10. South, T. L., Blake, P. R., Hare, D. R. & Summers, M. F. C-terminal retroviral-type zinc finger domain from the HIV-1 nucleocapsid protein is structurally similar to the N-terminal zinc finger domain. *Biochemistry* **30,** 6342–6349 (1991).
11. Luisi, B. F. *et al*. Crystallographic analysis of the interaction of the glucorticoid receptor with DNA. *Nature* **352,** 497–505 (1991).
12. Marmorstein, R., Carey, M., Ptashne, M. & Harrison, S. C. DNA recognition by GAL4: structure of a protein–DNA complex. *Nature* **356,** 408–414 (1992).
13. Everett, R. D. *et al*. A novel arrangement of zinc-binding residues and secondary structure in the C3HC4 motif of an alpha herpes virus protein family. *J. Mol. Biol.* **234,** 1038–1047 (1993).
14. Barlow, P. N., Luisi, B., Milner, A., Elliot, M. & Everett, R. Structure of the C_3HC_4 domain by ^{1}H-nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **237,** 201–211 (1994).
15. Borden, K. L. B. *et al*. The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncoprotein PML. *EMBO J.* **14,** 1532–1541 (1995).
16. Phillips, S. E. V. The β-ribbon DNA recognition motif. *Ann. Rev. Biophys. Biomol. Struct.* **23,** 671–701 (1994).
17. Kim, J. L., Nikolov, D. B. & Burley, S. K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365,** 520–527 (1993).
18. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP–TATA-box complex. *Nature* **365,** 512–520 (1993).
19. Schumacher, M. A., Choi, K. Y., Zalkin, H. & Brennan, R. G. Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by α-helices. *Science* **266,** 763–770 (1994).
20. Flick, K. E. *et al*. Crystallization and preliminary X-ray studies of I-PpoI: a nuclear, intron-encoded homing endonuclease from *Physarum polycephalum*. *Protein Sci.* **6,** 1–4 (1997).
21. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
22. Leslie, A. G. W. in *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography* (Daresbury Laboratory, Warrington, UK, 1992).
23. CCP4 *The SERC (UK) Collaborative Computing Project No. 4, a Suite of Programs for Protein Crystallography* (Daresbury Laboratory, Warrington, UK, 1979).
24. QUANTA96 *X-ray Structure Analysis User's Reference* (Molecular Simulations, San Diego, 1996).
25. Brünger, A. *XPLOR version 3.1: A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992).
26. Laskowski, R. J., Macarthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26,** 383–290 (1993).
27. Evans, S. V. SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics* **11,** 134–138 (1993).

Acknowledgements. We thank D. McHugh, K. Stephens and J. D. Heath for initial subcloning, purification and crystallization studies; R. Strong, K. Zhang and B. Scott for advice during the crystallographic analysis; and the beamline staff at the Advanced Light Source (NLBL laboratories), beamline 5.0.2, particularly T. Earnest, for assistance. B.L.S. and R.J.M. are funded for this project by the NIH. K.E.F. was supported by an NIH training grant and the American Heart Associaiton. M.S.J. was supported by an NSF fellowship and an NIH training grant.

Correspondence and requests for materials and coordinates should be addressed to B.L.S. (e-mail: bstoddar@fred.fhcrc.org). Coordinates have been deposited in the Brookhaven Protein Data Bank (accession nos lipp, 1a73, 1a74).

corrections

Emergence of symbiosis in peptide self-replication through a hypercyclic network

David H. Lee, Kay Severin, Yohei Yokobayashi & M. Reza Ghadiri

Nature **390,** 591–594 (1997)

Hypercycles are based on second-order (or higher) autocatalysis and defined by two or more replicators that are connected by another superimposed autocatalytic cycle. Our study describes a mutualistic relationship between two replicators, each catalysing the formation of the other, that are linked by a superimposed catalytic cycle. Although the kinetic data suggest the intermediary of higher-order species in the autocatalytic processes, the present system should not be referred to as an example of a minimal hypercycle in the absence of direct experimental evidence for the autocatalytic cross-coupling between replicators. □

The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*

Hans-Peter Klenk, Rebecca A. Clayton, Jean-Francois Tomb, Owen White, Karen E. Nelson, Karen A. Ketchum, Robert J. Dodson, Michelle Gwinn, Erin K. Hickey, Jeremy D. Peterson, Delwood L. Richardson, Anthony R. Kerlavage, David E. Graham, Nikos C. Kyrpides, Robert D. Fleischmann, John Quackenbush, Norman H. Lee, Granger G. Sutton, Steven Gill, Ewen F. Kirkness, Brian A. Dougherty, Keith McKenney, Mark D. Adams, Brendan Loftus, Scott Peterson, Claudia I. Reich, Leslie K. McNeil, Jonathan H. Badger, Anna Glodek, Lixin Zhou, Ross Overbeek, Jeannine D. Gocayne, Janice F. Weidman, Lisa McDonald, Teresa Utterback, Matthew D. Cotton, Tracy Spriggs, Patricia Artiach, Brian P. Kaine, Sean M. Sykes, Paul W. Sadow, Kurt P. D'Andrea, Cheryl Bowman, Claire Fujii, Stacey A. Garland, Tanya M. Mason, Gary J. Olsen, Claire M. Fraser, Hamilton O. Smith, Carl R. Woese & J. Craig Venter

Nature **390,** 364–370 (1997)

The pathway for sulphate reduction is incorrect as published: in Fig. 3 on page 367, adenylyl sulphate 3-phosphotransferase (*cysC*) is not needed in the pathway as outlined, as adenylyl sulphate reductase (*aprAB*) catalyses the first step in the reduction of adenylyl sulphate. The correct sequence of reactions is: sulphate is first activated to adenylyl sulphate, then reduced to sulphite and subsequently to sulphide. The enzymes catalysing these reactions are: sulphate adenylyltransferase (*sat*), adenylylsulphate reductase (*aprAB*), and sulphite reductase (*dsrABD*). We thank Jens-Dirk Schwenn for bringing this error to our attention. □

The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*

Hans-Peter Klenk*, Rebecca A. Clayton*, Jean-Francois Tomb*, Owen White*, Karen E. Nelson*, Karen A. Ketchum*, Robert J. Dodson*, Michelle Gwinn*, Erin K. Hickey*, Jeremy D. Peterson*, Delwood L. Richardson*, Anthony R. Kerlavage*, David E. Graham†, Nikos C. Kyrpides†, Robert D. Fleischmann*, John Quackenbush*, Norman H. Lee*, Granger G. Sutton*, Steven Gill*, Ewen F. Kirkness*, Brian A. Dougherty*, Keith McKenney*, Mark D. Adams*, Brendan Loftus*, Scott Peterson*, Claudia I. Reich†, Leslie K. McNeil†, Jonathan H. Badger†, Anna Glodek*, Lixin Zhou*, Ross Overbeek‡, Jeannine D. Gocayne*, Janice F. Weidman*, Lisa McDonald*, Teresa Utterback*, Matthew D. Cotton*, Tracy Spriggs*, Patricia Artiach*, Brian P. Kaine†, Sean M. Sykes*, Paul W. Sadow*, Kurt P. D'Andrea*, Cheryl Bowman*, Claire Fujii*, Stacey A. Garland*, Tanya M. Mason*, Gary J. Olsen†, Claire M. Fraser*, Hamilton O. Smith*, Carl R. Woese† & J. Craig Venter*

* *The Institute for Genomic Research (TIGR), Rockville, Maryland 20850, USA*
† *Department of Microbiology, University of Illinois, Champaign-Urbana, Illinois 61801, USA*
‡ *Mathematics and Computer Science Division, Argonne National Laboratory, Illinois 60439, USA*

Archaeoglobus fulgidus **is the first sulphur-metabolizing organism to have its genome sequence determined. Its genome of 2,178,400 base pairs contains 2,436 open reading frames (ORFs). The information processing systems and the biosynthetic pathways for essential components (nucleotides, amino acids and cofactors) have extensive correlation with their counterparts in the archaeon** ***Methanococcus jannaschii*****. The genomes of these two Archaea indicate dramatic differences in the way these organisms sense their environment, perform regulatory and transport functions, and gain energy. In contrast to** ***M. jannaschii*****,** ***A. fulgidus*** **has fewer restriction–modification systems, and none of its genes appears to contain inteins. A quarter (651 ORFs) of the** ***A. fulgidus*** **genome encodes functionally uncharacterized yet conserved proteins, two-thirds of which are shared with** ***M. jannaschii*** **(428 ORFs). Another quarter of the genome encodes new proteins indicating substantial archaeal gene diversity.**

Biological sulphate reduction is part of the global sulphur cycle, ubiquitous in the earth's anaerobic environments, and is essential to the basal workings of the biosphere. Growth by sulphate reduction is restricted to relatively few groups of prokaryotes; all but one of these are Eubacteria, the exception being the archaeal sulphate reducers in the Archaeoglobales[1,2]. These organisms are unique in that they are unrelated to other sulphate reducers, and because they grow at extremely high temperatures[3]. The known Archaeoglobales are strict anaerobes, most of which are hyperthermophilic marine sulphate reducers found in hydrothermal environments[2,4] and in subsurface oil fields[5]. High-temperature sulphate reduction by *Archaeoglobus* species contributes to deep subsurface oil-well 'souring' by producing iron sulphide, which causes corrosion of iron and steel in oil- and gas-processing systems[5].

Archaeoglobus fulgidus VC-16 (refs 2, 4) is the type strain of the Archaeoglobales. Cells are irregular spheres with a glycoprotein envelope and monopolar flagella. Growth occurs between 60 and 95 °C, with optimum growth at 83 °C and a minimum division time of 4 h. The organism grows organoheterotrophically using a variety of carbon and energy sources, but can grow lithoautotrophically on hydrogen, thiosulphate and carbon dioxide[6]. We sequenced the genome of *A. fulgidus* strain VC-16 as an example of a sulphur-metabolizing organism and to gain further insight into the Archaea[7,8] through genomic comparison with *Methanococcus jannaschii*[9].

General features of the genome

The genome of *A. fulgidus* consists of a single, circular chromosome of 2,178,400 base pairs (bp) with an average of 48.5% G+C content (Fig. 1). There are three regions with low G+C content (<39%), two rich in genes encoding enzymes for lipopolysaccharide (LPS) biosynthesis, and two regions of high G+C content (>53%), containing genes for large ribosomal RNAs, proteins involved in haem biosynthesis (*hemAB*), and several transporters (Table 1). Because the origins of replication in Archaea are not characterized, we arbitrarily designated base pair one within a presumed non-coding region upstream of one of three areas containing multiple short repeat elements.

Open reading frames. Two independent coding analysis programs and BLASTX[10] searches (see Methods) predicted 2,436 ORFs (Figs 1, 2, Tables 1, 2) covering 92.2% of the genome. The average size of the *A. fulgidus* ORFs is 822 bp, similar to that of *M. jannaschii* (856 bp), but smaller than that in the completely sequenced eubacterial genomes (949 bp). All ORFs were searched against a non-redundant protein database, resulting in 1,797 putative identifications that were assigned biological roles within a classification system adapted from ref. 11. Predicted start codons are 76% ATG, 22% GTG and 2% TTG. Unlike *M. jannaschii*, where 18 inteins were found in coding regions, no inteins were identified in *A. fulgidus*. Compared with *M. jannaschii*, *A. fulgidus* contains a large number of gene duplications, contributing to its larger genome size. The average protein relative molecular mass (M_r) in *A. fulgidus* is 29,753, ranging from 1,939 to 266,571, similar to that observed in other prokaryotes. The isoelectric point (pI) of predicted proteins among sequenced prokaryotes exhibits a bimodal distribution with peaks at pIs of approximately 5.5 and 10.5. The exceptions to this are *Mycoplasma genitalium* in which the distribution is skewed towards high pI

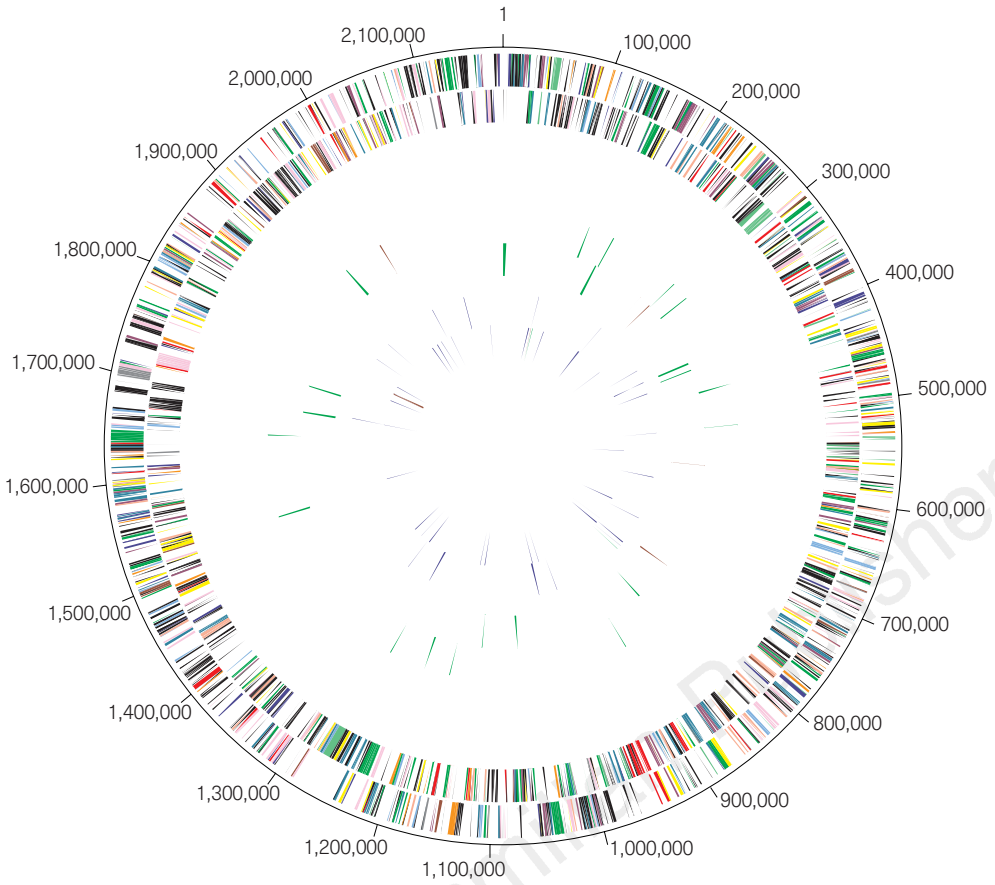

Figure 1 Circular representation of the *A. fulgidus* genome. The outer circle shows predicted protein-coding regions on the plus strand classified by function according to the colour code in Fig. 2 (except for unknowns and hypotheticals, which are in black). Second circle shows predicted protein-coding regions on the minus strand. Third and fourth circles show IS elements (red) and other repeats (green) on the plus and minus strand. Fifth and sixth circles show tRNAs (blue), rRNAs (red) and sRNAs (green) on the plus and minus strand, respectively.

Table 1 Genome features

General		
Chromosome size:	2,178,400 bp	
Protein coding regions:	92.2%	
Stable RNAs:	0.4%	
Predicted protein coding sequences:	2,436 (1.1 per kb)	
Identified by database match:	1,797	
putative function assigned:	1,096	
homologues of *M. jannaschii* ORFs:	916	
conserved hypothetical proteins:	651	
No database match:	639	
Members of 242 paralogous families:	719	
Members of 158 families with known functions:	475	
Stable RNAs	*Coordinates*	
16S rRNA:	1,790,478–1,788,987	
23S rRNA	1,788,751–1,785,820	
5S rRNA:	81,144–81,021	
7S RNA:	798,067–798,376	
RNase P:	86,281–86,032	
46 species of tRNA:	no significant clusters	
tRNAs with 15–62 bp introns:	Asp^{GUC}, Glu^{UUC}, Leu^{CAA}, Trp^{CCA}, Tyr^{GUA}	
Distinct G+C content regions	*Coordinates*	
HGC-1, >53% G+C	1,786,000–1,797,000	
HGC-2, >53% G+C	2,158,000–2,159,000	
LGC-1, <39% G+C	281,000–284,000	
LGC-2, <39% G+C	544,000–550,000	
LGC-3, <39% G+C	1,175,000–1,177,000	
Short, non-coding repeats	*Coordinates*	
SR-1A, CTTTCAATCCCATTTTGGTCTGATTTCAAC	147–4,213	
SR-1B, CTTTCAATCCCATTTTGGTCTGATTTCAAC	398,368–401,590	
SR-2, CTTTCAATCTCCATTTTCAGGGCCTCCCTTTCTTA	1,690,930–1,694,104	
Long, coding repeats	*Length*	*Copy number*
LR-01 NADH-flavin oxidoreductase	1,886 bp	2 copies
LR-02 NifS, NifU + ORF	1,549 bp	2 copies
LR-03 ISA1214 putative transposase + ISORF2	1,214 bp	6 copies
LR-04 ISA1083 putative transposase + ISORF2	1,083 bp	3 copies
LR-05 type II secretion system protein	1,014 bp	4 copies
LR-06 ISA0963 putative transposase	963 bp	7 copies
LR-07 homologue of MJ0794	836 bp	3 copies
LR-08 conserved hypothetical protein	696 bp	2 copies
LR-09 conserved hypothetical protein	628 bp	2 copies

(median, 9.8) and *A. fulgidus* where the skew is toward low pI (median, 6.3).

Multigene families. In *A. fulgidus* 719 genes (30% of the total) belong to 242 families with two or more members (Table 1). Of these families, 157 contained genes with biological roles. Most of these families contain genes assigned to the 'energy metabolism', 'transport and binding proteins', and 'fatty acid and phospholipid metabolism' categories (Table 2). The superfamily of ATP-binding subunits of ABC transporters is the largest, containing 40 members. The importance of catabolic degradation and signal recognition systems is reflected by the presence of two large superfamilies: acyl-CoA ligases and signal-transducing histidine kinases. *A. fulgidus* does not contain a homologue of the large 16-member family found in *M. jannaschii*[9].

Repetitive elements. Three regions of the *A. fulgidus* genome contain short (<40 bp) direct repeats (Table 1). Two regions (SR-1A and SR-1B) contain 48 and 60 copies, respectively, of an identical 30-bp repeat interspersed with unique sequences averaging 40 bp. The third region (SR-2) contains 42 copies of a 37-bp repeat similar in sequence to the SR-1 repeat and interspersed with unique sequence averaging 41 bp. These repeated sequences are similar to the short repeated sequences found in *M. jannaschii*.

Nine classes of long (>500 bp) repeated sequences with ≥95% sequence identity were found (LR1-LR9; Table 1). LR-3 is a novel element with 14-bp inverted repeats and two genes, one of which has weak similarity to a transposase from *Halobacterium salinarium*. One copy of LR-3 interrupts AF2090, a homologue of a large *M. jannaschii* gene encoding a protein of unknown function. LR-4 and LR-6 encode putative transposases not identified in *M. jannaschii* that may represent IS elements. The remaining LR elements are not similar to known IS elements.

Central intermediary and energy metabolism

Sulphur oxide reduction may be the dominant respiratory process in anaerobic marine and freshwater environments, and is an important aspect of the sulphur cycle in anaerobic ecosystems[12]. In this pathway, sulphate (SO_4^{2-}) is first activated to adenylylsulphate (adenosine-5′-phosphosulphate; APS), then reduced to sulphite and subsequently to sulphide[1,13] (Fig. 3). The most important enzyme in dissimilatory sulphate reduction, adenylylsulphate reductase, reduces the activated sulphate to sulphite, releasing AMP. In *A. fulgidus*, the APS reductase has a high degree of similarity and identical physiological properties to APS reductases in sulphate-reducing delta proteobacteria[14]. A desulphoviridin-type sulphite reductase then adds six electrons to sulphite to produce sulphide. As in the Eubacteria, three sulphite-reductase genes, *dsrABD*, constitute an operon. The genes for adenylylsulphate reductase and sulphate adenylyltransferase reside in a separate operon. In *A. fulgidus*, sulphate can be replaced as an electron acceptor by both thiosulphate ($S_2O_3^{2-}$) and sulphite (SO_3^{2-}), but not by elemental sulphur.

A. fulgidus VC-16 has been shown to use lactate, pyruvate, methanol, ethanol, 1-propanol and formate as carbon and energy sources[2]. Glucose has been described as a carbon source[1], but neither an uptake-transporter nor a catabolic pathway could be identified. Although it has been reported that *A. fulgidus* is incapable of growth on acetate[6], multiple genes for acetyl-CoA synthetase (which converts acetate to acetyl-CoA) were found. The organism may degrade a variety of hydrocarbons and organic acids because of the presence of 57 β-oxidation enzymes, at least one lipase, and a minimum of five types of ferredoxin-dependent oxidoreductases (Fig. 3). The predicted β-oxidation system is similar to those in Eubacteria and mitochondria, and has not previously been described in the Archaea. *Escherichia coli* requires both the *fadD* and *fadL* gene products to import long-chain fatty acids across the cell envelope into the cytosol[15]. *A. fulgidus* has 14 acyl-CoA ligases related to FadD, but as expected given that it has no outer membrane, no FadL. In *E. coli*, FadB has several metabolic functions, but in *A. fulgidus* these functions seem to be distributed among separate enzymes. For example, AF0435 encodes an orthologue of enoyl-CoA hydratase and resembles the amino-terminal domain of FadB. This gene is immediately upstream of a gene encoding an orthologue of 3-hydroxyacyl-CoA dehydrogenase that resembles the carboxy-terminal domain of FadB.

Acetyl-CoA is degraded by *A. fulgidus* through a C_1-pathway, not by the citric acid cycle or glyoxylate bypass[6,16,17]. This degradation is catalysed through the carbon monoxide dehydrogenase (CODH) pathway that consists of a five-subunit acetyl-CoA decarboxylase/synthase complex (ACDS) and five enzymes that are typically involved in methanogenesis[18]. In *A. fulgidus*, however, reverse methanogenesis occurs, resulting in CO_2 production. All of the enzymes and cofactors of methanogenesis from formylmethanofuran to N^5-methyltetrahydromethanopterin are used, but the absence of methyl-CoM reductase eliminates the possibility of methane production by conventional pathways. Production of trace amounts of methane ($<0.1\ \mu mol\ ml^{-1}$)[19] is probably a result of the reduction of N^5-methyltetrahydromethanopterin to methane and tetrahydromethanopterin by carbon monoxide (CO) dehydrogenase.

A. fulgidus also contains genes suggesting it has a second CO dehydrogenase system, homologous to that which enables *Rhodospirillum rubrum* to grow without light using CO as its sole energy source. Genes were detected for the nickel-containing CO dehydrogenase (CooS), an iron–sulphur redox protein, and a protein associated with the incorporation of nickel in CooS. These represent elements of a system that could catalyse the conversion of CO and H_2O to CO_2 and H_2.

In contrast to *M. jannaschii*, *A. fulgidus* contains genes representing multiple catabolic pathways. Systems include CoA-SH-dependent ferredoxin oxidoreductases specific for pyruvate, 2-ketoisovalerate, 2-ketoglutarate and indolepyruvate, as well as a 2-oxoacid with little substrate specificity[20,21]. Four genes with similarity to the tungsten-containing aldehyde ferredoxin oxidoreductase were also found[22].

Biochemical pathways characteristic of eubacterial metabolism, including the pentose-phosphate pathway, the Entner–Doudoroff pathway, glycolysis and gluconeogenesis, are either completely absent or only partly represented (Fig. 3). *A. fulgidus* does not have typical eubacterial polysaccharide biosynthesis machinery, yet it has been shown to produce a protein and carbohydrate-containing biofilm[23]. Nitrogen is obtained by importing inorganic molecules or degrading amino acids (Fig. 3); neither a glutamate dehydrogenase nor a relevant *fix* or *nif* gene is present.

The $F_{420}H_2$:quinone oxidoreductase complex[24] is recognized as

Figure 2 Linear representation of the *A. fulgidus* genome illustrating the location of each predicted protein-coding region, RNA gene, and repeat element in the genome. Symbols for the transporters are as follows: AsO, arsenite; COH, sugar; P_i, phosphate; aa2, dipeptide; NH_4^+, ammonium; a/o, arginine/lysine/ornithine; s/p, spermidine/putrescine; glyc, glycerol; Cl^-, chloride; Fe^{2+}, iron(II); Fe^{3+}, iron(III); I, L, V, branched-chain amino acids; P, proline; pan, pantothenate; rib, ribose; lac, lactate; Mg^{2+}/Co^{2+}, magnesium and cobalt; gln, glutamine; NO^{3-}, nitrate; ox/for, oxalate/formate; maln, malonic acid; Hg^{2+}, mercury; phs, polysaccharide; SO_4^{2-}, sulphate; OCN^-, cyanate; hex, hexuronate; phs, polysialic acid; K^+, potassium channel; H^+/Na^+, sodium/proton antiporter; Na^+/Cl^-, sodium- and chloride-dependent transporter; P/G, osmoprotection protein; Cu^{2+}, copper-transporting ATPase; +?, cation-transporting ATPase; ?, ABC-transporter without known function. Triplets associated with tRNAs represent the anticodon sequence. Numbers associated with GES represent the number of membrane-spanning domains (MSDs) according to Goldman, Engelman and Steiz scale as determined by TopPred[39]. Genes whose identification is based on genes in *M. jannaschii* are indicated by circles. Of the 236 proteins containing at least one MSD, 124 of these had two or more MSDs.

the main generator of proton-motive force. However, our analysis indicates the presence of heterodisulphide reductase and several molybdopterin-binding oxidoreductases, with polysulphide, nitrate, dimethyl sulphoxide, and thiosulphate as potential substrates, which might contribute to energizing the cell membrane. *A. fulgidus* contains a large number of flavoproteins, iron–sulphur proteins and iron-binding proteins that contribute to the general intracellular flow of electrons (Fig. 3). Detoxification enzymes include a peroxidase/catalase, an alkyl-hydroperoxide reductase, arsenate reductase, and eight NADH oxidases, presumably catalysing the

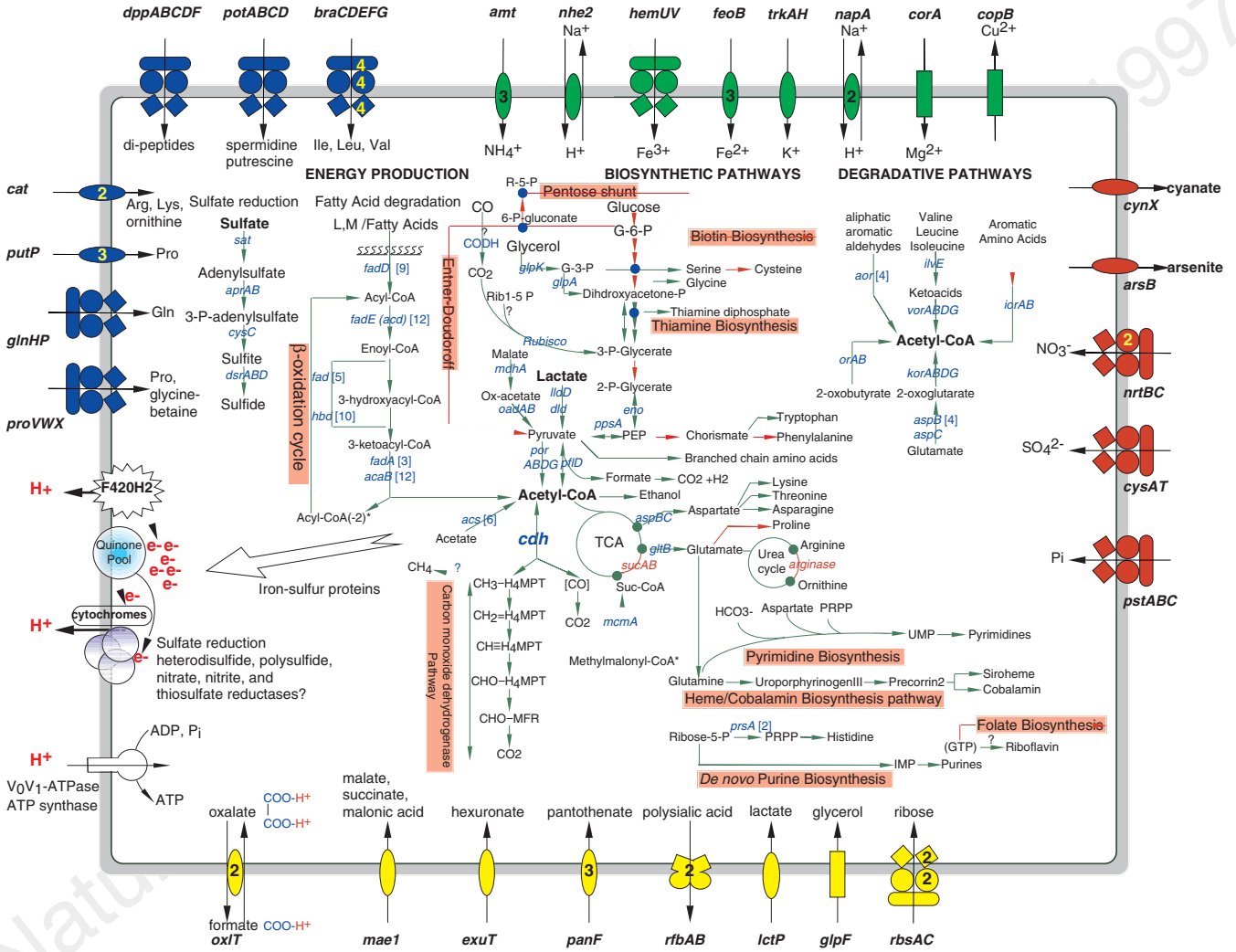

Figure 3 An integrated view of metabolism and solute transport in *A. fulgidus*. Biochemical pathways for energy production, biosynthesis of organic compounds, and degradation of amino acids, aldehydes and acids are shown with the central components of *A. fulgidus* metabolism, sulphate, lactate and acetyl-CoA highlighted. Pathways or steps for which no enzymes were identified are represented by a red arrow. A question mark is attached to pathways that could not be completely elucidated. Macromolecular biosynthesis of RNA, DNA and ether lipids have been omitted. Membrane-associated reactions that establish the proton-motive force (PMF) and generate ATP (electron transport chain and V_1V_0-ATPase) are linked to cytosolic pathways for energy production. The oxalate-formate antiporters (*oxlT*) may also contribute to the PMF by mediating electrogenic anion exchange. Each gene product with a predicted function in ion or solute transport is illustrated. Proteins are grouped by substrate specificity with transporters for cations (green), anions (red), carbohydrates/organic alcohols/acids (yellow), and amino acids/peptides/amines (blue) depicted. Ion-coupled permeases are represented by ovals (*mae1*, *exuT*, *panF*, *lctP*, *arsB*, *cynX*, *napA*/*nhe2*, *amt*, *feoB*, *trkAH*, *cat* and *putP* encode transporters for malate, hexuronate, pantothenate, lactate, arsenite, cyanate, sodium, ammonium, iron (II), potassium, arginine/lysine and proline, respectively). ATP-binding cassette (ABC) transport systems are shown as composite figures of ovals, diamonds and circles (*proVWX*, *glnHPQ*, *dppABCDF*, *potABCD*, *braCDEFG*, *hemUV*, *nrtBC*, *cysAT*, *pstABC*, *rbsAC*, *rfbAB* correspond to gene products for proline, glutamine, dipeptide, spermidine/putrescine, branch-chain amino acids, iron (III), nitrate, sulphate, phosphate, ribose and polysialic acid transport, respectively). All other porters drawn as rectangles (*glpF*, glycerol uptake facilitator; *copB*, copper transporting ATPase; *corA*, magnesium and cobalt transporter). Export and import of solutes is designated by arrows. The number of paralogous genes encoding each protein is indicated in brackets for cytoplasmic enzymes, or within the figure for transporters. Abbreviations: *acs*, acetyl-CoA synthetase; *aor*, aldehyde ferredoxin oxidoreductase; *aprAB*, adenylylsulphate reductase; *aspBC*, aspartate aminotransferase; *cdh*, acetyl-CoA decarbonylase/synthase complex; *cysC*, adenylylsulphate 3-phosphotransferase; *dld*, D-lactate dehydrogenase; *dsrABD*, sulphite reductase; *eno*, enolase; *fadA/acaB*, 3-ketoacyl-CoA thiolase; *fadD*, long-chain-fatty-acid-CoA ligase; *fad*, enoyl-CoA hydratase; *fadE* (*acd*), acyl-CoA dehydrogenase; *glpA*, glycerol-3-phosphate dehydrogenase; *glpK*, glycerol kinase; *gltB*, glutamate synthase; *hbd*, 3-hydroxyacyl-CoA dehydrogenase; *ilvE*, branched-chain amino-acid aminotransferase; *iorAB*, indolepyruvate ferredoxin oxidoreductase; *korABDG*, 2-ketoglutarate ferredoxin oxidoreductase; *lldD*, L-lactate dehydrogenase; *mcmA*, methylmalonyl-CoA mutase; *mdhA*, L-malate dehydrogenase; *oadAB*, oxaloacetate decarboxylase; *orAB*, 2-oxoacid ferredoxin oxidoreductase; *pflD*, pyruvate formate lysase 2; *porABDG*, pyruvate ferredoxin oxidoreductase; *ppsA*, phosphoenolpyruvate synthase; *prsA*, ribose-phosphate pyrophosphokinase; *sucAB*, 2-ketoglutarate dehydrogenase; *sat*, sulphate adenylyltransferase; TCA, tricarboxylic acid cycle; *vorABDG*, 2-ketoisovalerate ferredoxin oxidoreductase.

four-electron reduction of molecular oxygen to water, with the concurrent regeneration of NAD.

Transporters

A. fulgidus may synthesize several transporters for the import of carbon-containing compounds, probably contributing to its ability to switch from autotrophic to heterotrophic growth[5]. Both *M. jannaschii* and *A. fulgidus* have branched-chain amino-acid ABC transport systems and a transporter for the uptake of arginine and lysine. *A. fulgidus* encodes proteins for dipeptide, spermidine/putrescine, proline/glycine-betaine and glutamine uptake, as well as transporters for sugars and acids, rather like the membrane systems described in eubacterial heterotrophs. These compounds provide the necessary substrates for numerous biosynthetic and degradative pathways (Fig. 3).

Many *A. fulgidus* redox proteins are predicted to require iron. Correspondingly, iron transporters have been identified for the import of both oxidized (Fe^{3+}) and reduced (Fe^{2+}) forms of iron. There are duplications in functional and regulatory genes in both systems. The uptake of Fe^{3+} may depend on haemin or a haemin-like compound because *A. fulgidus* has orthologues to the eubacterial *hem* transport system proteins, HemU and HemV. *A. fulgidus* may also use the regulatory protein Fur to modulate Fe^{3+} transport; this protein is not present in *M. jannaschii*. Fe^{2+} uptake occurs through a modified Feo system containing FeoB. This is the third example of an isolated *feoB* gene: *M. jannaschii* and *Helicobacter pylori* also appear to lack *feoA*, implying that FeoA is not essential for iron transport in these organisms.

A complex suite of proteins regulates ionic homeostasis. Ten distinct transporters facilitate the flux of the physiological ions K^+, Na^+, NH_4^+, Mg^{2+}, Fe^{2+}, Fe^{3+}, NO_3^-, SO_4^{2-} and inorganic phosphate (P_i). Most of these transporters have homologues in *M. jannaschii* and are therefore likely to be critical for nutrient acquisition during autotrophic growth. *A. fulgidus* has additional ion transporters for the elimination of toxic compounds including copper, cyanate and arsenite. As in *M. jannaschii*, the *A. fulgidus* genome contains two paralogous operons of cobalamin biosynthesis-cobalt transporters, *cbiMQO*.

Sensory functions and regulation of gene expression

Consistent with its extensive energy-producing metabolism and versatile system for carbon utilization, *A. fulgidus* has complex sensory and regulatory networks. These networks contain over 55 proteins with presumed regulatory functions, including members of the ArsR, AsnC and Sir2 families, as well as several iron-dependent repressor proteins. There are at least 15 signal-transducing histidine kinases, but only nine response regulators; this difference suggests there is a high degree of cross-talk between kinases and regulators. Only four response regulators appear to be in operons with histidine kinases, including those in the methyl-directed chemotaxis system (Che), which lies adjacent to the flagellar biosynthesis operon. Although rich in regulatory proteins, *A. fulgidus* apparently lacks regulators for response to amino-acid and carbon starvation as well as to DNA damage. Finally, *A. fulgidus* contains a homologue of the mammalian mitochondrial benzodiazepine receptor, which functions as a sensor in signal-transduction pathways[25]. These receptors have been previously identified only in Proteobacteria and Cyanobacteria[25].

Replication, repair and cell division

A. fulgidus possesses two family B DNA polymerases, both related to the catalytic subunit of the eukaryal delta polymerase, as previously observed in the Sulfolobales[26]. It also has a homologue of the proofreading ε subunit of *E. coli* Pol III, not previously observed in the Archaea. The DNA repair system is more extensive than that found in *M. jannaschii*, including a homologue of the eukaryal Rad25, a 3-methyladenine DNA glycosylase, and exodeoxynuclease III. As well as reverse gyrase, topoisomerase I (ref. 9), and topoisomerase VI (ref. 27), the genes for the first archaeal DNA gyrase were identified.

A. fulgidus lacks a recognizable type II restriction-modification system, but contains one type I system. In contrast, two type II and three type I systems were identified in *M. jannaschii*. No homologue of the *M. jannaschii* thermonuclease was identified.

The cell-division machinery is similar to that of *M. jannaschii*, with orthologues of eubacterial *fts* and eukaryal *cdc* genes. However, several *cdc* genes found in *M. jannaschii*, including homologues of *cdc23*, *cdc27*, *cdc47* and *cdc54*, appear to be absent in *A. fulgidus*.

Transcription and translation

A. fulgidus and *M. jannaschii* have transcriptional and translational systems distinct from their eubacterial and eukaryal counterparts. In both, the RNA polymerase contains the large universal subunits and five smaller subunits found in both Archaea and eukaryotes. Transcription initiation is a simplified version of the eukaryotic mechanism[28,29]. However, *A. fulgidus* alone has a homologue of eukaryotic TBP-interacting protein 49 not seen in *M. jannaschii*, but apparently present in *Sulfolobus solfactaricus*.

Translation in *A. fulgidus* parallels *M. jannaschii* with a few exceptions. The organism has only one rRNA operon with an Ala-tRNA gene in the spacer and lacks a contiguous 5S rRNA gene. Genes for 46 tRNAs were identified, five of which contain introns in the anticodon region that are presumably removed by the intron excision enzyme EndA. The gene for selenocysteine tRNA (SelC) was not found, nor were the genes for SelA, SelB and SelD. With the exception of Asp-$tRNA^{GTC}$ and Val-$tRNA^{CAC}$, tRNA genes are not linked in the *A. fulgidus* genome. The RNA component of the tRNA maturation enzyme RNase P is present. Both *A. fulgidus* and *M. jannaschii* appear to possess an enzyme that inserts the tRNA-modified nucleoside archaeosine, but only *A. fulgidus* has the related enzyme that inserts the modified base queuine.

Both *A. fulgidus* and *M. jannaschii* lack glutamine synthetase and asparagine synthetase; the relevant tRNAs are presumably amino-acylated with glutamic and aspartic acids, respectively. An enzymatic *in situ* transamidation then converts the amino acid to its amide form, as seen in other Archaea and in Gram-positive Eubacteria[30]. Indeed, genes for the three subunits of the Glu-tRNA amidotransferase (*gatABC*) have been identified in *A. fulgidus*. The Lys aminoacyl-tRNA synthetase in both organisms is a class I-type, not a class II-type[31]. *A. fulgidus* possesses a normal tRNA synthetase for both Cys and Ser, unlike *M. jannaschii* in which the former was not identifiable and the latter was unusual[9].

M. jannaschii has a single gene belonging to the TCP-1 chaperonin family, whereas *A. fulgidus* has two that encode subunits α and β of the thermosome. Phylogenetic analysis of the archaeal TCP-1 family indicates that these *A. fulgidus* genes arose by a recent species-specific gene duplication, as is the case for the two subunits of the *Thermoplasma acidophilum* thermosome[32] and the *Sulfolobus shibatae* rosettasome[33]. As in *M. jannaschii*, no *dnaK* gene was identified.

Biosynthesis of essential components

Like most autotrophic microorganisms, *A. fulgidus* is able to synthesize many essential compounds, including amino acids, cofactors, carriers, purines and pyrimidines. Many of these biosynthetic pathways show a high degree of conservation between *A. fulgidus* and *M. jannaschii*. These two Archaea are similar in their biosynthetic pathways for siroheme, cobalamin, molybdopterin, riboflavin, thiamin and nictotinate, the role category with greatest conservation between these two organisms being amino-acid biosynthesis. Of 78 *A. fulgidus* genes assigned to amino-acid biosynthetic pathways, at least 73 (94%) have homologues in *M. jannaschii*. For both archaeal species, amino-acid biosynthetic pathways resemble those of *Bacillus subtilis* more closely than

those of *E. coli*. For example, in *A. fulgidus* and *M. jannaschii*, tryptophan biosynthesis is accomplished by seven enzymes, TrpA, B, C, D, E, F, G as in *B. subtilis*, rather than by five enzymes, TrpA, B, C, D, E (including the bifunctional TrpC and TrpD) as found in *E. coli*.

No biotin biosynthetic genes were identified, yet biotin can be detected in *A. fulgidus* cell extracts[34], and several genes encode a biotin-binding consensus sequence. Similarly, *A. fulgidus* lacks the genes for pyridoxine biosynthesis although pyridoxine can be found in cell extracts (albeit at lower levels than seen in *E. coli* and several Archaea[34]). No gene encoding ferrochelatase, the terminal enzyme in haem biosynthesis, has been identified, although *A. fulgidus* is known to use cytochromes[34]. These cofactors may be obtained by mechanisms that we have not recognized. Although all of the enzymes required for pyrimidine biosynthesis appear to be present, three enzymes in the purine pathway (GAR transformylase, AICAR formyltransferase and the ATPase subunit of AIR carboxylase) have not been identified, presumably because they exist as new isoforms.

The Archaea share a unique cell membrane composed of ether lipids containing a glycerophosphate backbone with a 2,3-*sn* stereochemistry[35] for which there are multiple biosynthetic pathways[36]. In the case of *Halobacterium cutirubrum*, the backbone is apparently obtained by enantiomeric inversion of *sn*-glycerol-3-phosphate; in *Sulfolobus acidocaldarius* and *Methanobacterium thermoautotrophicum*, *sn*-glycerol-1-phosphate dehydrogenase builds the backbone from dihydroxyacetonephosphate. An orthologue of *sn*-glycerol-1-phosphate dehydrogenase has been identified in *A. fulgidus*, suggesting that the latter pathway is present.

Conclusions

Although *A. fulgidus* has been studied since its discovery ten years ago[1], the completed genome sequence provides a wealth of new information about how this unusual organism exploits its environment. For example, its ability to reduce sulphur oxides has been well characterized, but genome sequence data demonstrate that *A. fulgidus* has a great diversity of electron transport systems, some of unknown specificity. Similarly, *A. fulgidus* has been characterized as a scavenger with numerous potential carbon sources, and its gene complement reveals the extent of this capability. *A. fulgidus* appears to obtain carbon from fatty acids through β-oxidation, from degradation of amino acids, aldehydes and organic acids, and perhaps from CO.

A. fulgidus has extensive gene duplication in comparison with other fully sequenced prokaryotes. For example, in the fatty acid and phospholipid metabolism category, there are 10 copies of 3-hydroxyacyl-CoA dehydrogenase, 12 copies of 3-ketoacyl-CoA thiolase, and 12 of acyl-CoA dehydrogenase. The duplicated proteins are not identical, and their presence suggests considerable metabolic differentiation, particularly with respect to the pathways for decomposing and recycling carbon by scavenging fatty acids. Other categories show similar, albeit less dramatic, gene redundancy. For example, there are six copies of acetyl-CoA synthetase and four aldehyde ferredoxin oxidoreductases for fermentation, as well as four copies of aspartate aminotransferase for amino-acid biosynthesis. These observations, together with the large number of paralogous gene families, suggest that gene duplication has been an important evolutionary mechanism for increasing physiological diversity in the Archaeoglobales.

A comparison of two archaeal genomes is inadequate to assess the diversity of the entire domain. Given this caveat, it is nevertheless possible to draw some preliminary conclusions from the comparison of *M. jannaschii* and *A. fulgidus*. A comparison of the gene content of these Archaea reveals that gene conservation varies significantly between role categories, with genes involved in transcription, translation and replication highly conserved; approximately 80% of the *A. fulgidus* genes in these categories have homologues in *M. jannaschii*. Biosynthetic pathways are also highly conserved, with approximately 80% of the *A. fulgidus* biosynthetic genes having homologues in *M. jannaschii*. In contrast, only 35% of the *A. fulgidus* central intermediary metabolism genes have homologues, reflecting their minimal metabolic overlap.

Over half of the *A. fulgidus* ORFs (1,290) have no assigned biological role. Of these, 639 have no database match. The remaining 651, designated 'conserved hypothetical proteins', have sequence similarity to hypothetical proteins in other organisms, two-thirds with apparent homologues in *M. jannaschii*. These shared hypothetical proteins will probably add to our understanding of the genetic repertoire of the Archaea. Analysis of the *A. fulgidus* and other archaeal and eubacterial genomes will provide the information necessary to begin to define a core set of archaeal genes, as well as to better understand prokaryotic diversity. □

Methods

Whole-genome random sequencing procedure. The type strain, *A. fulgidus* VC-16, was grown from a culture derived from a single cell isolated by optical tweezers[37] and provided by K. O. Stetter (University of Regensburg). Cloning, sequencing and assembly were essentially as described previously for genomes sequenced by TIGR[9,38–40]. One small-insert and one medium-insert plasmid library were generated by random mechanical shearing of genomic DNA. One large-insert lambda (λ) library was generated by partial *Tsp*509I digestion and ligation to λ-DASHII/*Eco*RI vector (Stratagene). In the initial random sequencing phase, 6.7-fold sequence coverage was achieved with 27,150 sequences from plasmid clones (average read length 500 bases) and 1,850 sequences from λ-clones. Both plasmid and λ-sequences were jointly assembled using TIGR assembler[41], resulting in 152 contigs separated by sequence gaps and five groups of contigs separated by physical gaps. Sequences from both ends of 560 λ-clones served as a genome scaffold, verifying the orientation, order and integrity and the contigs. Sequence gaps were closed by editing the ends of sequence traces and/or primer walking on plasmid or λ-clones clones spanning the respective gap. Physical gaps were closed by combinatorial polymerase chain reaction (PCR) followed by sequencing of the PCR product. At the end of gap closure, 90 regions representing 0.33% of the genome had only single-sequence coverage. These regions were confirmed with terminator reactions to ensure a minimum of 2-fold sequence coverage for the whole genome. The final genome sequence is based on 29,642 sequences, with a 6.8-fold sequence coverage. The linkage between the terminal sequences of 2,101 clones from the small-insert plasmid library (average size 1,419 bp) and 8,726 clones from the medium-insert plasmid library (average size 2,954 bp) supported the genome scaffold formed by the λ-clones (average size 16,381 bp), with 96.9% of the genome covered by λ-clones. The reported sequence differs in 20 positions from the 14,389 bp of DNA in a total of 11 previously published *A. fulgidus* genes.

ORF prediction and gene family identification. Coding regions (ORFs) were identified using a combination strategy based on two programs. Initial sets of ORFs were derived with GeneSmith (H.O.S., unpublished), a program that evaluates ORF length, separation and overlap between ORFs, and with CRITICA (J.H.B. & G.J.O., unpublished), a coding region identification tool using comparative analysis. The two largely overlapping sets of ORFs were merged into one joint set containing all members of both initial sets. ORFs were searched against a non-redundant protein database using BLASTX[10] and those shorter than 30 codons 'coding' for proteins without a database match were eliminated. Frameshifts were detected and corrected where appropriate as described previously[40]. Remaining frameshifts are considered authentic and corresponding regions were annotated as 'authentic frameshift'. In total, 527 hidden Markov models, based upon conserved protein families (PFAM version 2.0), were searched with HMMER to determine ORF membership in families and superfamilies[42]. Families of paralogous genes were constructed as described previously[40]. TopPred[43] was used to identify membrane-spanning domains in proteins.

Received 9 September; accepted 4 November 1997.

1. Stetter, K. O., Lauerer, G., Thomm, M. & Neuner, A. Isolation of extremely thermophilic sulfate reducers: Evidence for a novel branch of archaebacteria. *Science* **236,** 822–824 (1987).
2. Stetter, K. O., in *The Prokaryotes* (eds Balows, A., Trüper, H. G., Dworkin, M., Harder, W. & Schleifer, K. H.) 707–711 (Springer, Berlin, 1992).
3. Stetter, K. O. Microbial life in hyperthermal environments: Microorganisms from exotic environments continue to provide surprises about life's extremities. *ASM News* **61,** 285–290 (1995).

4. Stetter, K. O. *Archaeoglobus fulgidus* gen. nov., sp. nov.: a new taxon of extremely thermophilic archaebacteria. *Syst. Appl. Microbiol.* **10,** 172–173 (1988).
5. Stetter, K. O. *et al*. Hyperthermophilic archaea are thriving in deep North Sea and Alaskan oil reservoirs. *Nature* **365,** 743–745 (1993).
6. Vorholt, J., Kunow, J., Stetter, K. O. & Thauer, R. K. Enzymes and coenzymes of the carbon monoxide dehydrogenase pathway for autotrophic CO_2 fixation in *Archaeoglobus lithotrophicus* and the lack of carbon monoxide dehydrogenase in the heterotrophic *A. profundus*. *Arch. Microbiol.* **163,** 112–118 (1995).
7. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl Acad. Sci. USA* **74,** 5088–5090 (1977).
8. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl Acad. Sci. USA* **87,** 4576–4579 (1990).
9. Bult, C. J. *et al*. Complete genome sequence of the methanogenic archaeon *Methanococcus jannaschii*. *Science* **273,** 1058–1073 (1996).
10. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
11. Riley, M. Functions of gene products of *Escherichia coli*. *Microbiol. Rev.* **57,** 862–952 (1993).
12. Cooling, F, B. III, Maloney, C. L., Nagel, E., Tabinowski, J. & Odom, J. M. Inhibition of sulfate respiration by 1,8-dehydroxyanthraquinone and other anthraquinone derivatives. *Appl. Environ. Microbiol.* **62,** 2999–3004 (1996).
13. Thauer, R. K. & Kunow, J. in *Sulfate Reducing Bacteria* (ed. Barton, L. L.) 33–48 (Plenum, New York, 1995).
14. Speich, D. *et al*. Adenylylsulfate reductase from the sulfate-reducing archaeon *Archaeoglobus fulgidus*: cloning and characterization of the genes and comparison of the enzyme with other iron-sulfur flavoproteins. *Microbiology* **140,** 1273–1284 (1994).
15. Clark, D. P. & Cronan, J. E. Jr in Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular biology* (ed Neidhardt, F. C.) 343–357 (ASM Press, Washington DC, 1996).
16. Möller-zirkhan, D. & Thauer, R. K. Anaerobic lactate oxidation to 3 CO_2 by *Archaeoglobus fulgidus via* the carbon monoxide dehydrogenase pathway: demonstration of the acetyl-CoA carbon-carbon cleavage reaction in cell extracts. *Arch. Microbiol.* **153,** 215–218 (1990).
17. Schauder, R., Eikmanns, B., Thauer, R. K., Widdel, F. & Fuchs, G. Acetate oxidation to CO_2 in anaerobic-bacteria *via* a novel pathway not involving reactions of the citric-acid cycle. *Arch. Microbiol.* **145,** 162–172 (1986).
18. Dai, Y.-R. *et al*. Acetyl-CoA decarbonylase/synthase complex from *Archaeoglobus fulgidus*: purification, characterization, and properties. *Arch. Microbiol.* (submitted).
19. Gorris, L. G. M., Voet, A. C. W. A. & van der Drift, C. Structural characteristics of methanogenic cofactors in the non-methanogenic archaebacterium *Archaeoglobus fulgidus*. *BioFactors* **3,** 29–35 (1991).
20. Zhang, Q., Iwasaki, T., Wakagi, T. & Oshima, T. 2-oxoacid:ferredoxin oxidoreductase from the thermoacidophilic archaeon, *Sulfolobus* sp. strain 7. *J. Biochem.* **120,** 587–599 (1996).
21. Tersteegen, A., Linder, D., Thauer, R. K. & Hedderich, R. Structures and functions of four anabolic 2-oxoacid oxidoreductases in *Methanobacterium thermoautotrophicum*. *Eur. J. Biochem.* **244,** 862–868 (1997).
22. Kletzin, A. & Adams, M. W. W. Molecular and phylogenetic characterization of pyruvate and 2-ketoisovalerate ferredoxin oxidoreductases from *Pyrococcus furiosus* and pyruvate ferredoxin oxidoreductase from *Thermotoga maritima*. *J. Bacteriol.* **178,** 248–257 (1996).
23. LaPaglia, C. & Hartzell, P. L. Stress-induced production of biofilm in the hyperthermophile *Archaeoglobus fulgidus*. *Appl. Environ. Microbiol.* **63,** 3158–3163 (1997).
24. Kunow, J., Linder, D., Stetter, K. O. & Thauer, R. K. $F_{420}H_2$: quinone oxidoreductase from *Archaeoglobus fulgidus*—characterization of a membrane-bound mutlisubunit complex containing FAD and iron–sulfur clusters. *Eur. J. Biochem.* **223,** 503–511 (1994).
25. Yeliseev, A. A., Krueger, K. E. & Kaplan, S. A mammalian mitochondrial drug receptor functions as a bacterial "oxygen" sensor. *Proc. Natl Acad. Sci. USA* **94,** 5101–5106 (1997).
26. Edgell, D. R., Klenk, H.-P. & Doolittle, W. F. Gene duplications in evolution of archaeal family B DNA polymerases. *J. Bacteriol.* **179,** 2632–2640 (1997).
27. Bergerat, A. *et al*. An atypical topoisomerase II from archaea with implications for meiotic recombination. *Nature* **386,** 414–417 (1997).
28. Marsh, T. L., Reich, C. I., Whitelock, R. B. & Olsen, G. J. Transcription factor IID in the Archaea: sequences in the *Thermococcus celer* genome would encode a product closely related to the TATA-binding protein of eukaryotes. *Proc. Natl Acad. Sci. USA* **91,** 4180–4184 (1994).
29. Kosa, P. F., Ghosh, G., DeDecker, B. S. & Sigler, P. B. The 2.1-A crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box. *Proc. Natl Acad. USA* **94,** 6042–6047 (1997).
30. Curnow, A. W. *et al*. Glu-tRNA^{Gln} amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc. Natl Acad. Sci. USA* **94,** 11819–11826 (1997).
31. Ibba, M., Bobo, J. L., Rosa, P. A. & Soll, D. Archaeal-type lysyl-tRNA synthetase in the Lyme disease spirochete *Borrelia burgdorferi*. *Proc. Natl Acad. Sci. USA* (submitted).
32. Waldmann, T., Lupas, A., Kellermann, J., Peters, J. & Baumeister, W. Primary structure of the thermosome from *Thermoplasma acidophilum*. *Hoppe-Seyler's Biol. Chem.* **376,** 119–126 (1995).
33. Kagawa, H. K. *et al*. The 60 kDa heat shock proteins in the hyperthermophilic archaeon *Sulfolobus shibatae*. *J. Mol. Biol.* **253,** 712–725 (1995).
34. Noll, K. M. & Barber, T. S. Vitamin contents of archaebacteria. *J. Bacteriol.* **170,** 4315–4321 (1988).
35. Thornebene, T. G. & Langworthy, T. A. Diphytanyl and dibiphytanyl glycerol ether lipids of methanogenic archaebacteria. *Science* **203,** 51–53 (1979).
36. Nishihara, M. & Koga, Y. sn-glycerol-1-phosphate dehydrogenase in *Methanobacterium thermoautotrophicum*: key enzyme in biosynthesis of the enantiomeric glycerophosphate backbone of ether phospholipids of archaebacteria. *J. Biochem.* **117,** 933–935 (1995).
37. Huber, R. *et al*. Isolation of a hyperthermophilic archaeum predicted by *in situ* RNA analysis. *Nature* **376,** 57–58 (1995).
38. Fleischmann, R. D. *et al*. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269,** 496–511 (1995).
39. Fraser, C. M. *et al*. The minimal gene complement of *Mycoplasma genitalium*. *Science* **270,** 397–403 (1995).
40. Tomb, J.-F. *et al*. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388,** 539–547 (1997).
41. Sutton, G. G., White, O., Adams, M. D. & Kerlavage, A. R. TIGR Assembler: A new tool for assembling large shotgun sequencing projects. *Genome Sequence Technol.* **1,** 9–19 (1995).
42. Sonnhammer, E. L., Eddy, S. R. & Durbin, R. Pfam: A comprehensive database of protein families based on seed alignments. *Proteins* **28,** 405–420 (1997).
43. Claros, M. G. & von Heijne, G. TopPred II: an improved software for membrane protein structure predictions. *Comput. Appl. Biosci.* **10,** 685–686 (1994).

Acknowledgements. We thank M. Heaney, J. Scott and R. Shirley for software and database support; V. Sapiro, B. Vincent, J. Meehan and D. Maas for computer system support; B. Cameron and D. J. Doyle for editorial assistance: and K. O. Stetter for providing *A. fulgidus* VC-16. This work was supported by the US Department of Energy.

Correspondence and requests for materials should be addressed to J.C.V. (e-mail: gaf@tigr.org). The annotated genome sequence and the gene family alignments are available on the World-Wide Web at http://www.tigr.org/tdb/mdb/afdb/afdb.html. The sequence has been deposited in GenBank with accession number AE000782.

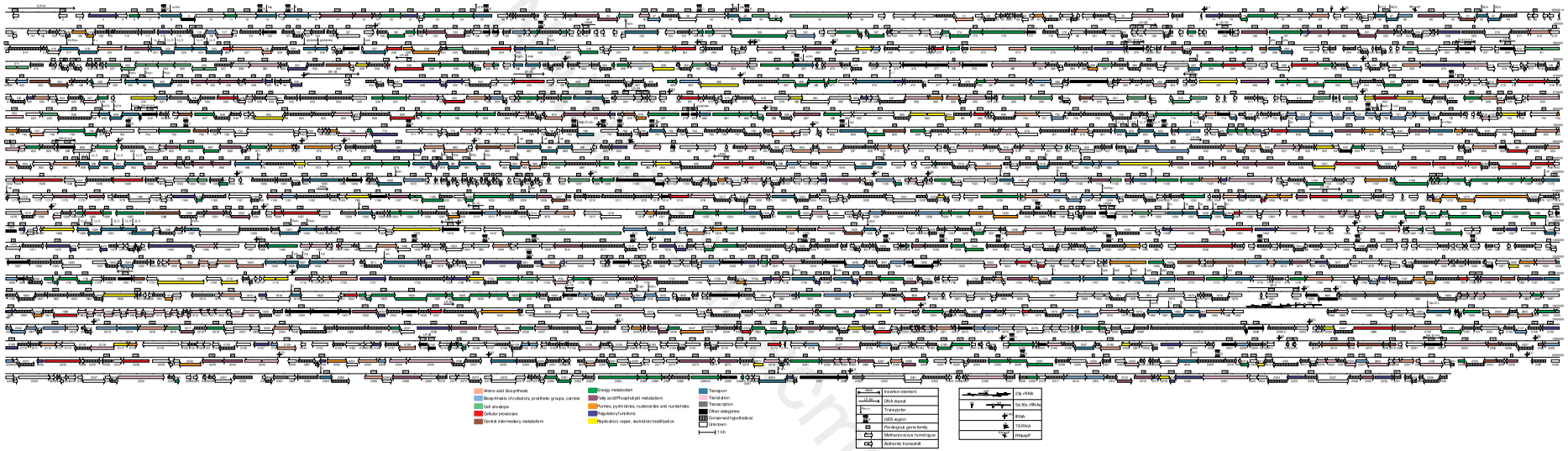

Table 2 . List of *A. fulgidus* genes with putative identification. Gene numbers correspond to those in Fig. 2. Percentages represent per cent identities.

AMINO ACID BIOSYNTHESIS		
General		
AF0906	hydantoin utilization protein A (hyuA)	27.4%
Aromatic amino acid family		
AF0228	3-dehydroquinate dehydratase (aroD)	36.8%
AF1497	5-enolpyruvylshikimate 3-phosphate synthase (aroA)	41.5%
AF1603	anthranilate synthase component I (trpE)	43.7%
AF1604	anthranilate synthase component II (trpD)	43.8%
AF1602	anthranilate synthase component II (trpG)	50.0%
AF0227	chorismate mutase/prephenate dehydratase (pheA)	32.2%
AF0670	chorismate synthase (aroC)	55.3%
AF1601	phosphoribosyl anthranilate isomerase (trpF)	37.1%
AF2327	shikimate 5-dehydrogenase (aroE)	43.1%
AF0343	tryptophan repressor binding protein (wrbA)	46.6%
AF1599	tryptophan synthase, subunit alpha (trpA)	39.5%
AF1240	tryptophan synthase, subunit beta (trpB-1)	39.4%
AF1600	tryptophan synthase, subunit beta (trpB-2)	64.1%
Aspartate family		
AF2112	5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase (metE)	28.1%
AF0882	asparaginase (asnA)	45.9%
AF1439	asparagine synthetase (asnB)	36.9%
AF2366	aspartate aminotransferase (aspB-1)	42.3%
AF2129	aspartate aminotransferase (aspB-2)	45.4%
AF1623	aspartate aminotransferase (aspB-3)	39.4%
AF0409	aspartate aminotransferase (aspB-4)	45.2%
AF1417	aspartate aminotransferase (aspC)	46.2%
AF0700	aspartate kinase (lysC)	49.1%
AF1422	aspartate racemase	48.0%
AF1506	aspartate-semialdehyde dehydrogenase (asd)	60.9%
AF0800	diaminopimelate decarboxylase (lysA)	45.6%
AF0747	diaminopimelate epimerase (dapF)	45.8%
AF0909	dihydrodipicolinate reductase (dapB)	48.6%
AF0910	dihydrodipicolinate synthase (dapA)	51.0%
AF0935	homoserine dehydrogenase (hom)	47.9%
AF0886	S-adenosylhomocysteinase hydrolase (ahcY-1)	31.7%
AF2000	S-adenosylhomocysteinase hydrolase (ahcY-2)	67.3%
AF0051	succinyl-diaminopimelate desuccinylase (dapE-1)	30.5%
AF0904	succinyl-diaminopimelate desuccinylase (dapE-2)	43.8%
AF0551	threonine synthase (thrC-1)	40.5%
AF1316	threonine synthase (thrC-2)	61.0%
Glutamate family		
AF1280	acetylglutamate kinase (argB)	56.1%
AF2288	acetylglutamate kinase, putative	29.0%
AF0080	acetylornithine aminotransferase (argD-1)	48.3%
AF1815	acetylornithine aminotransferase (argD-2)	36.2%
AF0522	acetylornithine deacetylase (argE)	29.4%
AF0883	argininosuccinate lyase (argH)	42.2%
AF2252	argininosuccinate synthetase (argG)	62.0%
AF1147	glutamate N-acetyltransferase (argJ)	47.8%
AF0953	glutamate synthase (gltB)	57.9%
AF0949	glutamine synthetase (glnA)	43.3%
AF2071	N-acetyl-gamma-glutamyl-phosphate reductase (argC)	53.3%
AF1255	ornithine carbamoyltransferase (argF)	51.7%
Pyruvate family		
AF0957	2-isopropylmalate synthase (leuA-1)	53.5%
AF0219	2-isopropylmalate synthase (leuA-2)	53.9%
AF2199	3-isopropylmalate dehydratase, large subunit (leuC)	49.3%
AF0629	3-isopropylmalate dehydratase, small subunit (leuD-1)	56.4%
AF1761	3-isopropylmalate dehydratase, small subunit (leuD-2)	57.1%
AF0628	3-isopropylmalate dehydrogenase (leuB)	59.2%
AF1720	acetolactate synthase, large subunit (ilvB-1)	57.5%
AF1780	acetolactate synthase, large subunit (ilvB-2)	32.1%
AF2015	acetolactate synthase, large subunit (ilvB-3)	34.1%
AF2100	acetolactate synthase, large subunit (ilvB-4)	38.4%
AF1719	acetolactate synthase, small subunit (ilvN)	60.4%
AF1672	acetolactate synthase, small subunit, putative	29.7%
AF0933	branched-chain amino acid aminotransferase (ilvE)	59.0%
AF1014	dihydroxy-acid dehydratase (ilvD)	54.5%
AF1985	ketol-acid reductoisomerase (ilvC)	61.8%
Serine family		
AF0813	phosphoglycerate dehydrogenase (serA)	48.8%
AF2138	phosphoserine phosphatase (serB)	50.7%
AF0273	sarcosine oxidase, subunit alpha (soxA)	31.1%
AF0274	sarcosine oxidase, subunit beta (soxB)	26.5%
AF0852	serine hydroxymethyltransferase (glyA)	56.1%
Histidine family		
AF0590	ATP phosphoribosyltransferase (hisG)	31.6%
AF0212	histidinol dehydrogenase (hisD)	51.6%
AF2002	histidinol-phosphate aminotransferase (hisC-1)	39.8%
AF2024	histidinol-phosphate aminotransferase (hisC-2)	36.8%
AF0985	imidazoleglycerol-phosphate dehydrogenase/histidinol-phosphatase (hisB)	42.2%
AF0819	imidazoleglycerol-phosphate synthase, cyclase subunit (hisF)	67.0%
AF2265	imidazoleglycerol-phosphate synthase, subunit H (hisH)	44.4%
AF0509	imidazoleglycerol-phosphate synthase, subunit H, putative	43.2%
AF1950	phosphoribosyl-AMP cyclohydrolase/ phosphoribosyl-ATP pyrophosphohydrolase (hisIE)	59.6%
AF0713	phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (hisA-1)	37.5%
AF0986	phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (hisA-2)	42.2%
BIOSYNTHESIS OF COFACTORS, PROSTHETIC GROUPS, AND CARRIERS		
General		
AF1855	2,3-dihydroxybenzoate-AMP ligase (entE)	27.2%
AF1070	coenzyme F390 synthetase (ftsA-1)	30.3%
AF1671	coenzyme F390 synthetase (ftsA-2)	31.9%
AF2013	coenzyme F390 synthetase (ftsA-3)	30.4%
AF2151	isochorismatase (entB)	31.2%
Folic acid		
AF1414	dihydropteroate synthase	40.8%
Heme and porphyrin		
AF1648	bacteriochlorophyll synthase, 33 kDa subunit	27.9%
AF0464	bacteriochlorophyll synthase, 43 kDa subunit (chlP-1)	29.7%
AF1023	bacteriochlorophyll synthase, 43 kDa subunit (chlP-2)	31.2%
AF1637	bacteriochlorophyll synthase, 43 kDa subunit (chlP-3)	27.0%
AF0037	cobalamin (5'-phosphate) synthase (cobS-1)	33.9%
AF2323	cobalamin (5'-phosphate) synthase (cobS-2)	34.4%
AF0725	cobalamin biosynthesis precorrin methylase (cbiG)	30.7%
AF0727	cobalamin biosynthesis precorrin-2 methyltransferase (cbiL)	31.5%
AF0726	cobalamin biosynthesis precorrin-3 methylase (cbiF)	49.2%
AF0724	cobalamin biosynthesis precorrin-3 methylase (cbiH)	49.0%
AF0722	cobalamin biosynthesis precorrin-6Y methylase (cbiE)	32.4%
AF0732	cobalamin biosynthesis precorrin-8W decarboxylase (cbiT)	30.8%
AF1336	cobalamin biosynthesis protein (cbiB)	38.4%
AF0723	cobalamin biosynthesis protein (cbiD)	36.3%
AF0728	cobalamin biosynthesis protein (cbiM-1)	51.4%
AF1843	cobalamin biosynthesis protein (cbiM-2)	41.2%
AF0731	cobalt transport ATP-binding protein (cbiO-1)	47.2%
AF1841	cobalt transport ATP-binding protein (cbiO-2)	41.1%
AF0729	cobalt transport protein (cbiN)	56.0%
AF0730	cobalt transport protein (cbiQ-1)	32.6%
AF1842	cobalt transport protein (cbiQ-2)	30.3%
AF1338	cobyric acid synthase (cbiP)	44.5%
AF2229	cobyrinic acid a,c-diamide synthase (cbiA)	42.3%
AF1241	glutamate-1-semialdehyde aminotransferase (hemL)	54.3%
AF1975	glutamyl-tRNA reductase (hemA)	42.7%
AF1594	heme biosynthesis protein (nirH)	25.2%
AF1125	heme biosynthesis protein (nirJ-1)	38.7%
AF2009	heme biosynthesis protein (nirJ-2)	31.8%
AF1593	heme d1 biosynthesis protein (nirD)	29.4%
AF1311	oxygen-independent coproporphyrinogen III oxidase, putative	27.1%
AF1242	porphobilinogen deaminase (hemC)	46.3%
AF1974	porphobilinogen synthase (hemB)	60.4%
AF1784	protoporphyrinogen oxidase (hemK)	33.5%
AF0422	uroporphyrin-III C-methyltransferase (cysG-1)	41.7%
AF1243	uroporphyrin-III C-methyltransferase (cysG-2)	52.5%
AF0116	uroporphyrinogen III synthase (hemD)	27.4%
Menaquinone and ubiquinone		
AF2176	4-hydroxybenzoate octaprenyltransferase (ubiA)	41.6%
AF0404	4-hydroxybenzoate octaprenyltransferase, putative	30.6%
AF2413	coenzyme PQQ synthesis protein (pqqE)	30.5%
AF1191	dihydroxynaphthoic acid synthase (menB)	54.6%
AF1551	octaprenyl-diphosphate synthase (ispB)	33.2%
AF0140	ubiquinone/menaquinone biosynthesis methyltransferase (ubiE)	31.0%
Molybdopterin		
AF2006	molybdenum cofactor biosynthesis protein (moaA)	47.8%
AF0265	molybdenum cofactor biosynthesis protein (moaB)	44.4%
AF2150	molybdenum cofactor biosynthesis protein (moaC)	62.0%
AF0931	molybdenum cofactor biosynthesis protein (moeA-1)	50.8%
AF0930	molybdenum cofactor biosynthesis protein (moeA-2)	44.8%
AF0161	molybdenum cofactor biosynthesis protein (moeA-3)	30.5%
AF0531	molybdenum cofactor biosynthesis protein (moeB)	44.0%
AF1022	molybdenum-pterin-binding protein (mopB)	39.3%
AF1624	molybdopterin converting factor, subunit 1 (moaD)	36.6%
AF2179	molybdopterin converting factor, subunit 2 (moaE)	33.3%
AF2005	molybdopterin-guanine dinucleotide biosynthesis protein A (mobA)	33.2%
AF2253	molybdopterin-guanine dinucleotide biosynthesis protein B (mobB)	40.0%
Pantothenate		
AF1645	pantothenate metabolism flavoprotein (dfp)	42.4%
Riboflavin		
AF0484	GTP cyclohydrolase II (ribA-1)	44.5%
AF2107	GTP cyclohydrolase II (ribA-2)	47.1%
AF1416	riboflavin synthase (ribC)	53.3%
AF2128	riboflavin synthase, subunit beta (ribE)	75.9%
AF2007	riboflavin-specific deaminase (ribG)	43.7%
Thiamine		
AF2075	hydroxyethylthiazole kinase (thiM)	33.6%
AF2208	hydroxymethylpyrimidine phosphate kinase (thiD)	35.5%
AF1695	thiamine biosynthesis protein (apbA)	36.9%
AF2412	thiamine biosynthesis protein (thiC)	60.2%
AF0553	thiamine biosynthesis protein (thiF)	38.1%
AF0088	thiamine biosynthesis protein, putative	28.2%
AF0702	thiamine biosynthetic enzyme (thi1)	50.0%
AF0733	thiamine monophosphate kinase (thiL)	30.4%
AF2074	thiamine phosphate pyrophosphorylase (thiE)	45.5%
Pyridine nucleotides		
AF1000	NH(3)-dependent NAD+ synthetase (nadE)	52.0%
AF1839	nicotinate-nucleotide pyrophosphorylase (nadC)	43.2%
AF1837	quinolinate synthetase (nadA), authentic frameshift	53.9%
CELL ENVELOPE		
Membranes, lipoproteins, and porins		
AF1420	membrane protein	51.8%
AF1354	membrane protein, putative	32.8%
Surface polysaccharides, lipopolysaccharides and antigens		
AF0324	dTDP-glucose 4,6-dehydratase (rfbB)	50.0%
AF0043	first mannosyl transferase (wbaZ-1)	30.0%
AF0606	first mannosyl transferase (wbaZ-2)	29.0%
AF1728	galactosyltransferase	26.9%
AF0044	GDP-D-mannose dehydratase (gmd-1), authentic frameshift	40.7%
AF1142	glucose-1-phosphate cytidylyltransferase (rfbF)	38.6%
AF0242	glucose-1-phosphate thymidylyltransferase (graD-1)	27.7%
AF0325	glucose-1-phosphate thymidylyltransferase (graD-2)	45.2%
AF0321	glycosyl transferase	30.7%
AF0387	glycosyltransferase, putative	33.8%
AF0467	immunogenic protein (bcsp31-1)	34.7%
AF0635	immunogenic protein (bcsp31-2)	44.3%
AF0988	immunogenic protein (bcsp31-3)	28.3%
AF0602	LPS biosynthesis protein, putative	29.6%
AF0617	LPS biosynthesis protein, putative	29.0%
AF0607	LPS glycosyltransferase, putative	29.7%
AF0326	mannose-1-phosphate guanylyltransferase (rfbM), authentic frameshift	42.4%
AF1097	mannose-6-phosphate isomerase/mannose-1-phosphate guanylyl transferase (manC)	43.1%
AF0035	mannosephosphate isomerase, putative	31.3%
AF0045	mannosyltransferase A (mtfA)	38.7%
AF0311	O-antigen biosynthesis protein (rfbC), authentic frameshift	30.6%
AF0458	phosphomannomutase (pmm)	39.5%
AF0595	polysaccharide biosynthesis protein, putative	24.1%
AF0322	rhamnosyl transferase (rfbQ)	27.5%
AF0323	spore coat polysaccharide biosynthesis protein (spsK-2), authentic frameshift	36.3%
AF0620	succinoglycan biosynthesis protein (exoM)	24.8%
AF0361	UDP-glucose 4-epimerase (galE-1)	38.6%
AF2016	UDP-glucose 4-epimerase (galE-2)	30.0%
AF0302	UDP-glucose dehydrogenase (ugd-1)	43.8%
AF0596	UDP-glucose dehydrogenase (ugd-2)	44.1%
Surface structures		
AF1054	flagellin (flaB1-1)	30.0%
AF1055	flagellin (flaB1-2)	31.1%
AF0275	surface layer protein B (slgB-1)	30.8%
AF1413	surface layer protein B (slgB-2)	29.9%
CELLULAR PROCESSES		
General		
AF1040	chemotaxis histidine kinase (cheA)	41.9%
AF1035	chemotaxis histidine kinase, putative	25.3%
AF1036	chemotaxis histidine kinase, putative	30.4%
AF1037	chemotaxis protein methyltransferase (cheR)	33.2%
AF1042	chemotaxis response regulator (cheY)	62.9%
AF1034	methyl-accepting chemotaxis protein (tlpC-1)	27.5%
AF1045	methyl-accepting chemotaxis protein (tlpC-2)	29.6%
AF1041	protein-glutamate methylesterase (cheB)	43.3%
AF1032	purine NTPase, putative	32.2%
AF1044	purine-binding chemotaxis protein (cheW)	40.4%
Cell division		
AF0517	cell division control protein 21 (cdc21)	32.8%
AF1297	cell division control protein 48, AAA family (cdc48-1)	69.1%
AF2098	cell division control protein 48, AAA family (cdc48-2)	62.0%
AF0244	cell division control protein 6, putative	27.5%
AF1285	cell division control protein, AAA family, putative	49.3%
AF0696	cell division inhibitor (minD-1)	55.0%
AF1937	cell division inhibitor (minD-2)	32.8%
AF2051	cell division protein (ftsJ)	40.8%
AF0535	cell division protein (ftsZ-1)	60.4%
AF0570	cell division protein (ftsZ-2)	61.4%
AF0837	cell division protein pelota (pelA)	41.7%
AF1215	cell division protein, putative	32.8%
AF0238	centromere/microtubule-binding protein (cbf5)	58.8%
AF1558	chromosome segregation protein (smc1)	32.8%
AF1822	serine/threonine phosphatase (ppa)	31.9%
Chaperones		
AF1296	small heat shock protein (hsp20-1)	52.3%
AF1971	small heat shock protein (hsp20-2)	38.1%
AF2238	thermosome, subunit alpha (thsA)	70.6%
AF1451	thermosome, subunit beta (thsB)	68.2%
Chromosome-associated protein		
AF0337	archaeal histone A1 (hpyA1-1)	64.6%
AF1493	archaeal histone A1 (hpyA1-2)	69.7%
Detoxification		
AF2173	2-nitropropane dioxygenase (ncd2)	39.7%
AF0270	alkyl hydroperoxide reductase	73.5%
AF1361	arsenate reductase (arsC)	30.5%
AF0550	N-ethylammeline chlorohydrolase (trzA-1)	45.9%
AF0997	N-ethylammeline chlorohydrolase (trzA-2)	44.5%
AF0254	NADH oxidase (noxA-1)	35.1%
AF0395	NADH oxidase (noxA-2)	35.5%
AF0400	NADH oxidase (noxA-3)	40.8%
AF0951	NADH oxidase (noxA-4)	36.7%
AF1858	NADH oxidase (noxA-5)	34.0%
AF0455	NADH oxidase (noxB-1)	43.3%
AF1262	NADH oxidase (noxB-2)	42.9%
AF0226	NADH oxidase (noxC)	38.4%
AF0515	NADH oxidase, putative	25.5%
AF2233	peroxidase / catalase (perA)	62.9%
Protein and peptide secretion		
AF1902	protein translocase, subunit SEC61 alpha (secY)	50.0%
AF0536	protein translocase, subunit SEC61 gamma (secE)	25.0%
AF2062	signal recognition particle receptor (dpa)	54.8%
AF1258	signal recognition particle, subunit SRP19 (srp19)	36.6%
AF0622	signal recognition particle, subunit SRP54 (srp54)	51.2%
AF1791	signal sequence peptidase (sec11)	36.3%
AF1657	signal sequence peptidase (spc21)	47.0%
AF1655	signal sequence peptidase, putative	34.5%
AF0338	type II secretion system protein (gspE-1)	38.5%
AF0659	type II secretion system protein (gspE-2)	38.2%
AF0996	type II secretion system protein (gspE-3)	41.7%
AF1049	type II secretion system protein (gspE-4)	46.5%
CENTRAL INTERMEDIARY METABOLISM		
Degradation of polysaccharides		
AF1207	2-deoxy-D-gluconate 3-dehydrogenase (kduD)	45.3%
AF1795	endoglucanase (celM)	55.4%
Phosphorus compounds		
AF0756	exopolyphosphatase (ppx1)	55.1%
Polyamine biosynthesis		
AF0646	agmatinase (speB)	33.3%
AF2334	spermidine synthase (speE)	37.1%
Polysaccharides - (cytoplasmic)		
AF0599	dolichol phosphate mannose synthase, putative	32.1%
Sulfur metabolism		
AF0288	adenylylsulfate 3-phosphotransferase (cysC)	52.0%
AF1670	adenylylsulfate reductase, subunit A (aprA)	96.0%
AF1669	adenylylsulfate reductase, subunit B (aprB)	97.3%
AF1667	sulfate adenylyltransferase (sat)	28.4%
AF2228	sulfite reductase, desulfoviridin-type subunit gamma (dsvC)	41.3%
AF0423	sulfite reductase, subunit alpha (dsrA)	100.0%
AF0424	sulfite reductase, subunit beta (dsrB)	100.0%
AF0425	sulfite reductase, subunit gamma (dsrD)	97.4%
Other		
AF1706	2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoic acid hydrolase (pcbD)	29.4%
AF0675	2-hydroxy-6-oxohepta-2,4-dienoate hydrolase (todF)	26.3%
AF0091	2-hydroxyhepta-2,4-diene-1,7-dioate isomerase (hpcE-1)	44.5%
AF2225	2-hydroxyhepta-2,4-diene-1,7-dioate isomerase (hpcE-2)	66.0%
AF0333	4-hydroxyphenylacetate-3-hydroxylase (hpaA-1)	22.4%
AF0885	4-hydroxyphenylacetate-3-hydroxylase (hpaA-2)	26.0%
AF1027	4-hydroxyphenylacetate-3-hydroxylase (hpaA-3)	21.0%
AF0669	4-oxalocrotonate tautomerase, putative	31.9%
AF0808	glycolate oxidase subunit (glcD)	32.0%
AF2216	methylmalonyl-CoA decarboxylase, biotin carboxyl carrier subunit (mmdC)	36.2%
AF2217	methylmalonyl-CoA decarboxylase, subunit alpha (mmdA)	62.5%
AF1288	methylmalonyl-CoA mutase, subunit alpha (mutB), authentic frameshift	46.1%
AF2219	methylmalonyl-CoA mutase, subunit alpha, C-terminus (mcmA2)	48.7%
AF2215	methylmalonyl-CoA mutase, subunit alpha, N-terminus (mcmA1)	51.2%
AF2099	muconate cycloisomerase II (clcB)	24.9%
AF1425	phosphonopyruvate decarboxylase (bcpC-1)	35.0%
AF1751	phosphonopyruvate decarboxylase (bcpC-2)	48.6%
ENERGY METABOLISM		
Amino acids and amines		
AF1958	2-hydroxyglutaryl-CoA dehydratase, subunit alpha (hgdA)	30.5%

AF1957	2-hydroxyglutaryl-CoA dehydratase, subunit beta (hgdB)	24.4%
AF0130	acetylpolyamine aminohydrolase (aphA)	38.7%
AF2290	acetylpolyamine aminohydrolase, putative	33.3%
AF0991	glutaryl-CoA dehydrogenase (gcdH)	48.7%
AF1323	group II decarboxylase	28.0%
AF2004	group II decarboxylase	46.1%
AF2295	group II decarboxylase	30.5%
AF1665	ornithine cyclodeaminase (arcB)	35.3%
Anaerobic		
AF1145	4-hydroxybutyrate CoA transferase (cat2-1)	46.5%
AF1854	4-hydroxybutyrate CoA transferase (cat2-2)	47.5%
AF0866	glycerol kinase (glpK)	33.8%
AF1328	glycerol-3-phosphate dehydrogenase (glpA)	27.8%
AF0871	glycerol-3-phosphate dehydrogenase (NAD(P)+) (gpsA)	36.3%
AF0020	L-carnitine dehydratase (caiB-1)	33.3%
AF0990	L-carnitine dehydratase (caiB-2)	31.2%
ATP-proton motive force interconversion		
AF1158	ATP synthase, subunit E, putative	47.1%
AF1166	H+-transporting ATP synthase, subunit A (atpA)	67.0%
AF1167	H+-transporting ATP synthase, subunit B (atpB)	72.6%
AF1164	H+-transporting ATP synthase, subunit C (atpC)	37.5%
AF1168	H+-transporting ATP synthase, subunit D (atpD)	47.1%
AF1163	H+-transporting ATP synthase, subunit E (atpE)	36.3%
AF1165	H+-transporting ATP synthase, subunit F (atpF)	45.0%
AF1159	H+-transporting ATP synthase, subunit I (atpI)	30.1%
AF1160	H+-transporting ATP synthase, subunit K (atpK-1)	46.3%
AF1162	H+-transporting ATP synthase, subunit K (atpK-2)	46.3%
Electron transport		
AF2036	cytochrome C oxidase folding protein (coxD)	33.3%
AF0144	cytochrome C oxidase, subunit II (cbaB)	34.2%
AF0142	cytochrome C oxidase, subunit II, putative	38.0%
AF0190	cytochrome C oxidase, subunit II, putative	31.7%
AF1057	cytochrome C-type biogenesis protein (ccdA)	30.7%
AF2192	cytochrome C-type biogenesis protein (nrfE)	36.1%
AF2296	cytochrome oxidase, subunit I (cydA-1)	22.9%
AF2297	cytochrome oxidase, subunit I (cydA-2)	31.5%
AF2046	cytochrome oxidase, subunit I, putative	25.1%
AF0528	cytochrome-c3 hydrogenase, subunit gamma	39.3%
AF0833	desulfoferrodoxin (dfx)	63.0%
AF0344	desulfoferrodoxin, putative	47.3%
AF0287	electron transfer flavoprotein, subunit alpha (etfA)	39.7%
AF0286	electron transfer flavoprotein, subunit beta (etfB)	38.8%
AF1380	F420-nonreducing hydrogenase (vhtA)	34.8%
AF1371	F420-nonreducing hydrogenase (vhtD-1)	30.9%
AF1378	F420-nonreducing hydrogenase (vhtD-2)	33.1%
AF1381	F420-nonreducing hydrogenase (vhtG)	46.1%
AF1824	F420H2:quinone oxidoreductase, 11.2 kDa subunit, putative	24.1%
AF1823	F420H2:quinone oxidoreductase, 16.5 kDa subunit, putative	25.7%
AF1832	F420H2:quinone oxidoreductase, 32 kDa subunit (nuoI)	95.5%
AF1833	F420H2:quinone oxidoreductase, 39 kDa subunit, putative	33.6%
AF1829	F420H2:quinone oxidoreductase, 39.7 kDa subunit, putative	43.8%
AF1831	F420H2:quinone oxidoreductase, 41.2 kDa subunit, putative	34.8%
AF1827	F420H2:quinone oxidoreductase, 43.2 kDa subunit, putative	26.9%
AF1830	F420H2:quinone oxidoreductase, 45 kDa subunit (nuoD)	80.0%
AF1825	F420H2:quinone oxidoreductase, 53.9 kDa subunit (nuoM)	32.1%
AF1826	F420H2:quinone oxidoreductase, 72.4 kDa subunit (nuoL)	33.2%
AF0156	ferredoxin (fdx-1)	45.3%
AF0166	ferredoxin (fdx-2)	49.2%
AF0355	ferredoxin (fdx-3)	53.2%
AF0427	ferredoxin (fdx-4)	56.1%
AF0923	ferredoxin (fdx-5)	56.9%
AF1010	ferredoxin (fdx-6)	44.4%
AF1239	ferredoxin (fdx-7)	29.0%
AF2142	ferredoxin (fdx-8)	38.0%
AF0164	ferredoxin-nitrite reductase (nirA)	29.7%
AF2332	flavodoxin, putative	30.3%
AF0167	flavoprotein (fprA-1)	33.2%
AF1520	flavoprotein (fprA-2)	47.2%
AF0557	flavoprotein reductase	26.2%
AF1463	fumarate reductase, flavoprotein subunit (fdrA)	27.0%
AF1536	glutaredoxin (grx-1)	34.3%
AF2145	glutaredoxin (grx-2)	38.8%
AF0663	heterodisulfide reductase, subunit A (hdrA-1)	42.2%
AF1377	heterodisulfide reductase, subunit A (hdrA-2)	46.8%
AF0662	heterodisulfide reductase, subunit A/ methylviologen reducing hydrogenase, subunit delta	34.2%
AF1238	heterodisulfide reductase, subunit A/methylviologen reducing hydrogenase, subunit delta	53.7%
AF1375	heterodisulfide reductase, subunit B (hdrB)	36.0%
AF0271	heterodisulfide reductase, subunit B, putative	35.3%
AF1376	heterodisulfide reductase, subunit C (hdrC)	33.3%
AF0502	heterodisulfide reductase, subunit D, putative	33.8%
AF0809	heterodisulfide reductase, subunit D, putative	100.0%
AF0661	heterodisulfide reductase, subunit E, putative	23.8%
AF0755	heterodisulfide reductase, subunits E and D, putative	31.8%
AF0506	iron-sulfur binding reductase	38.5%
AF1773	iron-sulfur binding reductase	33.3%
AF1998	iron-sulfur binding reductase	29.6%
AF0627	iron-sulfur cluster binding protein	45.5%
AF0688	iron-sulfur cluster binding protein	44.8%
AF1153	iron-sulfur cluster binding protein	27.9%
AF1185	iron-sulfur cluster binding protein	36.7%
AF1263	iron-sulfur cluster binding protein	42.1%
AF2380	iron-sulfur cluster binding protein	35.3%
AF2381	iron-sulfur cluster binding protein	34.4%
AF2409	iron-sulfur cluster binding protein	28.2%
AF0076	iron-sulfur cluster binding protein	32.7%
AF1461	iron-sulfur cluster binding protein, putative	51.0%
AF1436	iron-sulfur flavoprotein (isf-1)	35.7%
AF1519	iron-sulfur flavoprotein (isf-2)	56.6%
AF1896	iron-sulfur flavoprotein (isf-3)	37.1%
AF1372	methylviologen-reducing hydrogenase, subunit alpha (vhuA)	39.4%
AF1374	methylviologen-reducing hydrogenase, subunit delta (vhuD)	41.7%
AF1373	methylviologen-reducing hydrogenase, subunit gamma (vhuG)	38.6%
AF0157	molybdopterin oxidoreductase, iron-sulfur binding subunit	38.6%
AF0174	molybdopterin oxidoreductase, membrane subunit	26.0%
AF0175	molybdopterin oxidoreductase, iron-sulfur binding subunit	42.0%
AF0176	molybdopterin oxidoreductase, molybdopterin binding subunit	32.6%
AF0499	molybdopterin oxidoreductase, iron-sulfur binding subunit	41.5%
AF0500	molybdopterin oxidoreductase, membrane subunit	27.9%
AF1202	molybdopterin oxidoreductase, iron-sulfur binding subunit	35.5%
AF1203	molybdopterin oxidoreductase, molybdopterin binding subunit	30.1%
AF2384	molybdopterin oxidoreductase, molybdopterin binding subunit	34.6%
AF2385	molybdopterin oxidoreductase, iron-sulfur binding subunit	46.9%
AF2386	molybdopterin oxidoreductase, membrane subunit	30.3%
AF0159	molybdopterin oxidoreductase, molybdopterin binding subunit, putative	30.9%
AF2267	NAD(P)H-flavin oxidoreductase	31.4%
AF0131	NAD(P)H-flavin oxidoreductase, putative	28.2%
AF2352	NADH dehydrogenase, subunit 1, putative	28.9%
AF1828	NADH dehydrogenase, subunit 3	24.3%
AF0248	NADH-dependent flavin oxidoreductase	36.7%
AF0342	nigerythrin, putative	33.3%
AF0546	nitrate reductase, gamma subunit (narI)	30.1%
AF0501	nitrate reductase, gamma subunit, putative	29.3%
AF1126	P450 cytochrome, putative	30.5%
AF0463	polyferredoxin (mvhB), authentic frameshift	32.2%
AF1379	quinone-reactive Ni/Fe-hydrogenase B-type cytochrome subunit (hydC)	29.0%
AF0173	reductase, assembly protein	30.0%
AF0547	reductase, iron-sulfur binding subunit	28.3%
AF0867	reductase, putative	33.3%
AF0880	rubredoxin (rd-1)	69.2%
AF1349	rubredoxin (rd-2)	67.9%
AF0832	rubrerythrin (rr1)	45.7%
AF0831	rubrerythrin (rr2)	63.7%
AF1640	rubrerythrin (rr3)	37.8%
AF2312	rubrerythrin (rr4)	41.4%
AF0711	thioredoxin (trx-1)	28.4%
AF0769	thioredoxin (trx-2)	38.5%
AF1284	thioredoxin (trx-3)	52.9%
AF2144	thioredoxin (trx-4)	48.9%
AF1339	ubiquinol-cytochrome C reductase complex, subunit VI requiring protein	60.9%
Fermentation		
AF1779	2-hydroxyacid dehydrogenase, putative	37.6%
AF0469	2-ketoglutarate ferredoxin oxidoreductase, subunit alpha (korA)	52.3%
AF0468	2-ketoglutarate ferredoxin oxidoreductase, subunit beta (korB)	51.2%
AF0470	2-ketoglutarate ferredoxin oxidoreductase, subunit delta (korD)	47.2%
AF0471	2-ketoglutarate ferredoxin oxidoreductase, subunit gamma (korG)	40.0%
AF2053	2-ketoisovalerate ferredoxin oxidoreductase, subunit alpha (vorA)	41.2%
AF2052	2-ketoisovalerate ferredoxin oxidoreductase, subunit beta (vorB)	42.7%
AF2054	2-ketoisovalerate ferredoxin oxidoreductase, subunit delta (vorD)	51.5%
AF2055	2-ketoisovalerate ferredoxin oxidoreductase, subunit gamma (vorG)	45.2%
AF0749	2-oxoacid ferredoxin oxidoreductase, subunit alpha (orA)	33.7%
AF0750	2-oxoacid ferredoxin oxidoreductase, subunit beta (orB)	49.2%
AF1286	acetoin utilization protein, putative	35.1%
AF0197	acetyl-CoA synthetase (acs-1)	27.1%
AF0366	acetyl-CoA synthetase (acs-2)	47.3%
AF0677	acetyl-CoA synthetase (acs-3)	40.9%
AF0975	acetyl-CoA synthetase (acs-4)	42.3%
AF0976	acetyl-CoA synthetase (acs-5)	36.2%
AF1287	acetyl-CoA synthetase (acs-6)	34.3%
AF0024	alcohol dehydrogenase, iron-containing	36.2%
AF0339	alcohol dehydrogenase, iron-containing	37.4%
AF2019	alcohol dehydrogenase, iron-containing	35.7%
AF2389-C	acetyl-CoA synthetase, putative	64.8%
AF2389-N	acetyl-CoA synthetase, putative	59.3%
AF2101	alcohol dehydrogenase, zinc-dependent	34.8%
AF0023	aldehyde ferredoxin oxidoreductase (aor-1)	41.1%
AF0077	aldehyde ferredoxin oxidoreductase (aor-2)	32.6%
AF0340	aldehyde ferredoxin oxidoreductase (aor-3)	38.4%
AF2281	aldehyde ferredoxin oxidoreductase (aor-4)	53.0%
AF0006	corrinoid methyltransferase protein (mtaC-1)	30.7%
AF0011	corrinoid methyltransferase protein (mtaC-2)	29.5%
AF0394	D-lactate dehydrogenase, cytochrome-type (dld)	31.9%
AF0560	formate dehydrogenase (fdhD1), authentic frameshift	32.9%
AF1199	glutaconate CoA-transferase, subunit A (gctA)	31.9%
AF1198	glutaconate CoA-transferase, subunit B (gctB), authentic frameshift	37.0%
AF1489	indolepyruvate ferredoxin oxidoreductase, subunit alpha (iorA)	48.1%
AF2030	indolepyruvate ferredoxin oxidoreductase, subunit beta (iorB)	41.1%
AF0807	L-lactate dehydrogenase, cytochrome-type (lldD)	39.4%
AF0855	L-malate dehydrogenase, NAD+-dependent (mdhA)	40.1%
AF2085	oxaloacetate decarboxylase, biotin carboxyl carrier subunit, putative	38.7%
AF2084	oxaloacetate decarboxylase, sodium ion pump subunit (oadB)	59.8%
AF1252	oxaloacetate decarboxylase, subunit alpha (oadA)	63.3%
AF1701	pyruvate ferredoxin oxidoreductase, subunit alpha (porA)	50.3%
AF1702	pyruvate ferredoxin oxidoreductase, subunit beta (porB)	50.7%
AF1700	pyruvate ferredoxin oxidoreductase, subunit delta (porD)	53.1%
AF1699	pyruvate ferredoxin oxidoreductase, subunit gamma (porG)	50.8%
Gluconeogenesis		
AF0710	phosphoenolpyruvate synthase (ppsA)	61.4%
Glycolysis		
AF1146	3-phosphoglycerate kinase (pgk)	48.8%
AF1132	enolase (eno)	53.9%
AF1732	glyceraldehyde 3-phosphate dehydrogenase (gap)	56.6%
AF1304	triosephosphate isomerase (tpiA)	56.4%
Pentose phosphate pathway		
AF0943	ribose 5-phosphate isomerase (rpi)	48.9%
Sugars		
AF0356	carbohydrate kinase, pfkB family	31.3%
AF0401	carbohydrate kinase, pfkB family	34.1%
AF1324	carbohydrate kinase, FGGY family	27.1%
AF1752	carbohydrate kinase, FGGY family	29.3%
AF0861	D-arabino 3-hexulose 6-phosphate formaldehyde lyase (hps-1)	30.6%
AF1305	D-arabino 3-hexulose 6-phosphate formaldehyde lyase (hps-2)	44.2%
AF0480	fuculose-1-phosphate aldolase (fucA)	31.8%
TCA cycle		
AF1963	aconitase (acn)	57.1%
AF1340	citrate synthase (citZ)	50.3%
AF1098	fumarase (fum-1)	49.1%
AF1099	fumarase (fum-2)	53.4%
AF0647	isocitrate dehydrogenase, NADP (icd)	57.2%
AF1727	malate oxidoreductase (mae)	52.3%
AF0681	succinate dehydrogenase, flavoprotein subunit A (sdhA)	48.2%
AF0682	succinate dehydrogenase, iron-sulfur subunit B (sdhB)	51.3%
AF0683	succinate dehydrogenase, subunit C (sdhC)	36.6%
AF0684	succinate dehydrogenase, subunit D (sdhD)	25.9%
AF1539	succinyl-CoA synthetase, alpha subunit (sucD-1)	56.9%
AF2185	succinyl-CoA synthetase, alpha subunit (sucD-2)	63.5%
AF1540	succinyl-CoA synthetase, beta subunit (sucC-1)	51.3%
AF2186	succinyl-CoA synthetase, beta subunit (sucC-2)	49.6%

FATTY ACID AND PHOSPHOLIPID METABOLISM

General		
AF1736	3-hydroxy-3-methylglutaryl-coenzyme A reductase (mvaA)	57.1%
AF0017	3-hydroxyacyl-CoA dehydrogenase (hbd-1)	41.1%
AF0285	3-hydroxyacyl-CoA dehydrogenase (hbd-2)	55.8%
AF0434	3-hydroxyacyl-CoA dehydrogenase (hbd-3)	40.7%
AF1025	3-hydroxyacyl-CoA dehydrogenase (hbd-4)	45.6%
AF1122	3-hydroxyacyl-CoA dehydrogenase (hbd-5)	45.2%
AF1177	3-hydroxyacyl-CoA dehydrogenase (hbd-6)	35.8%
AF1190	3-hydroxyacyl-CoA dehydrogenase (hbd-7)	46.5%
AF1206	3-hydroxyacyl-CoA dehydrogenase (hbd-8)	36.3%
AF2017	3-hydroxyacyl-CoA dehydrogenase (hbd-9)	35.4%
AF2273	3-hydroxyacyl-CoA dehydrogenase (hbd-10)	39.4%
AF0018	3-ketoacyl-CoA thiolase (acaB-1)	41.0%
AF0034	3-ketoacyl-CoA thiolase (acaB-2)	38.3%
AF0133	3-ketoacyl-CoA thiolase (acaB-3)	32.3%
AF0134	3-ketoacyl-CoA thiolase (acaB-4)	32.5%
AF0201	3-ketoacyl-CoA thiolase (acaB-5)	26.9%
AF0202	3-ketoacyl-CoA thiolase (acaB-6)	33.5%
AF0283	3-ketoacyl-CoA thiolase (acaB-7)	42.0%
AF0438	3-ketoacyl-CoA thiolase (acaB-8)	42.4%
AF0967	3-ketoacyl-CoA thiolase (acaB-9)	33.7%
AF0968	3-ketoacyl-CoA thiolase (acaB-10)	28.0%
AF1291	3-ketoacyl-CoA thiolase (acaB-11)	40.1%
AF2416	3-ketoacyl-CoA thiolase (acaB-12)	49.9%
AF1028	3-ketoacyl-CoA thiolase (fadA-1)	38.8%
AF1197	3-ketoacyl-CoA thiolase (fadA-2)	47.2%
AF2243	3-ketoacyl-CoA thiolase (fadA-3)	40.3%
AF0033	acyl carrier protein synthase (acaA-1)	28.6%
AF2415	acyl carrier protein synthase (acaA-2)	58.7%
AF0199	acyl-CoA dehydrogenase (acd-1)	35.9%
AF0436	acyl-CoA dehydrogenase (acd-2)	44.1%
AF0498	acyl-coA dehydrogenase (acd-3)	22.9%
AF0671	acyl-CoA dehydrogenase (acd-4)	37.9%
AF0845	acyl-CoA dehydrogenase (acd-5)	44.6%
AF0964	acyl-CoA dehydrogenase (acd-6)	35.8%
AF1026	acyl-CoA dehydrogenase (acd-7)	42.6%
AF1141	acyl-CoA dehydrogenase (acd-8)	43.2%
AF1293	acyl-CoA dehydrogenase (acd-9)	45.8%
AF2057	acyl-CoA dehydrogenase (acd-10)	44.6%
AF2244	acyl-CoA dehydrogenase (acd-11)	42.6%
AF2275	acyl-CoA dehydrogenase (acd-12)	38.9%
AF1175	acyl-CoA dehydrogenase, short chain-specific (acdS)	30.1%
AF0818	acylphosphatase (acyP)	36.8%
AF0868	alkyldihydroxyacetonephosphate synthase	33.6%
AF2286	bifunctional short chain isoprenyl diphosphate synthase (idsA)	42.7%
AF0220	biotin carboxylase (acc)	59.1%
AF0865	carboxylesterase (est-1)	27.1%
AF1537	carboxylesterase (est-2)	29.0%
AF2336	carboxylesterase (est-3)	30.4%
AF1716	carboxylesterase (estA)	40.4%
AF1744	CDP-diacylglycerol–glycerol-3-phosphate 3-phosphatidyltransferase (pgsA-2)	26.7%
AF1143	CDP-diacylglycerol–glycerol-3-phosphate-3-phosphatidyltransferase (pgsA-1)	27.0%
AF2044	CDP-diacylglycerol–serine O-phosphatidyltransferase (pssA)	36.6%
AF0435	enoyl-CoA hydratase (fad-1)	47.6%
AF0685	enoyl-CoA hydratase (fad-2)	39.9%
AF0963	enoyl-CoA hydratase (fad-3)	48.6%
AF1641	enoyl-CoA hydratase (fad-4)	41.7%
AF2429	enoyl-CoA hydratase (fad-5)	34.7%
AF1763	lipase, putative	33.5%
AF0089	long-chain-fatty-acid–CoA ligase (fadD-1)	31.9%
AF0200	long-chain-fatty-acid–CoA ligase (fadD-2)	34.8%
AF0687	long-chain-fatty-acid–CoA ligase (fadD-3)	31.1%
AF0840	long-chain-fatty-acid–CoA ligase (fadD-4)	38.1%
AF1029	long-chain-fatty-acid–CoA ligase (fadD-5)	37.8%
AF1510	long-chain-fatty-acid–CoA ligase (fadD-6)	36.0%
AF1772	long-chain-fatty-acid–CoA ligase (fadD-7)	38.7%
AF1932	long-chain-fatty-acid–CoA ligase (fadD-8)	31.0%
AF2368	long-chain-fatty-acid–CoA ligase (fadD-9)	38.7%
AF1753	lysophospholipase	33.5%
AF0196	medium-chain acyl-CoA ligase (alkK-1)	34.6%
AF0262	medium-chain acyl-CoA ligase (alkK-2)	38.6%
AF0672	medium-chain acyl-CoA ligase (alkK-3)	31.0%
AF1261	medium-chain acyl-CoA ligase (alkK-4)	42.7%
AF2033	medium-chain acyl-CoA ligase (alkK-5)	33.5%
AF2289	mevalonate kinase (mvk)	40.6%
AF1794	myo-inositol-1-phosphate synthase (ino1)	32.2%
AF2045	phosphatidylserine decarboxylase (psd2)	42.5%
AF1674	sn-glycerol-1-phosphate dehydrogenase (gldA)	44.0%

AUTOTROPHIC METABOLISM

General		
AF1100	acetyl-CoA decarbonylase/synthase, subunit alpha (cdhA-1)	50.4%
AF2397	acetyl-CoA decarbonylase/synthase, subunit alpha (cdhA-2)	54.0%
AF0379	acetyl-CoA decarbonylase/synthase, subunit beta (cdhC)	62.7%
AF0377	acetyl-CoA decarbonylase/synthase, subunit delta (cdhD)	57.4%
AF1101	acetyl-CoA decarbonylase/synthase, subunit epsilon (cdhB-1)	40.0%
AF2398	acetyl-CoA decarbonylase/synthase, subunit epsilon (cdhB-2)	38.9%
AF0376	acetyl-CoA decarbonylase/synthase, subunit gamma (cdhE)	55.4%
AF1849	carbon monoxide dehydrogenase, catalytic subunit (cooS)	39.9%
AF0950	carbon monoxide dehydrogenase, iron sulfur subunit (cooF)	38.9%
AF1535	ferredoxin-thioredoxin reductase, catalytic subunit (ftrB)	38.6%
AF2073	formylmethanofuran:tetrahydromethanopterin formyltransferase (ftr-1)	46.0%
AF2207	formylmethanofuran:tetrahydromethanopterin formyltransferase (ftr-2)	68.4%

Nature © Macmillan Publishers Ltd 1997

AF1935	N5,N10-methenyltetrahydromethanopterin cyclohydrolase (mch)	97.3%
AF0714	N5,N10-methylenetetrahydromethanopterin dehydrogenase (mtd)	61.8%
AF1066	N5,N10-methylenetetrahydromethanopterin reductase (mer-1)	59.1%
AF1196	N5,N10-methylenetetrahydromethanopterin reductase (mer-2)	37.4%
AF0009	N5-methyltetrahydromethanopterin:coenzyme M methyltransferase (mtr)	42.1%
AF1587	ribulose bisphosphate carboxylase, large subunit (rbcL-1)	40.6%
AF1638	ribulose bisphosphate carboxylase, large subunit (rbcL-2)	44.9%
AF1930	tungsten formylmethanofuran dehydrogenase, subunit A (fwdA)	48.9%
AF1650	tungsten formylmethanofuran dehydrogenase, subunit B (fwdB-1)	37.0%
AF1929	tungsten formylmethanofuran dehydrogenase, subunit B (fwdB-2)	49.4%
AF1931	tungsten formylmethanofuran dehydrogenase, subunit C (fwdC)	44.1%
AF1651	tungsten formylmethanofuran dehydrogenase, subunit D (fwdD-1)	32.6%
AF1928	tungsten formylmethanofuran dehydrogenase, subunit D (fwdD-2)	52.6%
AF0177	tungsten formylmethanofuran dehydrogenase, subunit E (fwdE)	29.7%
AF1644	tungsten formylmethanofuran dehydrogenase, subunit F (fwdF)	38.2%
AF1649	tungsten formylmethanofuran dehydrogenase, subunit G (fwdG)	45.6%

PURINES, PYRIMIDINES, NUCLEOSIDES, AND NUCLEOTIDES

2'-Deoxyribonucleotide metabolism

AF1108	deoxycytidine triphosphate deaminase, putative	38.1%
AF1664	ribonucleotide reductase (nrd)	59.7%
AF1554	thioredoxin reductase (trxB)	45.2%
AF2047	thymidylate synthase, putative	33.1%

Nucleotide and nucleoside interconversions

AF0876	5'-nucleotidase (nt5)	30.9%
AF0676	adenylate kinase (adk)	56.1%
AF1900	cytidylate kinase (cmk)	48.6%
AF0767	nucleoside diphosphate kinase (ndk)	56.4%
AF0061	thymidylate kinase (tmk)	34.9%
AF1308	thymidylate kinase, putative	26.3%
AF2042	uridylate kinase (pyrH)	53.6%

Purine ribonucleotide biosynthesis

AF2242	adenylosuccinate lyase (purB)	52.3%
AF0841	adenylosuccinate synthetase (purA)	70.8%
AF0873	amidophosphoribosyltransferase (purF)	55.8%
AF0253	GMP synthase (guaA-1)	59.8%
AF1320	GMP synthase (guaA-2)	49.4%
AF1811	inosine monophosphate cyclohydrolase	38.3%
AF0847	inosine monophosphate dehydrogenase (guaB-1)	41.6%
AF2118	inosine monophosphate dehydrogenase (guaB-2)	31.9%
AF1259	inosine monophosphate dehydrogenase, putative	51.6%
AF1157	phosphoribosylamine–glycine ligase (purD)	40.9%
AF1271	phosphoribosylaminoimidazole carboxylase (purE)	42.8%
AF1272	phosphoribosylaminoimidazolesuccinocarboxamide synthase (purC)	34.7%
AF1693	phosphoribosylformylglycinamidine cyclo-ligase (purM)	53.8%
AF1260	phosphoribosylformylglycinamidine synthase I (purQ)	40.9%
AF1940	phosphoribosylformylglycinamidine synthase II (purL)	41.5%
AF0589	ribose-phosphate pyrophosphokinase (prsA-1)	35.0%
AF1419	ribose-phosphate pyrophosphokinase (prsA-2)	41.1%

Pyrimidine ribonucleotide biosynthesis

AF0106	aspartate carbamoyltransferase, catalytic subunit (pyrB)	60.7%
AF0107	aspartate carbamoyltransferase, regulatory subunit (pyrI)	48.2%
AF1274	carbamoyl-phosphate synthase, large subunit (carB)	65.1%
AF1273	carbamoyl-phosphate synthase, small subunit (carA)	55.2%
AF0252	CTP synthase (pyrG)	58.3%
AF2250	dihydroorotase (pyrC)	37.2%
AF0745	dihydroorotase dehydrogenase (pyrD)	44.8%
AF1741	orotate phosphoribosyl transferase (pyrE)	49.0%
AF0386	orotate phosphoribosyl transferase, putative	39.0%

Salvage of nucleosides and nucleotides

AF0240	adenine deaminase (adeC)	39.5%
AF1764	dCMP deaminase, putative	39.0%
AF1788	methylthioadenosine phosphorylase (mtaP)	40.0%
AF1341	thymidine phosphorylase (deoA-1)	46.7%
AF1342	thymidine phosphorylase (deoA-2)	40.7%
AF0239	xanthine-guanine phosphoribosyltransferase (gptA-1)	25.7%
AF1789	xanthine-guanine phosphoribosyltransferase (gptA-2)	28.2%

REGULATORY FUNCTIONS

AF1959	(R)-hydroxyglutaryl-CoA dehydratase activator (hgdC)	51.2%
AF0168	arsenical resistance operon repressor, putative	36.7%
AF2204	arylsulfatase regulatory protein, putative	29.9%
AF0074	biotin operon repressor/biotin–[acetyl CoA carboxylase] ligase (birA)	36.6%
AF1724	dinitrogenase reductase activating glycohydrolase (draG)	37.9%
AF2232	ferric uptake regulation protein (fur)	25.8%
AF1785	iron-dependent repressor	42.0%
AF2395	iron-dependent repressor	40.0%
AF0245	iron-dependent repressor (desR)	28.2%
AF1984	iron-dependent repressor (troR)	28.3%
AF2430	lacZ expression regulatory protein (icc)	29.6%
AF1622	leucine responsive regulatory protein (lrp)	29.1%
AF0673	mercuric resistance operon regulatory protein (merR)	37.6%
AF2425	methanol dehydrogenase regulatory protein (moxR)	48.3%
AF1475	mitochondrial benzodiazepine receptor/sensory transduction protein	38.4%
AF0198	monoamine oxidase regulatory protein, putative	41.7%
AF1933	monoamine oxidase regulatory protein, putative	38.9%
AF0978	nitrogen regulatory protein P-II (glnB-1)	61.7%
AF1747	nitrogen regulatory protein P-II (glnB-2)	58.0%
AF1750	nitrogen regulatory protein P-II (glnB-3)	60.7%
AF0331	pheromone shutdown protein (traB)	40.5%
AF1797	phosphate regulatory protein, putative	30.7%
AF0521	protease synthase and sporulation regulator Pai1, putative	52.4%
AF1627	repressor protein	59.1%
AF1793	repressor protein	54.5%
AF0449	response regulator	38.1%
AF1063	response regulator	36.3%
AF1256	response regulator	42.5%
AF1384	response regulator	44.7%
AF1473	response regulator	32.5%
AF1898	response regulator	48.7%
AF2249	response regulator	44.8%
AF2419	response regulator	37.9%
AF0004	RNase L inhibitor	54.5%
AF0021	signal-transducing histidine kinase	26.1%
AF0208	signal-transducing histidine kinase	27.9%
AF0450	signal-transducing histidine kinase	32.4%
AF0770	signal-transducing histidine kinase	26.9%
AF0893	signal-transducing histidine kinase	28.7%
AF1184	signal-transducing histidine kinase	29.8%
AF1452	signal-transducing histidine kinase	28.5%
AF1467	signal-transducing histidine kinase	37.4%
AF1472	signal-transducing histidine kinase	30.4%
AF1483	signal-transducing histidine kinase	27.7%
AF1515	signal-transducing histidine kinase	32.0%
AF1639	signal-transducing histidine kinase	29.9%
AF1721	signal-transducing histidine kinase	34.5%
AF2109	signal-transducing histidine kinase	31.6%
AF0881	signal-transducing histidine kinase, authentic frameshift	26.5%
AF0277	signal-transducing histidine kinase, putative	29.8%
AF0410	signal-transducing histidine kinase, putative	27.1%
AF0448	signal-transducing histidine kinase, putative	26.1%
AF1620	signal-transducing histidine kinase, putative	26.2%
AF2032	signal-transducing histidine kinase, putative	22.5%
AF2420	signal-transducing histidine kinase, putative	28.4%
AF0442	succinoglycan biosynthesis regulator (exsB)	37.2%
AF1516	sugar fermentation stimulation protein (sfsA)	31.0%
AF1270	transcriptional regulatory protein, ArsR family	35.4%
AF1544	transcriptional regulatory protein, ArsR family	32.3%
AF1853	transcriptional regulatory protein, ArsR family	34.9%
AF2136	transcriptional regulatory protein, ArsR family	39.8%
AF0439	transcriptional regulatory protein, AsnC family	29.8%
AF0474	transcriptional regulatory protein, AsnC family	51.0%
AF0584	transcriptional regulatory protein, AsnC family	35.3%
AF1121	transcriptional regulatory protein, AsnC family	35.8%
AF1148	transcriptional regulatory protein, AsnC family	32.6%
AF1404	transcriptional regulatory protein, AsnC family	45.1%
AF1448	transcriptional regulatory protein, AsnC family	30.6%
AF1723	transcriptional regulatory protein, AsnC family	46.4%
AF1743	transcriptional regulatory protein, AsnC family	34.9%
AF2127	transcriptional regulatory protein, LysR family	30.8%
AF0114	transcriptional regulatory protein, putative	35.6%
AF1968	transcriptional regulatory protein, Rok family	32.9%
AF0112	transcriptional regulatory protein, Sir2 family	38.9%
AF1676	transcriptional regulatory protein, Sir2 family	40.6%
AF1817	transcriptional regulatory protein, TetR family	24.5%
AF0363	transcriptional repressor (cinR)	27.5%

REPLICATION

DNA replication, restriction, modification, recombination, and repair

AF2117	3-methyladenine DNA glycosylase (alkA)	30.0%
AF2060	activator 1, replication factor C, 35 KDa subunit	66.3%
AF1195	activator 1, replication factor C, 53 KDa subunit	43.7%
AF0465	DNA gyrase, subunit A (gyrA)	48.4%
AF0530	DNA gyrase, subunit B (gyrB)	58.4%
AF1388	DNA helicase, putative	46.8%
AF1960	DNA helicase, putative	32.7%
AF0623	DNA ligase (lig)	44.4%
AF1725	DNA ligase, putative	32.7%
AF0497	DNA polymerase B1 (polB)	45.1%
AF0693	DNA polymerase B2 (boxA), authentic frameshift	32.3%
AF0972	DNA polymerase III, subunit epsilon (dnaQ)	31.9%
AF2277	DNA polymerase, bacteriophage-type	36.9%
AF0742	DNA primase, putative	26.8%
AF0264	DNA repair protein RAD2 (rad2)	44.4%
AF0358	DNA repair protein RAD25	32.5%
AF1031	DNA repair protein RAD32 (rad32)	37.6%
AF0993	DNA repair protein RAD51 (radA)	59.3%
AF2096	DNA repair protein REC	40.0%
AF2418	DNA repair protein, putative	28.9%
AF1806	DNA topoisomerase I (topA)	36.2%
AF0940	DNA topoisomerase VI, subunit A (top6A)	39.8%
AF0652	DNA topoisomerase VI, subunit B (top6B)	43.9%
AF1692	endonuclease III (nth)	44.3%
AF0580	exodeoxyribonuclease III (xthA)	41.3%
AF2314	methylated-DNA-protein-cysteine methyltransferase (ogt)	55.3%
AF1409	modification methylase, type III R/M system	31.4%
AF1234	mutator protein MutT (mutT)	63.6%
AF2200	mutator protein MutT, putative	42.0%
AF0335	proliferating-cell nuclear antigen (pol30)	33.7%
AF0694	replication control protein A, putative	30.2%
AF1024	reverse gyrase (top-RG)	40.7%
AF0621	ribonuclease HII (rnhB)	39.3%
AF1715	type I restriction-modification enzyme, M subunit, authentic frameshift	63.0%
AF1708	type I restriction-modification enzyme, R subunit	38.2%
AF1710	type I restriction-modification enzyme, S subunit	33.0%

TRANSCRIPTION

DNA-dependent RNA polymerase

AF1888	DNA-directed RNA polymerase, subunit A' (rpoA1)	63.6%
AF1889	DNA-directed RNA polymerase, subunit A'' (rpoA2)	55.7%
AF1887	DNA-directed RNA polymerase, subunit B' (rpoB1)	65.3%
AF1886	DNA-directed RNA polymerase, subunit B'' (rpoB2)	57.1%
AF2282	DNA-directed RNA polymerase, subunit D (rpoD)	34.6%
AF1117	DNA-directed RNA polymerase, subunit E' (rpoE1)	48.4%
AF1116	DNA-directed RNA polymerase, subunit E'' (rpoE2)	40.0%
AF1885	DNA-directed RNA polymerase, subunit H (rpoH)	59.5%
AF1131	DNA-directed RNA polymerase, subunit K (rpoK)	61.5%
AF0207	DNA-directed RNA polymerase, subunit L (rpoL)	42.0%
AF1130	DNA-directed RNA polymerase, subunit N (rpoN)	58.8%

Transcription factors

AF1813	TBP-interacting protein TIP49	45.7%
AF1299	transcription initiation factor IIB	60.4%
AF0373	transcription initiation factor IID	59.4%
AF0757	transcription initiation factor IIE, subunit alpha, putative	23.5%
AF1891	transcription termination-antitermination factor NusA, putative	48.9%
AF1235	transcription-associated protein TFIIS	59.0%

RNA processing

AF1783	dimethyladenosine transferase (ksgA)	44.7%
AF2087	fibrillarin (fib)	49.3%
AF0482	mRNA 3'-end processing factor, putative	55.5%
AF0532	mRNA 3'-end processing factor, putative	39.1%
AF2361	mRNA 3'-end processing factor, putative	30.5%
AF2399	rRNA methylase, putative	36.4%
AF0362	snRNP, putative	32.0%
AF0875	snRNP, putative	35.7%

TRANSLATION

Amino acyl tRNA synthetases

AF2255	alanyl-tRNA synthetase (alaS)	47.1%
AF0894	arginyl-tRNA synthetase (argS)	48.8%
AF0920	aspartyl-tRNA synthetase (aspS)	62.5%
AF0411	cysteinyl-tRNA synthetase (cysS)	46.1%
AF0260	glutamyl-tRNA synthetase (gltX)	44.9%
AF0916	glycyl-tRNA synthetase (glyS)	51.2%
AF1642	histidyl-tRNA synthetase (hisS)	46.0%
AF0633	isoleucyl-tRNA synthetase (ileS)	48.9%
AF2421	leucyl-tRNA synthetase (leuS)	49.7%
AF1216	lysyl-tRNA synthetase (lysS)	43.6%
AF1453	methionyl-tRNA synthetase (metS)	45.2%
AF1955	phenylalanyl-tRNA synthetase, subunit alpha (pheS)	44.4%
AF1424	phenylalanyl-tRNA synthetase, subunit beta (pheT)	42.6%
AF1609	prolyl-tRNA synthetase (proS)	56.8%
AF2035	seryl-tRNA synthetase (serS)	45.4%
AF0548	threonyl-tRNA synthetase (thrS)	46.9%
AF1694	tryptophanyl-tRNA synthetase (trpS)	52.4%
AF0776	tyrosyl-tRNA synthetase (tyrS)	57.6%
AF2224	valyl-tRNA synthetase (valS)	54.5%

Degradation of proteins, peptides, and glycopeptides

AF1976	26S protease regulatory subunit 4	66.0%
AF1653	alkaline serine protease (aprM)	44.5%
AF0578	aminopeptidase, putative	27.8%
AF0364	ATP-dependent protease La (lon)	36.6%
AF1946	cysteine proteinase, putative	36.2%
AF1281	intracellular protease (pfpI)	56.0%
AF1112	O-sialoglycoprotein endopeptidase (gcp)	57.6%
AF0665	O-sialoglycoprotein endopeptidase, putative	35.6%
AF2086	protease inhibitor, putative	37.0%
AF0490	proteasome, subunit alpha (psmA)	60.8%
AF0481	proteasome, subunit beta (psmB)	58.3%
AF2034	X-pro aminopeptidase (pepQ)	34.6%

Protein modification

AF0656	antibiotic maturation protein (pmbA)	32.7%
AF0378	CODH nickel-insertion accessory protein (cooC-1)	35.7%
AF1685	CODH nickel-insertion accessory protein (cooC-2)	47.4%
AF1615	cofactor modifying protein (cmo)	27.2%
AF2195	deoxyhypusine synthase (dys1-1)	32.6%
AF2300	deoxyhypusine synthase (dys1-2)	34.9%
AF0381	diphthine synthase (dph5)	40.8%
AF2324	fmu and fmv protein	40.0%
AF1367	hydrogenase expression/formation protein (hypA)	40.4%
AF1368	hydrogenase expression/formation protein (hypB)	54.4%
AF1369	hydrogenase expression/formation protein (hypC)	40.5%
AF1370	hydrogenase expression/formation protein (hypD)	46.0%
AF1365	hydrogenase expression/formation protein (hypE)	51.5%
AF1366	hydrogenase expression/formation regulatory protein (hypF)	45.1%
AF0036	L-isoaspartyl protein carboxyl methyltransferase (pcm-1)	60.7%
AF2322	L-isoaspartyl protein carboxyl methyltransferase (pcm-2)	59.3%
AF1840	methionyl aminopeptidase (map)	48.6%
AF1989	peptidyl-prolyl cis-trans isomerase (slyD)	34.4%
AF0853	proliferating-cell nucleolar antigen P120, putative	35.7%
AF2039	proliferating-cell nucleolar antigen P120, putative	44.2%
AF1449	pyruvate formate-lyase 2 (pflD)	37.8%
AF1450	pyruvate formate-lyase 2 activating enzyme (pflC)	38.8%
AF0117	pyruvate formate-lyase activating enzyme (act-1)	25.5%
AF0918	pyruvate formate-lyase activating enzyme (act-2)	42.3%
AF1330	pyruvate formate-lyase activating enzyme (act-3)	45.8%
AF2278	pyruvate formate-lyase activating enzyme (act-4)	42.5%
AF1961	pyruvate formate-lyase activating enzyme (pflX)	50.2%
AF0380	transmembrane oligosaccharyl transferase, putative	27.8%
AF0329	transmembrane oligosaccharyl transferase, putative	29.3%

Ribosomal proteins: synthesis and modification

AF1490	LSU ribosomal protein L1P (rpl1P)	48.6%
AF1922	LSU ribosomal protein L2P (rpl2P)	60.4%
AF1925	LSU ribosomal protein L3P (rpl3P)	56.5%
AF1924	LSU ribosomal protein L4P (rpl4P)	56.4%
AF1912	LSU ribosomal protein L5P (rpl5P)	51.7%
AF1909	LSU ribosomal protein L6P (rpl6P)	53.7%
AF0764	LSU ribosomal protein L7AE (rpl7AE)	60.7%
AF1491	LSU ribosomal protein L10E (rpl10E)	45.6%
AF0538	LSU ribosomal protein L11P (rpl11P)	67.8%
AF1492	LSU ribosomal protein L12A (rpl12A)	76.0%
AF1128	LSU ribosomal protein L13P (rpl13P)	47.4%
AF1915	LSU ribosomal protein L14P (rpl14P)	66.7%
AF2319	LSU ribosomal protein L15E (rpl15E)	70.3%
AF1903	LSU ribosomal protein L15P (rpl15P)	53.8%
AF1127	LSU ribosomal protein L18E (rpl18E)	53.8%
AF1906	LSU ribosomal protein L18P (rpl18P)	57.8%
AF1907	LSU ribosomal protein L19E (rpl19E)	55.5%
AF1529	LSU ribosomal protein L21E (rpl21E)	53.2%
AF1920	LSU ribosomal protein L22P (rpl22P)	55.2%
AF1923	LSU ribosomal protein L23P (rpl23P)	55.6%
AF0537	LSU ribosomal protein L24A (rpl24A)	51.4%
AF0766	LSU ribosomal protein L24E (rpl24E)	66.1%
AF1914	LSU ribosomal protein L24P (rpl24P)	57.8%
AF1918	LSU ribosomal protein L29P (rpl29P)	44.6%
AF1890	LSU ribosomal protein L30E (rpl30E)	41.7%
AF1904	LSU ribosomal protein L30P (rpl30P)	55.9%
AF2066	LSU ribosomal protein L31E (rpl31E)	50.6%
AF1908	LSU ribosomal protein L32E (rpl32E)	51.2%
AF0057	LSU ribosomal protein L37AE (rpl37AE)	47.6%
AF0874	LSU ribosomal protein L37E (rpl37E)	57.9%
AF2067	LSU ribosomal protein L39E (rpl39E)	56.9%
AF1430	LSU ribosomal protein L40E (rpl40E)	73.3%
AF1333	LSU ribosomal protein L44E (rpl44E)	46.8%
AF2064	LSU ribosomal protein LXA (rplXA)	53.8%
AF0739	ribosomal protein S18 alanine acetyltransferase	38.5%
AF2303	ribosomal protein S6 modification protein (rimK)	32.2%
AF1133	SSU ribosomal protein S2P (rps2P)	58.3%
AF1919	SSU ribosomal protein S3P (rps3P)	50.0%
AF1913	SSU ribosomal protein S4E (rps4E)	48.9%
AF2284	SSU ribosomal protein S4P (rps4P)	59.1%
AF1905	SSU ribosomal protein S5P (rps5P)	60.0%
AF0611	SSU ribosomal protein S6E (rps6E)	50.8%
AF1893	SSU ribosomal protein S7P (rps7P)	59.6%
AF2152	SSU ribosomal protein S8E (rps8E)	61.6%
AF1910	SSU ribosomal protein S8P (rps8E)	64.6%
AF1129	SSU ribosomal protein S9P (rps9P)	59.5%
AF0938	SSU ribosomal protein S10P (rps10P)	71.0%
AF2283	SSU ribosomal protein S11P (rps11P)	71.1%
AF1892	SSU ribosomal protein S12P (rps12P)	74.1%
AF2285	SSU ribosomal protein S13P (rps13P)	52.1%
AF1911	SSU ribosomal protein S14P (rps14P)	61.5%
AF0801	SSU ribosomal protein S15P (rps15P)	62.0%
AF0911	SSU ribosomal protein S17E (rps17E)	52.6%
AF1916	SSU ribosomal protein S17P (rps17P)	59.0%
AF2069	SSU ribosomal protein S19E (rps19E)	64.2%
AF1921	SSU ribosomal protein S19P (rps19P)	60.9%
AF1114	SSU ribosomal protein S24E (rps24E)	40.2%
AF1113	SSU ribosomal protein S27AE (rps27AE)	60.0%
AF1334	SSU ribosomal protein S27E (rps27E)	49.0%
AF0765	SSU ribosomal protein S28E (rps28E)	55.6%
AF2320	SSU ribosomal protein S3AE (rps3AE)	38.9%

tRNA modification

AF0588	archaeosine tRNA-ribosyltransferase (tgtA)	52.0%
AF1954	Glu-tRNA amidotransferase, subunit A (gatA-1)	38.6%
AF2329	Glu-tRNA amidotransferase, subunit A (gatA-2)	53.5%
AF1440	Glu-tRNA amidotransferase, subunit B (gatB-1)	54.7%
AF2116	Glu-tRNA amidotransferase, subunit B (gatB-2)	46.4%

AF2328	Glu-tRNA amidotransferase, subunit C (gatC)	35.1%
AF0815	N2,N2-dimethylguanosine tRNA methyltransferase (trm1)	38.2%
AF1730	pseudouridylate synthase I (truA)	37.4%
AF1485	queuine tRNA-ribosyltransferase (tgtB)	44.1%
AF0493	ribonuclease PH (rph)	30.8%
AF0900	tRNA intron endonuclease (endA)	41.8%
AF2156	tRNA nucleotidyltransferase (cca)	43.9%
Translation factors		
AF2350	ATP-dependent RNA helicase HepA, putative	31.5%
AF2254	ATP-dependent RNA helicase, DEAD-family (deaD)	52.2%
AF0071	ATP-dependent RNA helicase, putative	29.6%
AF1458	ATP-dependent RNA helicase, putative	48.1%
AF2406	ATP-dependent RNA helicase, putative	35.2%
AF1149	large helicase-related protein (lhr-1)	34.5%
AF2177	large helicase-related protein (lhr-2), authentic frameshift	56.0%
AF1220	peptide chain release factor eRF, subunit 1	51.2%
AF2245	SKI2-family helicase, authentic frameshift	45.7%
AF0937	translation elongation factor EF-1, subunit alpha (tuf)	74.4%
AF0574	translation elongation factor EF-1, subunit beta	31.3%
AF1894	translation elongation factor EF-2 (fus)	62.5%
AF0777	translation initiation factor eIF-1A (eif1A)	57.5%
AF0527	translation initiation factor eIF-2, subunit alpha (eif2A)	51.1%
AF2326	translation initiation factor eIF-2, subunit beta, putative	45.5%
AF0592	translation initiation factor eIF-2, subunit gamma (eif2G)	64.4%
AF0370	translation initiation factor eIF-2B, subunit delta (eif2BD)	53.3%
AF2037	translation initiation factor eIF-2B, subunit delta (eif2BD)	57.9%
AF0645	translation initiation factor eIF-5A (eif5A)	50.4%
AF0768	translation initiation factor IF-2 (infB)	52.2%

TRANSPORT AND BINDING PROTEINS

General		
AF0393	ABC transporter, ATP-binding protein	34.5%
AF0984	ABC transporter, ATP-binding protein	35.2%
AF1006	ABC transporter, ATP-binding protein	35.1%
AF1018	ABC transporter, ATP-binding protein	57.7%
AF1021	ABC transporter, ATP-binding protein	37.8%
AF1136	ABC transporter, ATP-binding protein	39.3%
AF1139	ABC transporter, ATP-binding protein	38.2%
AF1300	ABC transporter, ATP-binding protein	34.1%
AF1469	ABC transporter, ATP-binding protein	43.5%
AF1819	ABC transporter, ATP-binding protein	51.1%
AF1982	ABC transporter, ATP-binding protein	41.3%
AF2364	ABC transporter, ATP-binding protein	53.5%
AF1005	ABC transporter, ATP-binding protein, putative	28.7%
AF1064	ABC transporter, ATP-binding protein, putative	36.0%
AF1983	ABC transporter, periplasmic binding protein	25.4%
AF1981	ABC transporter, permease protein	29.9%
AF1995	sodium- and chloride-dependent transporter	52.5%
Amino acids, peptides and amines		
AF1766	amino-acid ABC transporter, periplasmic binding protein/protein kinase	27.4%
AF0222	branched-chain amino acid ABC transporter, ATP-binding protein (braF-1)	42.7%
AF0822	branched-chain amino acid ABC transporter, ATP-binding protein (braF-2)	44.7%
AF0959	branched-chain amino acid ABC transporter, ATP-binding protein (braF-3)	37.6%
AF1390	branched-chain amino acid ABC transporter, ATP-binding protein (braF-4)	59.7%
AF0221	branched-chain amino acid ABC transporter, ATP-binding protein (braG-1)	48.2%
AF0823	branched-chain amino acid ABC transporter, ATP-binding protein (braG-2)	42.9%
AF0958	branched-chain amino acid ABC transporter, ATP-binding protein (braG-3)	34.1%
AF1389	branched-chain amino acid ABC transporter, ATP-binding protein (braG-4)	64.6%
AF0223	branched-chain amino acid ABC transporter, periplasmic binding protein (braC-1)	34.3%
AF0827	branched-chain amino acid ABC transporter, periplasmic binding protein (braC-2)	26.8%
AF0962	branched-chain amino acid ABC transporter, periplasmic binding protein (braC-3)	25.6%
AF1391	branched-chain amino acid ABC transporter, periplasmic binding protein (braC-4)	50.1%
AF0224	branched-chain amino acid ABC transporter, permease protein (braD-1)	25.4%
AF0825	branched-chain amino acid ABC transporter, permease protein (braD-2)	30.8%
AF0961	branched-chain amino acid ABC transporter, permease protein (braD-3)	23.9%
AF1392	branched-chain amino acid ABC transporter, permease protein (braD-4)	65.4%
AF0225	branched-chain amino acid ABC transporter, permease protein (braE-1)	28.7%
AF0824	branched-chain amino acid ABC transporter, permease protein (braE-2)	31.3%
AF0960	branched-chain amino acid ABC transporter, permease protein (braE-3)	30.1%
AF1393	branched-chain amino acid ABC transporter, permease protein (braE-4)	60.5%
AF1612	cationic amino acid transporter (cat-1)	29.5%
AF1774	cationic amino acid transporter (cat-2)	38.0%
AF1770	dipeptide ABC transporter, ATP-binding protein (dppD)	47.8%
AF1771	dipeptide ABC transporter, ATP-binding protein (dppF)	43.1%
AF1767	dipeptide ABC transporter, dipeptide-binding protein (dppA)	33.1%
AF1768	dipeptide ABC transporter, permease protein (dppB)	39.3%
AF1769	dipeptide ABC transporter, permease protein (dppC)	40.8%
AF0680	glutamine ABC transporter, ATP-binding protein (glnQ)	63.8%
AF0231	glutamine ABC transporter, periplasmic glutamine-binding protein (glnH)	38.0%
AF0232	glutamine ABC transporter, permease protein (glnP)	39.3%
AF0981	osmoprotection protein (proV)	39.0%
AF0979	osmoprotection protein (proW-1)	32.8%
AF0980	osmoprotection protein (proW-2)	36.8%
AF0982	osmoprotection protein (proX)	28.7%
AF0015	proline permease (putP-1)	26.2%
AF0969	proline permease (putP-2)	27.4%
AF1222	proline permease (putP-3)	27.0%
AF1608	spermidine/putrescine ABC transporter, ATP-binding protein (potA)	50.2%
AF1605	spermidine/putrescine ABC transporter, periplasmic spermidine/putrescine-binding protein (potD), authentic frameshift	31.0%
AF1607	spermidine/putrescine ABC transporter, permease protein (potB)	38.0%
AF1606	spermidine/putrescine ABC transporter, permease protein (potC)	38.7%
Anions		
AF2308	arsenite transport protein (arsB)	27.3%
AF1415	chloride channel, putative	27.3%
AF0025	cyanate transport protein (cynX)	24.5%
AF0087	nitrate ABC transporter, ATP-binding protein (nrtC-1)	47.4%
AF0638	nitrate ABC transporter, ATP-binding protein (nrtC-2)	55.5%
AF0640	nitrate ABC transporter, ATP-binding protein, putative	32.5%
AF0086	nitrate ABC transporter, permease protein (nrtB-1)	35.4%
AF0639	nitrate ABC transporter, permease protein (nrtB-2)	37.4%
AF1359	phosphate ABC transporter, ATP-binding protein (pstB)	66.0%
AF1356	phosphate ABC transporter, periplasmic phosphate-binding protein (phoX)	25.1%
AF1358	phosphate ABC transporter, permease protein (pstA)	34.1%
AF1357	phosphate ABC transporter, permease protein (pstC)	33.7%
AF1360	phosphate ABC transporter, regulatory protein (phoU)	26.9%
AF0791	phosphate permease, putative	31.1%
AF1798	phosphate permease, putative	52.9%
AF0092	sulfate ABC transporter, ATP-binding protein (cysA)	54.2%
AF0093	sulfate ABC transporter, permease protein (cysT)	44.1%
Carbohydrates, organic alcohols, and acids		
AF0347	C4-dicarboxylate transporter (mae1)	24.5%
AF1426	glycerol uptake facilitator, MIP channel (glpF)	36.2%
AF0013	hexuronate transporter (exuT)	25.1%
AF0806	L-lactate permease (lctP)	31.7%
AF0008	oxalate/formate antiporter (oxlT-1)	25.7%
AF0367	oxalate/formate antiporter (oxlT-2)	33.2%
AF1069	pantothenate permease (panF-1)	28.9%
AF1205	pantothenate permease (panF-2)	24.8%
AF0237	pantothenate permease (panF-3)	25.1%
AF0041	polysaccharide ABC transporter, ATP-binding protein (rfbB-1)	42.5%
AF0290	polysaccharide ABC transporter, ATP-binding protein (rfbB-2)	43.9%
AF0042	polysaccharide ABC transporter, permease protein (rfbA-1)	27.5%
AF0289	polysaccharide ABC transporter, permease protein (rfbA-2)	28.5%
AF0887	ribose ABC transporter, ATP-binding protein (rbsA-1)	33.3%
AF1170	ribose ABC transporter, ATP-binding protein (rbsA-2)	27.9%
AF0888	ribose ABC transporter, permease protein (rbsC-1)	24.1%
AF0889	ribose ABC transporter, permease protein (rbsC-2)	31.2%
AF2014	sugar transporter, putative	26.0%
Cations		
AF0977	ammonium transporter (amt-1)	44.3%
AF1746	ammonium transporter (amt-2)	49.0%
AF1749	ammonium transporter (amt-3)	41.5%
AF0473	cation-transporting ATPase, P-type (pacS)	44.0%
AF0152	copper-transporting ATPase, P-type (copB)	44.5%
AF0246	iron (II) transporter (feoB-1)	33.3%
AF2394	iron (II) transporter (feoB-2)	48.0%
AF0561	iron (II) transporter (feoB-3), authentic frameshift	29.4%
AF0430	iron (III) ABC transporter, ATP-binding protein (hemV-1)	50.4%
AF0432	iron (III) ABC transporter, ATP-binding protein (hemV-2)	58.7%
AF1401	iron (III) ABC transporter, ATP-binding protein (hemV-3)	35.2%
AF1397	iron (III) ABC transporter, periplasmic hemin-binding protein (hemT), authentic frameshift	28.2%
AF0431	iron (III) ABC transporter, permease protein (hemU-1)	36.2%
AF1402	iron (III) ABC transporter, permease protein (hemU-2)	35.2%
AF0786	magnesium and cobalt transporter (corA)	40.1%
AF0346	mercuric transport protein periplasmic component (merP)	35.2%
AF0217	Na+/H+ antiporter (napA-1)	28.2%
AF1245	Na+/H+ antiporter (napA-2)	28.4%
AF0846	Na+/H+ antiporter (nhe2)	33.1%
AF0715	potassium channel, putative	39.5%
AF1673	potassium channel, putative	36.3%
AF2197	potassium channel, putative	24.6%
AF0218	TRK potassium uptake system protein (trkA-1)	30.2%
AF0838	TRK potassium uptake system protein (trkA-2)	42.9%
AF0839	TRK potassium uptake system protein (trkH)	39.8%
Other		
AF0834	ferritin, putative	39.8%
AF1980	heme exporter protein C (helC)	29.0%
AF1144	multidrug resistance protein	29.2%
AF1325	multidrug resistance protein	29.9%
AF2258	multidrug resistance protein	31.3%

OTHER CATEGORIES

Adaptations and atypical conditions		
AF0508	ethylene-inducible protein	74.5%
AF0235	heat shock protein (htpX)	32.9%
AF0942	surE stationary-phase survival protein (surE)	50.2%
AF1996	virulence associated protein C (vapC-1)	50.0%
AF1690	virulence associated protein C (vapC-2)	30.0%
Drug and analog sensitivity		
AF1884	daunorubicin resistance ATP-binding protein (drrA)	47.1%
AF1883	daunorubicin resistance membrane protein (drrB)	27.0%
AF0487	penicillin G acylase	31.7%
AF1214	phenylacrylic acid decarboxylase (pad1)	43.2%
AF2194	rRNA (adenine-N6)-methyltransferase, putative	29.2%
AF1696	small multidrug export protein (qacE)	39.0%
Transposon-related functions		
AF0120	insertion sequence ISH S1, authentic frameshift	34.5%
AF0193	ISA0963-1, putative transposase, authentic frameshift	34.3%
AF0309	ISA0963-2, putative transposase	33.5%
AF1310	ISA0963-3, putative transposase	33.5%
AF1383	ISA0963-4, putative transposase	33.5%
AF1410	ISA0963-5, putative transposase	33.5%
AF1705	ISA0963-6, putative transposase	33.5%
AF1836	ISA0963-7, putative transposase, authentic frameshift	20.0%
AF0678	ISA1083-1, ISORF2	33.6%
AF0679	ISA1083-1, putative transposase	37.2%
AF1351	ISA1083-2, ISORF2	30.8%
AF1352	ISA1083-2, putative transposase	31.5%
AF2140	ISA1083-3, ISORF2	30.8%
AF2139	ISA1083-3, putative transposase	31.5%
AF0278	ISA1214-1, ISORF2	27.7%
AF0279	ISA1214-1, putative transposase	33.3%
AF0305	ISA1214-2, ISORF2	27.7%
AF0306	ISA1214-2, putative transposase	33.3%
AF0641	ISA1214-3, ISORF2	26.5%
AF0642	ISA1214-3, putative transposase	33.3%
AF0857	ISA1214-4, ISORF2	27.7%
AF0858	ISA1214-4, putative transposase	33.3%
AF2091	ISA1214-5, ISORF2	26.5%
AF2092	ISA1214-5, putative transposase	33.3%
AF2223	ISA1214-6, ISORF2	26.5%
AF2222	ISA1214-6, putative transposase	25.6%
AF0138	transposase IS240-A	43.3%
AF0895	transposase IS240-A	46.2%
AF2390	transposase, authentic frameshift	24.0%
AF0137	transposase, putative	29.6%
AF1628	transposase, putative	32.8%

UNKNOWN

AF0477	AAA superfamily ATPase	35.0%
AF0513	allene oxide synthase, putative	39.5%
AF0478	ATP-binding protein PhnP (phnP)	30.9%
AF1775	atrazine chlorohydrolase, putative	34.4%
AF0973	bile acid-inducible operon protein F (baiF-1)	30.8%
AF0974	bile acid-inducible operon protein F (baiF-2)	29.9%
AF1315	bile acid-inducible operon protein F (baiF-3)	31.3%
AF2063	c-myc binding protein, putative	21.7%
AF1992	calcium-binding protein, putative	31.2%
AF2287	carotenoid biosynthetic gene ERWCRTS, putative	49.4%
AF0512	chloroplast inner envelope membrane protein	42.5%
AF2251	competence-damage protein, putative	28.0%
AF0090	dehydrase	34.1%
AF1498	dehydrase, putative	29.4%
AF1518	DNA/pantothenate metabolism flavoprotein, putative	51.4%
AF0039	dolichol-P-glucose synthetase, putative	33.7%
AF0328	dolichol-P-glucose synthetase, putative	39.0%
AF0581	dolichol-P-glucose synthetase, putative	27.5%
AF0569	DR-beta chain MHC class II	37.7%
AF0383	endonuclease III, putative	47.1%
AF1150	erpK protein, putative	54.9%
AF2372	extragenic suppressor (suhB)	37.0%
AF1418	glycerol-3-phosphate cytidyltransferase (taqD)	56.6%
AF0744	GTP-binding protein	33.4%
AF1181	GTP-binding protein	36.3%
AF1364	GTP-binding protein	57.5%
AF2146	GTP-binding protein	65.9%
AF0428	GTP-binding protein, GTP1/OBG-family	43.9%
AF2237	HAM1 protein	31.4%
AF2211	HIT family protein (hit)	29.6%
AF0216	L-isoaspartyl protein carboxyl methyltransferase PimT, putative	35.5%
AF2313	maoC protein (maoC)	43.0%
AF0429	methyltransferase	43.8%
AF0186	nifS protein, class-V aminotransferase (nifS-1)	46.1%
AF0564	nifS protein, class-V aminotransferase (nifS-2)	45.1%
AF0185	nifU protein (nifU-1)	55.6%
AF0565	nifU protein (nifU-2)	55.6%
AF0632	nifU protein (nifU-3)	47.4%
AF1781	nodulation protein NfeD (nfeD)	33.4%
AF2269	nucleotide-binding protein	48.7%
AF2382	nucleotide-binding protein	49.1%
AF0374	p-nitrophenyl phosphatase (pho2)	31.7%
AF1978	periplasmic divalent cation tolerance protein (cutA)	31.3%
AF1652	prepro-subtilisin sendai, putative	35.6%
AF2021	rod shape-determining protein (mreB)	26.6%
AF1778	stage V sporulation protein (spoVG)	43.9%
AF1970	TPR domain-containing protein	29.0%
AF2202	tryptophan-specific permease, putative	25.2%
AF0816	vtpJ-therm, putative	42.1%
AF1679	vtpJ-therm, putative	45.1%