothetical protein	PIR:E64488	63.31	40.43
PH1888	1,642,117	1,641,128	<	329	hypothetical protein	PIR:F64368	68.95	36.21
PH1889	1,642,351	1,643,952	>	533	hypothetical protein	Swiss_Prot:P44263	91.24	38.4

Table 4. Continued.

ORF ID	Position		Direction	Length (aa)	Predicted Product	Locus	ZScore	% Identity
	Initiation	Termination						
PH1896	1,650,167	1,648,215	<	650	hypothetical protein	PIR:H64487	137.08	42.42
PH1897	1,650,928	1,650,164	<	254	hypothetical protein	PIR:E64313	101.51	45.08
PH1898	1,651,734	1,650,931	<	267	hypothetical protein	Swiss_Prot:P29085	72.32	83.12
PH1903	1,658,635	1,657,529	<	368	hypothetical protein	PIR:G64428	49.07	42.94
PH1905	1,659,043	1,659,888	>	281	hypothetical protein	PIR:E64316	109.67	51.31
PH1909	1,661,279	1,661,818	>	179	hypothetical protein	PIR:C64349	46.43	32.9
PH1912	1,662,835	1,664,109	>	424	hypothetical protein	Swiss_Prot:P46838	115.47	43.03
PH1915	1,665,742	1,664,552	<	396	hypothetical protein	PIR:C64506	126.09	51.19
PH1917	1,666,857	1,666,297	<	186	hypothetical protein	PIR:C64328	84.60	51.4
PH1921	1,669,069	1,669,974	>	301	hypothetical protein	owl:S51901	89.17	49.3
PH1928	1,675,294	1,673,894	<	466	hypothetical protein	PIR:F64373	115.34	43.21
PH1931	1,676,213	1,675,908	<	101	hypothetical protein	PIR:H64339	74.48	62.77
PH1933	1,677,463	1,676,984	<	159	hypothetical protein	PIR:F64450	55.80	39.22
PH1938	1,681,682	1,680,462	<	406	hypothetical protein	owl:MAGHPS	72.62	44.61
PH1940	1,682,726	1,682,397	<	109	hypothetical protein	Swiss_Prot:P54027	66.29	66.3
PH1945	1,686,305	1,685,451	<	284	hypothetical protein	PIR:E64471	103.10	45.64
PH1948	1,690,704	1,690,081	<	207	hypothetical protein	PIR:E64335	51.76	39.39
PH1950	1,692,423	1,693,874	>	483	hypothetical protein	PIR:A64498	106.23	42.83
PH1952	1,694,652	1,693,861	<	263	hypothetical protein	owl:BSARALMNP	48.66	34.43
PH1966	1,708,939	1,709,850	>	303	hypothetical protein	PIR:D64400	61.86	39.78
PH1968	1,709,897	1,710,718	>	273	hypothetical protein	PIR:F64499	113.30	48.3
PH1969	1,711,910	1,710,696	<	404	hypothetical protein	PIR:D64370	42.15	29.12
PH1970	1,712,673	1,711,888	<	261	hypothetical protein	owl:DSU964878	146.15	72.76
PH1971	1,713,446	1,712,670	<	258	hypothetical protein	owl:DSU964877	202.18	78.43
PH1976	1,718,773	1,718,462	<	103	hypothetical protein	owl:DSU964873	81.85	66.67
PH1977	1,719,882	1,718,770	<	370	hypothetical protein	owl:DSU964872	172.73	58.79
PH1978	1,720,483	1,719,887	<	198	hypothetical protein	owl:DSU964871	98.23	61.11
PH1980	1,721,003	1,720,515	<	162	hypothetical protein	owl:DSU96487	61.72	84.81
PH1984	1,724,135	1,723,449	<	228	hypothetical protein	PIR:H64437	64.91	41.78
PH1985	1,725,663	1,724,140	<	507	hypothetical protein	PIR:G64313	74.02	52.15
PH1986	1,726,541	1,725,660	<	293	hypothetical protein	PIR:D64456	68.48	50.3
PH1989	1,729,356	1,728,349	<	335	hypothetical protein	PIR:B64499	42.33	25.32
PH1991	1,730,522	1,729,353	<	389	hypothetical protein	PIR:B64303	55.52	44.64
PH1992	1,730,861	1,731,295	>	144	hypothetical protein	owl:AF008220188	58.76	47.79
PH1993	1,731,282	1,731,998	>	238	hypothetical protein	PIR:A64363	67.11	42.93
PHS001	53919	53683	<	78	hypothetical protein	Swiss_Prot:P42979	28.79	41.33
PHS004	195255	195431	>	58	hypothetical protein	owl:MTCCGNME6	14.68	50.00
PHS005	262723	263001	>	92	hypothetical protein	owl:BPCP1XX7	12.31	24.66
PHS007	340120	339941	<	59	hypothetical protein	Swiss_Prot:P44812	13.13	32.69
PHS008	365707	365928	>	73	hypothetical protein	PIR:S74464	29.17	40.98
PHS011	369527	369748	>	73	hypothetical protein	PIR:S74783	15.77	26.92
PHS013	370694	370422	<	90	hypothetical protein	PIR:G64308	43.9	43.18
PHS016	373401	373691	>	96	hypothetical protein	PIR:G64308	31.13	44.32
PHS017	373943	373692	<	83	hypothetical protein	PIR:F64308	14.63	33.90
PHS018	376085	375807	<	92	hypothetical protein	owl:MTCY19H522	14.79	36.36
PHS020	430977	430684	<	97	hypothetical protein	PIR:H64330	15.29	34.83
PHS024	679121	678933	<	62	hypothetical protein	PIR:E64509	17.14	28.57
PHS025	690845	691069	>	74	hypothetical protein	owl:EHARPQTOU2	15.05	34.48
PHS026	738610	738900	>	96	hypothetical protein	PIR:G64352	32.71	57.14
PHS027	820081	820305	>	74	hypothetical protein	Swiss_Prot:P08874	16.41	41.86
PHS029	845061	844852	<	69	hypothetical protein	PIR:H64501	23.04	58.14
PHS031	965807	965965	>	52	hypothetical protein	PIR:A64440	22.97	57.58
PHS033	1120932	1120681	<	83	hypothetical protein	PIR:A64347	48.12	62.20
PHS036	1163258	1163548	>	96	hypothetical protein	Swiss_Prot:P20297	14.91	57.58
PHS038	1181782	1181940	>	52	hypothetical protein	PIR:B64334	35.32	70.73
PHS039	1283813	1284007	>	64	hypothetical protein	PIR:E64354	21.53	33.93
PHS053	1598556	1598275	<	93	hypothetical protein	owl:SSHTOPR1	65.81	57.14
PHS054	1613780	1613481	<	99	hypothetical protein	PIR:A64429	20.91	44.19
PHS059	1670033	1670314	>	93	hypothetical protein	PIR:B64477	46.2	53.09

Assigned ORFs (ORF ID) were classified into four categories: similarities to genes with known function, similarities to genes with unknown function, protein sequence motifs and no information. And ORFs with similarities were listed in the sections according to these categories. The initiation and termination position, direction of reading and length of amino acid residues of ORFs are shown in the left four columns. The right four columns represent the probable gene products, the gene 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.

Table 5. List of ORFs containing protein motifs.

ORF ID	Position		Direction	Length (aa)	Motifs assigned
	Initiation	Termination			
PH0019	17334	18062	>	242	prokaryotic membrane lipoprotein lipid attachment site
PH0021	18773	20608	<	611	ATP/GTP-binding site motif A (P-loop)
PH0024	22753	23607	<	284	binding-protein-dependent transport systems inner membrane components signature
PH0025	23644	24924	<	426	prokaryotic membrane lipoprotein lipid attachment site
PH0040	37046	37480	>	144	prokaryotic membrane lipoprotein lipid attachment site
PH0041	37177	37599	>	140	prenyl group binding site (CAAX box) ; prokaryotic membrane lipoprotein lipid attachment site
PH0050	43963	44295	<	110	prokaryotic membrane lipoprotein lipid attachment site
PH0081	67104	67826	<	240	ATP/GTP-binding site motif A (P-loop) ; tonB-dependent receptor proteins signatures
PH0088	71825	72835	>	336	tubulin-beta mRNA autoregulation signal
PH0093	77775	78518	>	247	zinc carboxypeptidases, zinc-binding regions signatures
PH0109	90236	93802	>	1188	ATP/GTP-binding site motif A (P-loop)
PH0142	124718	125587	>	289	soybean trypsin inhibitor (Kunitz) protease inhibitors family signature
PH0150	130347	131654	<	435	prokaryotic membrane lipoprotein lipid attachment site
PH0162	140313	142385	>	690	tubulin-beta mRNA autoregulation signal
PH0167	144540	145697	>	385	ATP/GTP-binding site motif A (P-loop)
PH0171	149208	149891	>	227	zinc carboxypeptidases, zinc-binding regions signatures
PH0187	164378	165133	>	251	ATP/GTP-binding site motif A (P-loop)
PH0189	165877	166998	>	373	tubulin-beta mRNA autoregulation signal
PH0206	180233	181867	<	544	lipocalin signature ; prokaryotic membrane lipoprotein lipid attachment site
PH0212	187392	188579	>	395	ATP/GTP-binding site motif A (P-loop)
PH0225	199958	200863	>	301	ATP/GTP-binding site motif A (P-loop)
PH0234	208555	209490	>	311	prokaryotic membrane lipoprotein lipid attachment site
PH0235	209479	210051	<	190	ATP/GTP-binding site motif A (P-loop)
PH0244	220724	221776	<	350	neutral zinc metalloproteinases, zinc-binding region signature
PH0248	222984	223301	<	105	prokaryotic membrane lipoprotein lipid attachment site
PH0308	271764	272411	<	215	prokaryotic membrane lipoprotein lipid attachment site
PH0315	281704	282198	<	164	prokaryotic membrane lipoprotein lipid attachment site
PH0317	285069	285686	>	205	prokaryotic membrane lipoprotein lipid attachment site
PH0327	296412	296726	<	104	eukaryotic putative RNA-binding region RNP-1 signature
PH0338	306498	307487	>	329	N-6 Adenine-specific DNA methylases signature
PH0364	326591	327112	>	173	prokaryotic membrane lipoprotein lipid attachment site
PH0370	333788	334588	>	266	crystallins beta and gamma 'Greek key' motif signature
PH0377	338764	339750	>	328	aminotransferases class-II pyridoxal-phosphate attachment site
PH0380	342439	343422	>	327	lipocalin signature
PH0398	358563	359567	>	334	prokaryotic membrane lipoprotein lipid attachment site
PH0400	359657	360052	<	131	tonB-dependent receptor proteins signatures
PH0403	361571	361972	>	133	mitochondrial energy transfer proteins signature
PH0407	363908	364939	>	343	mitochondrial energy transfer proteins signature
PH0427	389252	389629	<	125	prenyl group binding site (CAAX box)
PH0437	399784	400887	>	367	ATP/GTP-binding site motif A (P-loop)
PH0442	403594	405276	>	560	ATP/GTP-binding site motif A (P-loop)
PH0444	405848	406354	<	168	prokaryotic membrane lipoprotein lipid attachment site
PH0460	420712	421278	<	188	CDP-alcohol phosphatidyltransferases signature
PH0462	421813	423405	<	530	prokaryotic membrane lipoprotein lipid attachment site
PH0485	441325	442248	<	307	prokaryotic membrane lipoprotein lipid attachment site
PH0492	446418	446762	<	114	prokaryotic membrane lipoprotein lipid attachment site
PH0513	463793	464095	<	100	prokaryotic membrane lipoprotein lipid attachment site
PH0514	464601	465659	>	352	prokaryotic membrane lipoprotein lipid attachment site
PH0515	465639	466226	<	195	ATP/GTP-binding site motif A (P-loop)
PH0524	472058	473035	>	325	prenyl group binding site (CAAX box)
PH0529	475635	476159	>	174	prokaryotic membrane lipoprotein lipid attachment site
PH0531	477064	477975	>	303	inorganic pyrophosphatase signature
PH0533	478382	478741	>	119	ATP/GTP-binding site motif A (P-loop)
PH0539	482014	483102	>	362	ATP/GTP-binding site motif A (P-loop)
PH0560	499586	501511	>	641	myc-type, 'helix-loop-helix' dimerization domain signature
PH0573	510838	511263	<	141	prokaryotic membrane lipoprotein lipid attachment site
PH0584	521768	522697	<	309	N-4 cytosine-specific DNA methylases signature
PH0610	548439	548936	<	165	DEAD and DEAH box families ATP-dependent helicases signatures
PH0611	548979	549737	<	252	ATP/GTP-binding site motif A (P-loop)
PH0626	562210	563367	<	385	tubulin-beta mRNA autoregulation signal
PH0638	572750	573214	>	154	prokaryotic membrane lipoprotein lipid attachment site
PH0642	575675	576463	>	262	ATP/GTP-binding site motif A (P-loop)
PH0659	591402	591878	>	158	prokaryotic membrane lipoprotein lipid attachment site
PH0672	598606	599535	<	309	prokaryotic membrane lipoprotein lipid attachment site
PH0688	611676	612116	<	146	prokaryotic membrane lipoprotein lipid attachment site
PH0692	616335	617195	>	286	aminoacyl-transfer RNA synthetases class-II signatures
PH0697	620786	622711	>	641	ATP/GTP-binding site motif A (P-loop)
PH0712	634778	635470	>	230	neutral zinc metalloproteinases, zinc-binding region signature
PH0753	664048	665442	>	464	prokaryotic membrane lipoprotein lipid attachment site
PH0770	679289	679753	>	154	mitochondrial energy transfer proteins signature
PH0788	696280	698730	>	816	ATP/GTP-binding site motif A (P-loop) ; sugar transport proteins signatures
PH0791	699729	700454	>	241	ATP/GTP-binding site motif A (P-loop)
PH0794	701861	702574	>	237	arginase family signatures

Table 5. Continued.

ORF ID	Position		Direction	Length (aa)	Motifs assigned
	Initiation	Termination			
PH0799	707382	708185	<	267	ATP/GTP-binding site motif A (P-loop)
PH0812	721811	722956	>	381	prokaryotic membrane lipoprotein lipid attachment site
PH0837	748438	749628	>	396	ATP/GTP-binding site motif A (P-loop)
PH0840	750009	750362	>	117	ATP/GTP-binding site motif A (P-loop)
PH0844	752014	752325	>	103	myb DNA-binding domain repeat signatures
PH0858	766077	767156	<	359	aspartokinase signature
PH0873	779958	782324	>	788	ATP/GTP-binding site motif A (P-loop)
PH0877	783301	783687	>	128	prokaryotic membrane lipoprotein lipid attachment site
PH0893	798863	799348	<	161	4Fe-4S ferredoxins, iron-sulfur binding region signature ; cytochrome c family heme-binding site signature
PH0899	805294	806739	>	481	ATP/GTP-binding site motif A (P-loop)
PH0905	813379	816369	<	996	N-6 Adenine-specific DNA methylases signature
PH0911	821342	822115	<	257	ABC-2 type transport system integral membrane proteins signature
PH0917	824928	826508	<	526	ATP/GTP-binding site motif A (P-loop)
PH0922	830046	831353	<	435	ATP/GTP-binding site motif A (P-loop) ; prokaryotic membrane lipoprotein lipid attachment site
PH0938	847693	848409	>	238	prenyl group binding site (CAAX box)
PH0943	852806	853294	<	162	sugar transport proteins signatures
PH0952	858412	860673	<	753	ATP/GTP-binding site motif A (P-loop)
PH0975	887481	887813	<	110	renal dipeptidase active site
PH0980	892654	893232	<	192	ATP/GTP-binding site motif A (P-loop)
PH0994	902821	903144	>	107	prokaryotic membrane lipoprotein lipid attachment site
PH1018	923431	923739	>	102	prenyl group binding site (CAAX box)
PH1025	931110	931820	<	236	ATP/GTP-binding site motif A (P-loop)
PH1039	946342	947604	<	420	prokaryotic membrane lipoprotein lipid attachment site
PH1040	947219	947731	>	170	prokaryotic membrane lipoprotein lipid attachment site
PH1045	952118	952699	<	193	eukaryotic putative RNA-binding region RNP-1 signature
PH1049	956285	956938	<	217	alpha-isopropylmalate and homocitrate synthases signatures
PH1053	958656	959075	>	139	prokaryotic membrane lipoprotein lipid attachment site
PH1067	972341	973405	<	354	ATP/GTP-binding site motif A (P-loop) ; G-protein coupled receptors signature
PH1079	982737	983402	>	221	prokaryotic membrane lipoprotein lipid attachment site
PH1086	986746	987786	>	346	aspartokinase signature
PH1098	997992	999203	<	403	binding-protein-dependent transport systems inner membrane componentssignature
PH1130	1025340	1025852	>	170	prokaryotic membrane lipoprotein lipid attachment site
PH1133	1028233	1028538	>	101	G-protein coupled receptors signature
PH1141	1036215	1037723	>	502	prokaryotic membrane lipoprotein lipid attachment site
PH1142	1037774	1039135	>	453	tubulin-beta mRNA autoregulation signal
PH1147	1042411	1042737	>	108	prokaryotic membrane lipoprotein lipid attachment site
PH1153	1047605	1048669	>	354	argE / dapE / ACY1 / CPG2 / yscS family signatures
PH1174	1062375	1062980	<	201	site-specific recombinases signatures
PH1181	1068183	1068614	>	143	prokaryotic membrane lipoprotein lipid attachment site
PH1188	1071845	1073299	<	484	ATP/GTP-binding site motif A (P-loop)
PH1190	1074266	1074694	<	142	prokaryotic membrane lipoprotein lipid attachment site
PH1211	1090616	1091200	<	194	prokaryotic membrane lipoprotein lipid attachment site
PH1212	1091185	1091715	<	176	prokaryotic membrane lipoprotein lipid attachment site ; thermonuclease family signatures
PH1219	1097995	1099266	>	423	CDP-alcohol phosphatidyltransferases signature
PH1230	1106970	1107731	>	253	ATP/GTP-binding site motif A (P-loop)
PH1246	1121204	1123387	<	727	ATP/GTP-binding site motif A (P-loop)
PH1248	1122682	1123041	>	119	prenyl group binding site (CAAX box)
PH1249	1123375	1124004	<	209	prokaryotic membrane lipoprotein lipid attachment site
PH1259	1131879	1133117	>	412	RNA methyltransferase trmA family signatures
PH1269	1139999	1140595	>	198	prokaryotic membrane lipoprotein lipid attachment site
PH1271	1141554	1143830	<	758	sugar transport proteins signatures
PH1272	1142347	1142652	>	101	prokaryotic membrane lipoprotein lipid attachment site
PH1279	1149140	1149709	>	189	prokaryotic membrane lipoprotein lipid attachment site
PH1286	1155865	1156206	>	113	prokaryotic membrane lipoprotein lipid attachment site
PH1288	1157631	1158998	<	455	argE / dapE / ACY1 / CPG2 / yscS family signatures
PH1293	1161633	1161971	<	112	prokaryotic membrane lipoprotein lipid attachment site
PH1298	1167611	1167964	<	117	sugar transport proteins signatures
PH1309	1179699	1180280	<	193	ATP/GTP-binding site motif A (P-loop)
PH1324	1191088	1191465	<	125	prokaryotic membrane lipoprotein lipid attachment site
PH1334	1199768	1201843	>	691	EF-hand calcium-binding domain
PH1335	1201856	1202953	>	365	tubulin subunits alpha, beta, and gamma signature
PH1352	1217527	1219206	<	559	binding-protein-dependent transport systems inner membrane componentssignature
PH1373	1240264	1241007	>	247	ATP/GTP-binding site motif A (P-loop)
PH1394	1256816	1257763	>	315	legume lectins signatures
PH1409	1267751	1269916	>	721	prokaryotic membrane lipoprotein lipid attachment site
PH1420	1277409	1278263	<	284	sugar transport proteins signatures
PH1428	1283395	1283760	>	121	G-protein coupled receptors signature
PH1431	1285086	1286618	>	510	prokaryotic membrane lipoprotein lipid attachment site
PH1436	1287660	1287965	<	101	ATP/GTP-binding site motif A (P-loop)
PH1440	1290261	1290671	>	136	4Fe-4S ferredoxins, iron-sulfur binding region signature
PH1443	1293938	1294282	>	114	prokaryotic membrane lipoprotein lipid attachment site
PH1445	1294198	1294542	<	114	prokaryotic membrane lipoprotein lipid attachment site
PH1446	1295215	1295835	<	206	4Fe-4S ferredoxins, iron-sulfur binding region signature

Table 5. Continued.

ORF ID	Position		Direction	Length (aa)	Motifs assigned
	Initiation	Termination			
PH1460	1307360	1307677	>	105	prokaryotic membrane lipoprotein lipid attachment site
PH1490	1330745	1331413	<	222	ABC transporters family signature ; ATP/GTP-binding site motif A (P-loop)
PH1507	1340308	1342209	>	633	ATP/GTP-binding site motif A (P-loop)
PH1513	1348563	1349135	<	190	short-chain dehydrogenases/reductases family signature ; prokaryotic membrane lipoprotein lipid attachment site
PH1523	1356256	1356966	<	236	prokaryotic membrane lipoprotein lipid attachment site
PH1525	1357794	1358126	>	110	prokaryotic membrane lipoprotein lipid attachment site
PH1532	1363575	1364255	<	226	ATP/GTP-binding site motif A (P-loop)
PH1533	1364305	1364706	<	133	aldehyde dehydrogenases active sites
PH1643	1451148	1451633	>	161	prokaryotic membrane lipoprotein lipid attachment site
PH1648	1455457	1456878	>	473	cytosolic fatty-acid binding proteins signature
PH1649	1456191	1456526	<	111	prokaryotic membrane lipoprotein lipid attachment site
PH1653	1458342	1459124	<	260	ABC transporters family signature ; ATP/GTP-binding site motif A (P-loop)
PH1659	1462917	1463954	<	345	prokaryotic membrane lipoprotein lipid attachment site
PH1690	1485871	1486293	>	140	prokaryotic membrane lipoprotein lipid attachment site
PH1707	1500279	1500650	<	123	prokaryotic membrane lipoprotein lipid attachment site
PH1708	1500886	1501989	>	367	prokaryotic membrane lipoprotein lipid attachment site
PH1714	1505907	1507124	<	405	prokaryotic membrane lipoprotein lipid attachment site
PH1715	1507478	1508455	<	325	argE / dapE / ACY1 / CPG2 / yscS family signatures
PH1732	1518865	1519305	<	146	tyrosine specific protein phosphatases signature and profiles
PH1745	1529491	1529865	<	124	prokaryotic membrane lipoprotein lipid attachment site
PH1749	1532245	1533693	>	482	2Fe-2S ferredoxins, iron-sulfur binding region signature
PH1750	1533690	1534196	>	168	4Fe-4S ferredoxins, iron-sulfur binding region signature
PH1784	1555029	1555667	<	212	neutral zinc metalloproteinases, zinc-binding region signature
PH1793	1560073	1560492	<	139	prokaryotic membrane lipoprotein lipid attachment site
PH1796	1561770	1562375	<	201	prokaryotic membrane lipoprotein lipid attachment site
PH1803	1567673	1568242	>	189	prenyl group binding site (CAAX box)
PH1806	1570882	1571640	<	252	ATP/GTP-binding site motif A (P-loop)
PH1814	1578457	1578786	<	109	prokaryotic membrane lipoprotein lipid attachment site
PH1835	1594371	1594898	>	175	methylated-DNA--protein-cysteine methyltransferase active site
PH1845	1603951	1604841	>	296	pfkB family of carbohydrate kinases signatures
PH1851	1608039	1608521	>	160	prokaryotic membrane lipoprotein lipid attachment site
PH1902	1656435	1657532	>	365	bacterial regulatory proteins, araC family signature and profile
PH1906	1659543	1659857	<	104	ATP/GTP-binding site motif A (P-loop)
PH1914	1664106	1664555	>	149	ATP/GTP-binding site motif A (P-loop)
PH1919	1667414	1668337	>	307	short-chain dehydrogenases/reductases family signature
PH1927	1672830	1673897	>	355	myb DNA-binding domain repeat signatures
PH1929	1674708	1675232	>	174	prokaryotic membrane lipoprotein lipid attachment site
PH1936	1679044	1679697	<	217	hypothetical cof family signatures
PH1943	1684723	1685454	>	243	mitochondrial energy transfer proteins signature ; myb DNA-binding domain repeat signatures
PH1979	1719912	1720460	>	182	prokaryotic membrane lipoprotein lipid attachment site
PHS002	193417	193599	>	59	Mitochondrial energy transfer proteins signature
PHS003	195076	195234	>	52	ATP/GTP-binding site motif A (P-loop)
PHS006	284711	284863	>	50	prenyl group binding site (CAAX box)
PHS009	367015	367212	>	65	prokaryotic membrane lipoprotein lipid attachment site
PHS010	367456	367656	>	66	prokaryotic membrane lipoprotein lipid attachment site
PHS012	369684	369887	<	67	prokaryotic membrane lipoprotein lipid attachment site
PHS014	370704	370907	<	67	EF-hand calcium-binding domain
PHS015	371086	371253	>	55	signal peptidases I signatures
PHS022	566879	567142	<	87	endoplasmic reticulum targeting sequence
PHS028	820551	820754	<	67	prokaryotic membrane lipoprotein lipid attachment site
PHS030	941443	941616	<	57	prenyl group binding site (CAAX box)
PHS049	1547968	1548228	<	86	phosphopantetheine attachment site
PHS058	1670020	1670253	<	77	Glyceraldehyde 3-phosphate dehydrogenase active site; Prokaryotic membrane lipoprotein lipid attachment site