# The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*

**Gerard Deckert*†, Patrick V. Warren*†, Terry Gaasterland‡, William G. Young*, Anna L. Lenox*, David E. Graham§, Ross Overbeek‡, Marjory A. Snead*, Martin Keller*, Monette Aujay*, Robert Huber‖, Robert A. Feldman*, Jay M. Short*, Gary J. Olsen§ & Ronald V. Swanson***

* *Diversa Corporation, 10665 Sorrento Valley Road, San Diego, California 92121, USA*
‡ *Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA*
§ *Department of Microbiology, University of Illinois, Urbana, Illinois 61801, USA*
‖ *Lehrstuhl für Mikrobiologie, Universität Regensburg W-8400, Regensburg W-8400, Germany*

***Aquifex aeolicus* was one of the earliest diverging, and is one of the most thermophilic, bacteria known. It can grow on hydrogen, oxygen, carbon dioxide, and mineral salts. The complex metabolic machinery needed for *A. aeolicus* to function as a chemolithoautotroph (an organism which uses an inorganic carbon source for biosynthesis and an inorganic chemical energy source) is encoded within a genome that is only one-third the size of the *E. coli* genome. Metabolic flexibility seems to be reduced as a result of the limited genome size. The use of oxygen (albeit at very low concentrations) as an electron acceptor is allowed by the presence of a complex respiratory apparatus. Although this organism grows at 95 °C, the extreme thermal limit of the Bacteria, only a few specific indications of thermophily are apparent from the genome. Here we describe the complete genome sequence of 1,551,335 base pairs of this evolutionarily and physiologically interesting organism.**

Complete genome sequences have been determined for a number of organisms, including Archaea[1], Bacteria[2–7], and Eukarya[8]. Here we present and explore the genome sequence of *Aquifex aeolicus*. With growth-temperature maxima near 95 °C, *Aquifex pyrophilus* and *A. aeolicus* are the most thermophilic bacteria known. Although isolated and described only recently[9], these species are related to filamentous bacteria first observed at the turn of the century, growing at 89 °C in the outflow of hot springs in Yellowstone National Park[10,11]. The observation of these macroscopic assemblages would later be instrumental in the drive to culture hyperthermophilic organisms[12].

The *Aquificaceae* represent the most deeply branching family within the bacterial domain on the basis of phylogenetic analysis of 16S ribosomal RNA sequences[13,14], although analyses of individual protein sequences vary in their placement of *Aquifex* relative to other groups[15–18]. The genera in this group, *Aquifex* and *Hydrogenobacter*, are thermophilic, hydrogen-oxidizing, microaerophilic, obligate chemolithoautotrophs[9,19–21]. *A. aeolicus* (isolated by R.H. and K. O. Stetter) was cultured at 85 °C under an $H_2/CO_2/O_2$ (79.5:19.5:1.0) atmosphere in a medium containing only inorganic components. *A. aeolicus* does not grow on a number of organic substrates, including sugars, amino acids, yeast extract or meat extract. Unlike its close relative *A. pyrophilus*, *A. aeolicus* has not been shown to grow anaerobically with nitrate as an electron acceptor in the laboratory.

From study of the physiology of the organism, several predictions can be made. As an autotroph, *A. aeolicus* must have genes encoding proteins for one or more modes of carbon fixation and a complete set of biosynthetic genes. As autotrophy is a feature that is distributed throughout the Archaea and Bacteria, most of the associated genes are expected to be of ancient origin and clearly related to those characterized elsewhere. The obligate autotrophy suggests a biosynthetic rather than a degradative character. Oxygen respiration implies the presence of corresponding utilization and tolerance genes. The early divergence of the *Aquificaceae* inferred from ribosomal RNA sequences leads to several questions. Are the machineries for oxygen usage and tolerance homologous to those found in mitochondria and well studied organisms such as *Escherichia coli*, or were they invented separately? If there was far less oxygen when the lineage originated, is there evidence for use of alternative oxidants?

## Genome

General features of the *A. aeolicus* genome are listed in Box 1. We classified 1,512 open-reading frames (ORFs) into one of three categories, namely, identified (Table 1), hypothetical, or unknown. Identified ORFs were further classified into one of 57 cellular role categories adapted from Riley[22] (Table 1). The relatively high G + C content of the two 16S-23S-5S rRNA operons (65%) is characteristic of thermophilic bacterial rRNAs[23]. The genome is densely packed: most genes are apparently expressed in polycistronic operons and many convergently transcribed genes overlap slightly. Nonetheless, many genes that are functionally grouped within operons in other organisms, such as the tryptophan or histidine biosynthesis pathways, are found dispersed throughout the *A. aeolicus* genome or appear in novel operons. Even when they encode subunits of the same enzyme, the genes are often separated on the chromosome (for example, *glt*B and *glt*D, the genes encoding the large and small subunits of glutamate synthase). Operon organization of genes for the biosynthesis of amino acids is found in both Archaea and Bacteria but it is not universal in either group. *A. aeolicus* is extreme in that no two amino acid biosynthetic genes are found in the same operon. In contrast, genes required for electron transport, hydrogenase subunits, transport systems, ribosomal subunits, and flagella are often in functionally related operons in *A. aeolicus* (Fig. 1). No introns or inteins (protein splicing elements) were detected in the genome.

A single extrachromosomal element (ECE) was identified during sequencing. Sequence redundancy for the total project was calculated to be 4.83. The ECE, however, is significantly over-represented

† Present addresses: Codex Bioinformatics Services, PO Box 90273, San Diego, California 92169, USA (G.D.); Department of Bioinformatics, SmithKline Beecham Pharmaceuticals, Collegeville, Philadelphia 19426, USA (P.V.W.)


Nature © Macmillan Publishers Ltd 1998

relative to the chromosome; when calculated independently for the final assemblies, redundancies are 4.73 and 8.76 for the chromosome and for the ECE, respectively. The ECE therefore appears to be present at roughly twice the copy number of the chromosome. Although no ORFs on the ECE can be assigned a function with confidence, except for a transposase, two of the predicted proteins show similarity to hypothetical proteins in the *Methanococcus jannaschii* genome[1]. One ORF on the ECE is also present in two identical copies on the *A. aeolicus* chromosome, providing evidence of genetic exchange between the chromosome and the ECE.

### Reductive tricarboxylic acid cycle

As an autotroph, *A. aeolicus* obtains all necessary carbon by fixing $CO_2$ from the environment. An assay for activity of the reductive tricarboxylic acid (TCA) cycle in *A. pyrophilus* cell extracts showed *in vitro* activities for each proposed reaction[24]. The reductive (reverse) TCA cycle fixes two molecules of $CO_2$ to form acetyl-coenzyme A (acetyl-CoA) and other biosynthetic intermediates[25]. The *A. aeolicus* genome contains genes encoding malate dehydrogenase, fumarate hydratase, fumarate reductase, succinate-CoA ligase, ferredoxin oxidoreductase, isocitrate dehydrogenase, aconitase and citrate synthase, which together could constitute the TCA pathway. There is no biochemical evidence for alternative carbon-fixation pathways in *A. pyrophilus*[24,25] nor is there sequence evidence for such pathways in *A. aeolicus*.

The TCA cycle is vital as it provides the substrates of many biosynthetic pathways. (It is beyond the scope of this report to detail these biosynthetic pathways, but they seem to be typically bacterial, and candidate genes for all or most of the enzymes have been identified in *A. aeolicus*.) The central role of the TCA cycle is emphasized by duplication of many of its constituent genes in *A. aeolicus*. Two genes encode proteins that are similar to malate dehydrogenase (in addition to a lactate dehydrogenase). The fumarate hydratase is split into amino- and carboxy-terminal subunits, as is the case in *M. jannaschii*[1]. Unlinked genes encoding two iron–sulphur proteins of fumarate reductase (alternatively succinate dehydrogenase) accompany a single flavoprotein subunit. Two sets of genes resembling succinate-CoA ligase (both the α- and β-subunits) are present. *A. aeolicus* has two putative operons encoding four-subunit (α, β, γ, δ) 2-acid ferredoxin oxidoreductases; members of this family catalyze reversible carboxylation/decarboxylation of pyruvate, 2-isoketovalerate, or 2-oxoglutarate with varying specificity[26]. These duplicated genes may encode paralogous proteins with unique substrate specificity, as opposed to redundant functions. For example, a paralogue of succinate-CoA ligase may activate citrate with coenzyme A to form citryl-CoA, which citrate synthase can cleave to produce oxaloacetate and acetyl-CoA.

### Gluconeogenesis through the Embden–Meyerhof–Parnas pathway

Growing autotrophically, *A. aeolicus* must synthesize pentose and hexose monosaccharides from products of the reductive TCA cycle. Pyruvate produced by pyruvate ferredoxin oxidoreductase or by pyruvate carboxylase (oxaloacetate decarboxylase)[24] may enter the Embden–Meyerhof–Parnas pathway of glycolysis and gluconeogenesis. Genes encoding fructose-1,6-bisphosphatase, an essential gluconeogenic enzyme in *E. coli*, have not been identified in the genomes of the autotrophs *A. aeolicus* or *M. jannaschii*[1], suggesting that an unidentified pathway may exist. The *A. aeolicus* genome also encodes enzymes of the pentose-phosphate pathway and enzymes for glycogen synthesis and catabolism. We found neither (phospho) gluconate dehydrase nor 2-keto-3-deoxy-(6-phospho)gluconate aldolase of the Entner–Doudoroff pathway.

### Respiration

*Aquifex* species are able to grow by using oxygen concentrations as low as 7.5 p.p.m. (R.H. and K. O. Stetter, unpublished observations). The enzymes for oxygen respiration are similar to those of other bacteria: ubiquinol cytochrome *c* oxidoreductase ($bc_1$ complex), cytochrome *c* (three different genes) and cytochrome *c* oxidase (with two different subunit I genes and two different subunit II genes). The alternative system, with cytochrome *bd* ubiquinol oxidase, is also present. Clearly, the *Aquifex* lineage did not independently invent oxygen respiration. This leaves at least three possibilities: consistent with the ability of *Aquifex* to use very low levels of oxygen, the oxygen-respiration system was highly developed when oxygen had only a small fraction of its present concentration before the advent of oxygenic photosynthesis; contrary to what is implied by the 16S phylogeny, the lineage including *Aquifex* originated after the rise in atmospheric oxygen; or oxygen respiration developed once, and was then laterally transferred among bacterial lineages and acquired by *Aquifex*.

Many other oxidoreductases are present in addition to those obviously involved in oxygen respiration. The physiological role of most of these oxidoreductases is unknown or ambiguous, but two deserve comment. There is a putative nitrate reductase in the genome, although *A. aeolicus* has not been observed to perform $NO_3^-$ respiration, unlike the closely related *A. pyrophilus*. The nitrate reductase gene is adjacent to a nitrate transporter, and may be involved in nitrogen assimilation rather than respiration. It is also possible that *A. aeolicus* has a latent ability to respire with nitrate but that the conditions required have not been found. Two gene sequences show strong similarities to Rieske proteins, even though the rest of the ubiquinol cytochrome *c* oxidoreductase subunits appear only once in the genome. One of these Rieske protein genes is adjacent to a sulphide dehydrogenase subunit, suggesting a role in sulphur respiration.

### Oxidative stress

*A. aeolicus* grows optimally under microaerophilic conditions and consequently possesses various protective enzymes to counter reactive oxygen species, particularly superoxide and peroxide. The genome contains three genes encoding superoxide dismutases, two of the copper/zinc family and one of the iron/manganese family. The latter has also been noted in *A. pyrophilus*[27]. One of the copper/zinc superoxide dismutase genes is located in a large gene cluster encoding formate dehydrogenase.

No catalase genes were identified. There are several genes in the genome that might encode proteins that catalyze the detoxification of $H_2O_2$, including cytochrome *c* peroxidase, thiol peroxidase, and two alkyl hydroperoxide reductase genes. All of these enzymes require an exogenous reductant and therefore do not evolve $O_2$. However, treatment of *A. pyrophilus*[9] or *A. aeolicus* biomass with $H_2O_2$ results in the rapid evolution of gas bubbles. This catalase activity may result from a novel enzyme that cannot yet be identified by sequence similarity.

### Motility

Like *A. pyrophilus*[9], *A. aeolicus* is motile and possesses monopolar polytrichous flagella. More than 25 genes encoding proteins involved in flagellar structure and biosynthesis have been identified in *A. aeolicus* (Box 1). However, no homologues of the bacterial chemotaxis system were identified. In enteric bacteria, membrane-bound receptors bind chemoattractants and repellents and mod-

**Figure 1** Linear map of the *A. aeolicus* circular chromosome. Genes are shown as arrows which denote the direction of transcription and are coloured to denote functional categorization according to the key below the figure. The sequences of the two rRNA gene clusters are identical. Here, the first base of the coding sequence of *fusA* was arbitrarily assigned as base number 1 as no origin of replication has been identified. ORF numbers are discontinuous because some ORFs representing 100 amino acids or more are not predicted to be coding and are not shown.


Nature © Macmillan Publishers Ltd 1998

ulate the activity of the histidine kinase CheA[28]. Phosphoryl groups from CheA are transferred to CheY, which then binds to the flagellar switch, altering the direction of flagellar rotation. Homologous chemotaxis systems are present in the archaea *Halobacterium salinarum*[29] and *Pyrococcus* sp. OT3 (H. Sizuya, personal communication), although the bacterial and archaeal flagellar apparatuses are not homologous[30]. The *M. jannaschii* genome also lacks homologues of known genes required for chemotaxis. Thus, either motility in *A. aeolicus* and *M. jannaschii* is undirected or input for controlling taxis is mediated through another, unidentified system. The most studied chemotaxis systems respond to sugars and amino acids, although responses to other inputs (for example, metals, redox potential, and light) may also occur. In contrast to all the organisms known to possess the classical chemotactic signal-transduction pathways, both *A. aeolicus* and *M. jannaschii* are obligate chemoautotrophs. Chemoautotrophs may respond to a different set of factors, such as concentrations of dissolved gas ($CO_2$, $H_2$ or $O_2$) or another critical parameter such as temperature.

In *E. coli*, the flagellar switch is essential for flagellar structure and function and coupling of chemotaxis signals. But the *A. aeolicus* genome encodes homologues of only two of the three *E. coli* proteins that make up the switch, FliG and FliN. Biochemical[31] and genetic[32] studies implicate the missing FliM protein as the receptor for phosphorylated CheY, the switch signal. The absence of both FliM and CheY in *A. aeolicus* supports the identification of FliM as the receptor for phosphorylated CheY in *E. coli*. This result also argues against a direct role for FliM in torque generation.

## DNA replication and repair

The *A. aeolicus* primary replicative DNA polymerase, corresponding to the DNA polymerase III holoenzyme in *E. coli*, probably consists

**Box 1 *Aquifex aeolicus* genome features**

**General**

Length 1,551, 335 bp
G + C content 43.4%
Protein-coding regions 93%
Stable RNA 0.8%
Non-coding repeats (none significant)
Intergenic sequences 6.2%

**RNA**

| Ribosomal RNA | Chromosome coordinates |
|---|---|
| 16S-23S-5S | 572785-567770 |
| 16S-23S-5S | 1192069-1197084 |

Transfer RNA
44 species (7 clusters, 28 single genes)

| Other RNAs | Chromosome coordinates |
|---|---|
| tmRNA | 1153844-1153498 |

**Chromosomal coding sequences**

849 similar to protein of known function (average length 1,066 bp)
256 similar to protein of unknown function (average length 898 bp)
407 unknown coding regions (average length 762 bp)
1,512 total (average length 956)

**Extrachromosmal element (ECE)**

Length 39,456 bp
G + C content 36.4%
Protein-coding regions 53.5%

**ECE-coding sequences**

1 similar to proteins of known function (length 948 bp)
4 similar to proteins of unknown function (average length 667 bp)
27 unknown coding regions (average length 648 bp)

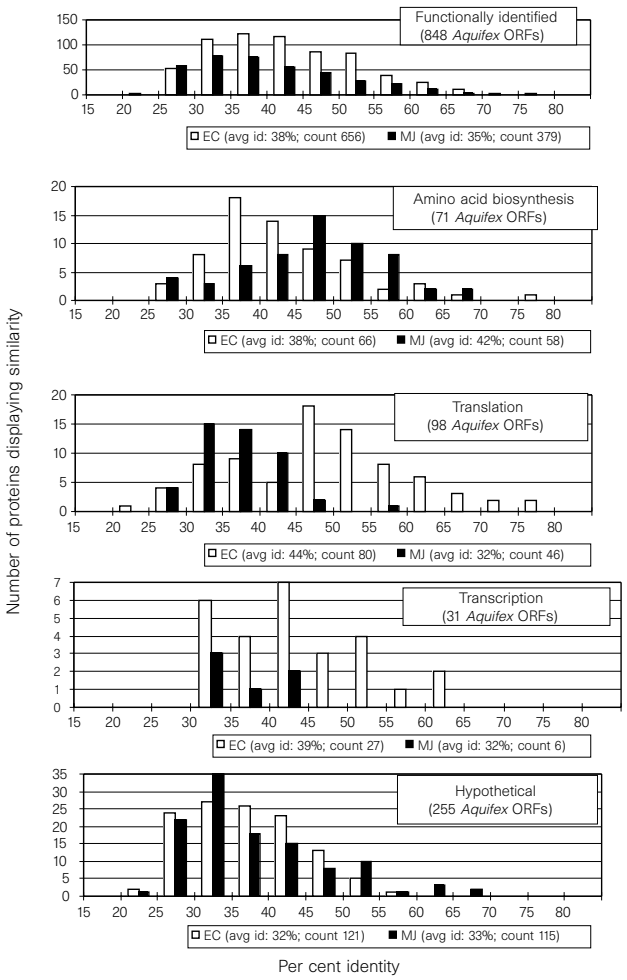


**Figure 2** Histogram representation of the similarity of selected classes of predicted proteins to predicted proteins from the *E. coli* (EC) and *M. jannaschii* (MJ) genomes. Predicted *A. aeolicus* proteins representing each category were independently compared to sets of all potential polypeptides (≥100 amino acids) from the two genomes using FASTA[44]. If the top scoring alignment covered ≥80% of the length of the *A. aeolicus* protein, the score was plotted. There were more positives found in the *E. coli* genome in nearly every category. Hypothetical proteins (those identified by database match but of unknown function) are very similarly represented by *M. jannaschii* and *E. coli.* There are a small number of very highly conserved hypotheticals that are shared between *A. aeolicus* and *M. jannaschii*. Generally, biosynthetic categories show less discrimination than information-processing categories, which are clearly more *E. coli*-like. The variation in the apparent rates of evolution in different categories suggests that different phylogenies may be inferred depending on the sequence analysed. Within each graph, correspondence to *E. coli* is shown in white and *M. jannaschii* is shown in black. Avg id, average identity; count, number of proteins analysed.

Nature © Macmillan Publishers Ltd 1998



Nature © Macmillan Publishers Ltd 1998

of a core structure containing α- and ε-subunits, a γ-τ-subunit and an additional member of the γ-τ/δ′-family. A gene encoding a protein homologous to the β-sliding clamp was also found. This minimalistic complex lacks homologous θ-, δ-, χ- and ψ-subunits, as does the *Mycoplasma genitalium* holoenzyme[3]. Translation of the 54K (relative molecular mass) γ-τ-ATPase subunit may proceed without a programmed frameshift to produce a protein similar to the N-terminal region of the *E. coli* γ-subunit. DNA polymerase I is present as separate Klenow fragment and 5′ → 3′ exonuclease subunits, encoded by two non-adjacent ORFs. Although the repair polymerase, DNA polymerase II, has not been found in *A. aeolicus*, one ORF (Aq1422) encodes a protein similar to the eukaryotic DNA repair polymerase-β. A member of the same family has been identified in *Thermus aquaticus*[33] and *Bacillus subtilis*.

## Transcriptional and translational apparatuses

The transcriptional apparatus of *A. aeolicus* is similar to that of *E. coli* and lacks any components specific to the Eukarya or Archaea (Fig. 2). In addition to the core RNA polymerase α-, β-, and β′-subunits, four σ-factors which determine promoter specificity are present (Table 1). Several different families of bacterial transcriptional regulators were also identified, including two-component systems. All of the ribosomal proteins and elongation factors common to other bacteria are present, indicating that all bacteria-specific ribosomal proteins were present in the common ancestor of *Aquifex* and other bacteria. Also present are the four *sel* genes required for the cotranslational incorporation of selenocysteine. These latter genes are clustered in a 15-kilobase-pair segment that also encodes the biosynthetic and structural proteins for formate dehydrogenase, the only selenocysteine-containing protein identified. The gene that encodes selenocysteine transfer RNA, *selC*, is apparently cotranscribed with the genes encoding the formate dehydrogenase structural proteins.

*A. aeolicus* lacks glutaminyl-tRNA and asparaginyl-tRNA synthetases. The genes required for transamidation of glutamyl-tRNA$^{Gln}$ are present[34]. Charging of asparaginyl-tRNA is likely to proceed through the analogous reaction, as shown in halobacteria[35], although the genes(s) for that transamidase are unknown. The canonical methionyl- and leucyl-tRNA synthetases have only been seen previously as single polypeptide enzymes; however, in *A. aeolicus* the homologues appear fragmented into two subunits. In both cases, the genes that encode the N- and C-terminal portions are widely separated on the chromosome. No complete three-dimensional structural data are available for either methionyl- or leucyl-aminoacyl tRNA synthetases, but the subunit organization in the *A. aeolicus* aminoacyl-tRNA synthetases may reflect domain organization in the homologous proteins.

## Thermophily

The *A. aeolicus* genome is the second completely sequenced genome of a hyperthermophile. By comparing the *A. aeolicus* and *M. jannaschii* genomes and contrasting them with the complete genomes of mesophiles, we can discover whether there are aspects of the genome or the encoded information that are diagnostic of hyperthermophiles. The G + C content of the stable RNAs is clearly indicative of the high growth temperature of the organism. This property can be used to identify stable RNAs against the relatively low G + C background of the *A. aeolicus* genome. The gene encoding tmRNA (or 10Sa RNA)[36], an RNA involved in tagging polypeptides translated from incomplete messenger RNAs for degradation, was located in this way.

Two genes for reverse gyrase are present in the genome. This is the only protein known to be present only in thermophiles. Other proteins, currently described as hypotheticals, may be diagnostic of hyperthermophiles but the data sets are not yet large enough to decide this with confidence.

Although features of stabilization may not be apparent in any given protein[37], a large enough data set may reveal general trends in amino-acid usage that are informative. Particularly important in this regard is inclusion of multiple genomes of hyperthermophiles so as not to allow the idiosyncracies of a single organism to bias the conclusions. As shown in Table 2, comparison of the amino-acid composition encoded by six genomes shows that use of individual amino acids can vary significantly from genome to genome. The data suggest trends that may be correlated with the thermostability of the encoded proteins. One apparent trend is that the hyperthemophile genomes encode higher levels of charged amino acids on average than mesophile genomes[38], primarily at the expense of uncharged polar residues. Glutamine in particular seems to be significantly discriminated against in the hyperthermophiles. Although this observation might be rationalized on the basis of

**Table 2 Comparison of relative amino acid compositions (in percentages) of mesophiles and thermophiles**

| | Mesophiles | | | | Thermophiles | |
|---|---|---|---|---|---|---|
| *Amino acid* | *H. influenzae* | *H. pylori* | *E. coli* | *Synechosystis* | *A. aeolicus* | *M. jannaschii* |
| A | 8.21 | 6.83 | 9.55 | 9.07 | 5.90 | 5.54 |
| C | 1.03 | 1.09 | 1.11 | 1.01 | 0.79 | 1.27 |
| D | 4.98 | 4.77 | 5.20 | 5.07 | 4.32 | 5.52 |
| E | 6.48 | 6.88 | 5.91 | 6.20 | 9.63 | 8.67 |
| F | 4.46 | 5.41 | 3.87 | 3.75 | 5.13 | 4.20 |
| G | 6.65 | 5.76 | 7.42 | 7.77 | 6.75 | 6.41 |
| H | 2.05 | 2.12 | 2.26 | 1.93 | 1.54 | 1.43 |
| I | 7.10 | 7.20 | 5.95 | 6.31 | 7.32 | 10.45 |
| K | 6.32 | 8.94 | 4.48 | 4.26 | 9.40 | 10.36 |
| L | 10.50 | 11.18 | 10.56 | 10.93 | 10.57 | 9.38 |
| M | 2.44 | 2.28 | 2.86 | 2.12 | 1.92 | 2.33 |
| N | 4.89 | 5.83 | 3.88 | 3.76 | 3.60 | 5.24 |
| P | 3.72 | 3.28 | 4.41 | 5.09 | 4.07 | 3.38 |
| Q | 4.64 | 3.70 | 4.42 | 5.26 | 2.04 | 1.44 |
| R | 4.47 | 3.46 | 5.58 | 5.18 | 4.91 | 3.85 |
| S | 5.84 | 6.81 | 5.67 | 5.46 | 4.79 | 4.46 |
| T | 5.20 | 4.37 | 5.35 | 5.53 | 4.21 | 4.06 |
| V | 6.68 | 5.59 | 7.11 | 7.10 | 7.93 | 6.85 |
| W | 1.12 | 0.70 | 1.48 | 1.30 | 0.93 | 0.71 |
| Y | 3.12 | 3.68 | 2.83 | 2.78 | 4.13 | 4.33 |

| | Mesophiles | Thermophiles |
|---|---|---|
| Charged residues (DEKRH) | 24.11 | 29.84 |
| Polar/uncharged residues (GSTNQYC) | 31.15 | 26.79 |
| Hydrophobic residues (LMIVWPAF) | 44.74 | 43.36 |

Nature © Macmillan Publishers Ltd 1998

an increased rate of deamidation of this residue at higher temperatures, aspargine does not appear subject to similar discrimination.

## Phylogeny

The placement of the *Aquifex* lineage as one of the earliest divergences in the eubacterial tree[13,14] is interesting because of the insights it could provide into the ancestral eubacterial phenotype, including the hypothesized thermophilic nature of the first bacteria. Protein-based phylogenies often do not support the original rRNA-based placement[15,16,18]. Thus, the availability of some 1,500 genes from an *Aquifex* species would seem to offer a definitive resolution of the phylogeny. However, our analyses of ribosomal proteins, amino-acyl-tRNA synthetases, and other proteins do not do so, showing no consistent picture of the organism's phylogeny. We cannot make a more complete analysis and discussion here, but some observations can be made. These proteins do not yield a statistically significant placement of the *Aquifex* lineage or of other major eubacterial lineages. This situation partially reflects the inadequacy of some protein sequences as indicators of distant molecular genealogy because of their particular evolutionary dynamic, including the patterns and rates of amino-acid replacements. In some cases (such as the aminoacyl-tRNA synthetases for arginine, cysteine, histidine, proline and tyrosine), the analyses are further complicated by the presence of paralogous genes and/or apparent lateral gene transfers. It seems that a more extensive survey of genes and a better sampling of major eubacterial taxa will be required to confidently confirm or refute an early divergence of the *Aquifex* lineage.

## Conclusions

Advances in sequencing techniques have allowed us to move beyond studies of single genes to studies of complete genomes only recently[2]. This rapid advance has created the opportunity to begin to characterize an organism with the full knowledge of the genome in hand. The complete genome summarized in this report represents our first view of *A. aeolicus*. The challenge now is to ask specific questions in ways which take advantage of the whole-genome data.

Beyond studies of any single organism in isolation, complete genomes allow comprehensive comparisons between organisms. For instance, comparisons of the similarity of genes can be made that reveal that genes in different categories vary in their relative conservation (Fig. 2). In addition, genome-wide trends are apparent. For example, why is there not more of a tendency to group functionally related genes (for example, biosynthetic pathways) into operons in *A. aeolicus*? This was also seen in the genome sequence of the autotroph *M. jannaschii*[1]. Is this because the autotrophic lifestyle decreases the need for selective regulation? There also seem to be a few multifunctional, fused proteins in *A. aeolicus* and *M. jannaschii*. Although this seems unlikely to be related to autotrophy, it might be associated with extreme thermophily. The large number of diverse genome sequences that will become available in the coming years will allow more detailed correlation of global genomic properties with particular physiologies. □

### Methods

**Sequencing strategy.** The sequencing strategy used to assemble the complete genome was based on the whole genome random (or 'shotgun') approach, which has been successfully used for other genomes of similar size[1–4]. Shotgun sequencing projects are characterized by two phases: an initial completely random phase in which the bulk of the data is collected, followed by a closure phase where directed techniques are used to close gaps and complete the assembly. By pursuing a strategy where only 97% coverage was initially achieved, we were able to limit the number of sequences needed for the random phase to only 10,500 (ref. 39).

Sequences were generated from a small insert library constructed in λ ZAP II vectors[40,41] (average insert length 2.9 kilobase pairs). Two different methods were used for sequencing: first, dye-primer M13-21 and M13 reverse primer ABI Prism CS$^+$ ready reaction kits, analysed on 48-cm 4% polyacrylamide gels; and second, dye-terminator (ABI Prism FS+) reactions using two pBluescript-specific primers. These reactions were analysed on 36-cm 5% Long-Ranger gels.

The sequence fragments were assembled on an Apple Power Macintosh computer using Sequencher (Gene Codes, Ann Arbor, MI), an assembly and editing program. Assembly was typically performed in batches of roughly 200–400 sequences, and was followed by inspection and editing of the assemblies. All sequences in the set were compared with all others through this process. After assembly, the sequences comprised ~750 contigs at the end of the random phase. Sequences were obtained from both ends of ~200 randomly chosen clones from a fosmid library[42,43]. These sequences were then assembled with consensus sequences derived from the contigs of random-phase sequences using Sequencher. Gaps between contigs were closed by direct sequencing on fosmids not wholly contained within a contig. The fosmid library thus served a purpose analogous to that of the λ-scaffold in other projects[1–4]. The final eight gaps were closed by direct sequencing of polymerase chain reaction (PCR) products generated with the TaqPlus Long PCR System (Stratagene Cloning Systems, La Jolla, CA).

Consequences of reducing the number of sequences in the random phase are the large number of gaps that remain to be closed in the directed phase, and the reduction in overall coverage. To ensure that reduced coverage did not compromise accuracy, ~200 oligonucleotide primers were synthesized to resequence regions of ambiguity identified by visual inspection of the entire assembly. 13,785 sequences, with an average edited read length of 557 base pairs, constitute the final assembly. On the basis of a relatively small number of errors identified during the annotation process, we estimate the error frequency to be <0.01%, comparable to other published genomic sequence estimates.

**Gene (ORF + RNA) identification and functional assignment approaches.** Coding regions of the *A. aeolicus* genome were analysed and assigned using primarily the programs BLASTP[44] and FASTA[45] to search against a non-redundant protein database. Many analyses were carried out within the context of MAGPIE[46,47], an integrated computing environment for genome analysis. The results of these analyses are available for user interpretation, validation, and categorization. Additional ORFs were identified and start sites refined using the program CRITICA (J. H. Badger and G.J.O., unpublished program). Finally, all presumed 'intergenic regions' were examined with BLASTX for similarities to known protein sequences[48]. Transfer RNA genes were identified with the program tRNAscan-SE[49].

Received 26 August 1997; accepted 3 February 1998.


1. Bult, C. *et al.* Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273,** 1058–1073 (1996).
2. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269,** 496–511 (1995).
3. Fraser, C. M. *et al.* The minimal gene complement of *Mycoplasma genitalium*. *Science* **270,** 397–403 (1995).
4. Tomb, J.-F. *et al.* The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388,** 539–547 (1997).
5. Himmelreich, R. *et al.* Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* **24,** 4420–4449 (1996).
6. Kaneko, T. *et al.* Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC7803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3,** 109–136 (1996).
7. Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277,** 1453–1462 (1997).
8. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274,** 546 (1996).
9. Huber, R. *et al. Aquifex pyrophilus* gen. nov. sp. nov. represents a novel group of marine hyperthermophilic hydrogen oxidizing bacteria. *Arch. Micrtobiol.* **15,** 340–351 (1992).
10. Reysenbach, L., Wickham, G. S. & Pace, N. R. Phylogenetic analysis of the hyperthermophilic pink filament community in Octopus Spring, Yellowstone National Park. *Appl. Environ. Microbiol.* **60,** 2113–2119 (1994).
11. Setchell, W. A. The upper temperature limits of life. *Science* **17,** 934–937 (1903).
12. Brock, T. D. The road to Yellowstone—and beyond. *Annu. Rev. Microbiol.* **49,** 1–28 (1995).
13. Burggraf, S., Olsen, G. J., Stetter, K. O. & Woese, C. R. A phylogenetic analysis of *Aquifex pyrophilus*. *Syst. Appl. Microbiol.* **15,** 353–356 (1992).
14. Pitulle, C. *et al.* Phylogenetic position of the genus *Hydrogenobacter*. *Int. J. Syst. Bacteriol.* **44,** 620–626 (1994).
15. Baldauf, S. L., Palmer, J. D. & Doolittle, W. F. The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proc. Natl Acad. Sci. USA* **93,** 7749–7754 (1996).
16. Klenk, H.-P., Palm, P. & Zillig, W. in *Molecular Biology of the Archaea* (eds Pfeifer, F., Palm, P. & Scleeifer, K. H.) 139–147 (Vch Pub, 1994).
17. Bocchetta, M. *et al.* Arrangement and nucleotide sequence of the gene (*fus*) encoding elongation factor G (EF-G) from the hyperthermophilic bacterium *Aquifex pyrophilus*: phylogenetic depth of hyperthermophilic bacteria inferred from analysis of the EF-G/*fus* sequences. *J. Mol. Evol.* **41,** 803–812 (1995).
18. Wetmur, J. G. *et al.* Cloning, sequencing, and expression of RecA proteins from three distantly related thermophilic eubacteria. *J. Biol. Chem.* **269,** 25928–25935 (1994).
19. Kawasumi, T., Igarashi, Y., Kodama, T. & Minoda, Y. *Hydrogenobacter thermophilus* gen. nov., sp. nov.



Nature © Macmillan Publishers Ltd 1998

an extremely thermophilic, aerobic, hydrogen-oxidizing bacterium. *Int. J. Syst. Bacteriol.* **34,** 5–10 (1984).
20. Kristjannson, J., Ingason, A. & Alfredsson, G. A. Isolation of thermophilic obligately autotrophic hydrogen-oxidizing bacteria, similar to *Hydrogenobacter thermophilus*, from Icelandic hotsprings. *Arch. Microbiol.* **140,** 321–325 (1985).
21. Kryukov, V. R., Savel'eva, N. D. & Pusheva, M. A. *Calderobacterium hydrogenophilum* gen. nov., sp. nov. an extreme thermophilic bacterium and its hydrogenase activity. *Microbiology* (Engl. Trans. *Mikrobiologiya*) **52,** 611–618 (1983).
22. Riley, M. Functions of the gene products of *Escherichia coli*. *Microbiol. Rev.* **57,** 862–952 (1993).
23. Weisburg, W. G., Giovannoni, S. J. & Woese, C. R. The *Deinococcus-Thermus* phylum and the effect of rRNA composition on phylogenetic tree construction. *Syst. Appl. Microbiol.* **11,** 128–134 (1989).
24. Beh, M., Strauss, G., Huber, R., Stetter, K. O. & Fuchs, G. Enzymes of the reductive citric acid cycle in the autotrophic eubacterium *Aquifex pyrophilus* and in the archaebacterium *Thermoproteus neutrophilus*. *Arch. Microbiol.* **160,** 306–311 (1993).
25. Fuchs, G. in *Autotrophic Bacteria* (eds Schegel, H. G & Bowein, B.) 365–382 (Springer, New York, 1987).
26. Mai, X. & Adams, M. W. Characterization of a fourth type of 2-keto acid-oxidizing enzyme from a hyperthermophilic archaeon: 2-ketoglutarate ferredoxin oxidoreductase from *Thermococcus litoralis*. *J. Bacteriol.* **178,** 5890–5896 (1996).
27. Lim, J. H. *et al.* Cloning and expression of superoxide dismutase from *Aquifex pyrophilus*, a hyperthermophilic bacterium. *FEBS Lett.* **406,** 142–146 (1997).
28. Bourret, R. B., Borkovich, K. A. & Simon, M. I. Signal transduction pathways involving protein phosphorylation in prokaryotes. *Annu. Rev. Biochem.* **60,** 401–441 (1991).
29. Rudolph, J., Tolliday, N., Schmitt, C., Schuster, S. C. & Oesterhelt, D. Phosphorylation in halobacterial signal transduction. *EMBO J.* **14,** 4249–4257 (1995).
30. Jarrell, K. F., Bayley, D. P. & Kostyukova, A. S. The archaeal flagellum: a unique motility structure. *J. Bacteriol.* **178,** 5057–5064 (1996).
31. Welch, M., Oosawa, K., Aizawa, S. I. & Eisenbach, M. Effects of phosphorylation, $Mg^{2+}$, and conformation of the chemotaxis protein CheY on its binding to the flagellar switch protein FliM. *Biochemistry* **33,** 10470–10467 (1994).
32. Sockett, H., Yamaguchi, S., Kihara, M., Irikura, V. M. & Macnab, R. M. Molecular analysis of the flagellar switch protein FliM of *Salmonella typhimurium*. *J. Bacteriol.* **174,** 793–806 (1992).
33. Motoshima, H. *et al*. Molecular cloning and nucleotide sequence of the aminopeptidase T gene of *Thermus aquaticus* YT-1 and its high-level expression in *Escherichia coli*. *Agric. Biol. Chem.* **54,** 2385–2392 (1990).
34. Curnow, A. W. *et al*. Glu-tRNAGln amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc. Natl Acad. Sci. USA* **94,** 11819–11826 (1997).
35. Curnow, A. W., Ibba, M. & Söll, D. tRNA-dependent asparagine formation. *Nature* **382,** 589–590 (1996).
36. Tu, G. F., Reid, G. E., Zhang, J. G., Moritz, R. L. & Simpson, R. J. C-terminal extension of truncated proteins in *Escherichia coli* with a 10Sa decapeptide. *J. Biol. Chem.* **270,** 9322–9326 (1995).
37. Böhm, G. & Jaenicke, R. Relevance of sequence statistics for the properties of extremophilic proteins. *Int. J. Pept. Protein Res.* **43,** 97–106 (1994).
38. Choi, I.-G. *et al.* Random sequence analysis of genomic DNA of a hyperthermophile: *Aquifex pyrophilus*. *Extremophiles* **1,** 125–134 (1997).
39. Lander, E. S. & Waterman, M. S. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2,** 231–239 (1988).
40. Short, J. M., Fernandez, J. M., Sorge, J. A. & Huse, W. D. Lambda ZAP: a bacteriophage lambda expression vector with *in vivo* excision properties. *Nucleic Acids Res.* **16,** 7583–7600 (1988).
41. Alting-Mees, M. A. & Short, J. M. pBluescript II: gene mapping vectors. *Nucleic Acids Res.* **17,** 9494 (1989).
42. Shizuya, H. *et al.* Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl Acad. Sci. USA* **89,** 8794–8797 (1992).
43. Kim, U.-J., Shizuya, H., de Jong, P. J., Birren, B. & Simon, M. I. Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids Res.* **20,** 1083–1085 (1992).
44. Altschul, S. F., Fish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
45. Pearson, W. R. & Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA* **85,** 2444–2448 (1988).
46. Gaasterland, T. & Sensen, C. W. MAGPIE: automated genome interpretation. *Trends Genet.* **12,** 76–78 (1996).
47. Gaasterland, T. & Sensen, C. W. Fully automated genome analysis that reflects user needs and preferences. A detailed introduction to the MAGPIE system architecture. *Biochimie* **78,** 302–310 (1996).
48. Gish, W. & States, D. J. Identification of protein coding regions by database similarity search. *Nature Genet.* **3,** 266–272 (1993).
49. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25,** 955–964 (1997).

**Acknowledgements.** This work was supported in part by Department of Energy Microbial Genome Program grants (to R.V.S., C. R. Woese and G.J.O.). We thank C. Woese for his cooperation in the analysis of the genome and interest in the project; K. Stetter for continuing interest; G. Frey, J. Holaska, S. Peralta, D. Hafenbrandl, S. Delk, T. Robinson, and J. Arnett for technical assistance; and D. Robertson, J. Stein, I. Sanyal, T. Richardson, G. Hauska, and K. Williams for discussions.

Correspondence should be addressed to R.V.S. (e-mail: rswanson@diversa.com). Requests for *Aquifex aeolicus* should be addressed to R.H. (e-mail: Robert.huber@biologie.uni-regensburg.de). The sequences have been deposited with GenBank and assigned accession numbers AE000657 (chromosome) and AE000667 (extrachromosomal element).

Nature © Macmillan Publishers Ltd 1998



Nature © Macmillan Publishers Ltd 1998


Nature © Macmillan Publishers Ltd 1998

Nature © Macmillan Publishers Ltd 1998

**Table 1 *Aquifex aeolicus* Open Reading Frame Identifications. Gene numbers (Aq) correspond to those in Fig.1. Percentages refer to the identity found in the best FASTA alignment. The percentage of the sequence covered by the alignment is displayed with bullets as follows 20—40% ·, 40—60% ··, 60—80% ···, 80—100% ····**

| | | | |
|---|---|---|---|
| **Amino Acid Biosynthesis** | | | |
| Aromatic amino acids | | | |
| Aq1536 | aroA | 5-enolpyruvylshikimate-3-phosphate synthetase | 43.0% •••• |
| Aq081 | aroC | chorismate synthase | 55.2% •••• |
| Aq021 | aroD | 3-dehydroquinate dehydratase | 33.3% •••• |
| Aq901 | aroE | shikimate 5-dehydrogenase | 46.1% •••• |
| Aq2177 | aroK | shikimate kinase | 36.5% •••• |
| Aq951 | pheA | chorismate mutase/prephenate dehydratase | 44.0% •••• |
| Aq1548 | trpA | tryptophan synthase alpha subunit | 44.5% •••• |
| Aq706 | trpB1 | tryptophan synthase beta subunit | 68.0% •••• |
| Aq1410 | trpB2 | tryptophan synthase beta subunit | 50.0% •••• |
| Aq1787 | trpC | indole-3-glycerol phosphate synthase | 43.3% •••• |
| Aq196 | trpD1 | phosphoribosylanthranilate transferase | 45.1% •••• |
| Aq209 | trpD2 | phosphoribosylanthranilate transferase | 24.9% •••• |
| Aq582 | trpE | anthranilate synthase component I | 50.0% •••• |
| Aq2076 | trpF | phosphoribosyl anthranilate isomerase | 45.6% •••• |
| Aq549 | trpG | anthranilate synthase component II | 59.2% •••• |
| Aq1755 | tyrA | prephenate dehydrogenase | 36.1% •••• |
| Aspartate family | | | |
| Aq1866 | asd | aspartate-semialdehyde dehydrogenase | 54.6% •••• |
| Aq1969 | aspC1 | aspartate aminotransferase | 53.5% •••• |
| Aq2094 | aspC2 | aminotransferase (AspC family) | 55.4% •••• |
| Aq421 | aspC3 | aminotransferase (AspC family) | 43.3% •••• |
| Aq273 | aspC4 | aminotransferase (AspC family) | 48.5% •••• |
| Aq1143 | dapA | dihydrodipicolinate synthase | 53.1% •••• |
| Aq916 | dapB | dihydrodipicolinate reductase | 44.2% •••• |
| Aq547 | dapE | succinyl-diaminopimelate desuccinylase | 25.8% •••• |
| Aq1838 | dapF | diaminopimelate epimerase | 35.5% •••• |
| Aq1208 | lysA | diaminopimelate decarboxylase | 47.4% •••• |
| Aq1152 | lysC | aspartokinase | 52.2% •••• |
| Aq1710 | metE | tetrahydropteroyltriglutamate methyltransferase | 45.9% •••• |
| Aq1812 | thrA | homoserine dehydrogenase | 40.4% •••• |
| Aq1309 | thrB | homoserine kinase | 38.3% •••• |
| Aq608 | thrC1 | threonine synthase | 64.3% •••• |
| Aq425 | thrC2 | threonine synthase | 61.9% •••• |
| Branched-chain family | | | |
| Aq451 | ilvB | acetolactate synthase large subunit | 53.1% •••• |
| Aq1245 | ilvC | acetohydroxy acid isomeroreductase | 64.3% •••• |
| Aq837 | ilvD | dihydroxyacid dehydratase | 58.0% •••• |
| Aq1893 | ilvE | branched-chain amino acid aminotransferase | 40.3% •••• |
| Aq1851 | ilvH | acetolactate synthase | 53.2% •••• |
| Aq356 | leuA1 | 2-isopropylmalate synthase | 52.1% •••• |
| Aq2090 | leuA2 | 2-isopropylmalate synthase | 49.9% •••• |
| Aq244 | leuB | 3-isopropylmalate dehydrogenase | 58.7% •••• |
| Aq940 | leuC | large subunit of isopropylmalate isomerase | 52.3% •••• |
| Aq1398 | leuD | 3-isopropylmalate dehydratase | 56.6% ••• |
| Glutamate family | | | |
| Aq2068 | argB | acetylglutamate kinase | 54.2% •••• |
| Aq1879 | argC | N-Acetyl-gamma-glutamylphosphate reductase | 40.6% •••• |
| Aq023 | argD | N-acetylornithine aminotransferase | 49.5% •••• |
| Aq1711 | argF | ornithine carbamoyltransferase | 46.2% •••• |
| Aq1140 | argG | argininosuccinate synthase | 54.9% •••• |
| Aq1372 | argH | argininosuccinate lyase | 46.4% •••• |
| Aq970 | argJ | glutamate N-acetyltransferase | 39.8% •••• |
| Aq111 | glnA | glutamine synthetase | 57.6% •••• |
| Aq109 | glnB | nitrogen regulatory PII protein | 73.2% •••• |
| Aq1774 | glnE | glutamate ammonia ligase adenylyl-transferase | 28.4% •••• |
| Aq1565 | gltB | glutamate synthase large subunit | 44.3% •••• |
| Aq2064 | gltD | glutamate synthase small subunit gltD | 37.7% •••• |
| Aq1071 | proA | gamma-glutamyl phosphate reductase | 47.9% •••• |
| Aq1134 | proB | glutamate 5-kinase | 43.2% •••• |
| Aq166 | proC | pyrroline carboxylate reductase | 35.1% •••• |
| Histidine | | | |
| Aq1303 | hisA | phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase | 40.9% •••• |
| Aq039 | hisB | imidazoleglycerolphosphate dehydratase | 46.4% •••• |
| Aq2084 | hisC | histidinol-phosphate aminotransferase | 33.7% •••• |
| Aq782 | hisD | histidinol dehydrogenase | 49.9% •••• |
| Aq181 | hisF | HisF (cyclase) | 59.9% •••• |
| Aq1613 | hisG | ATP phosphoribosyltransferase | 40.3% •••• |
| Aq732 | hisH | amidotransferase HisH | 47.7% •••• |
| Aq1968 | hisIE | phosphoribosyl-ATP pyrophosphohydrolase | 43.8% •••• |
| Selenocysteine | | | |
| Aq1031 | selA | L-seryl-tRNA(ser) selenium transferase | 42.7% •••• |
| Aq1030 | selD | selenophosphate synthase | 37.7% •••• |
| Serine family | | | |
| Aq1556 | cysM | cysteine synthase, O-acetylserine (thiol) lyase B | 45.8% •••• |
| Aq479 | glyA | serine hydroxymethyl transferase | 62.7% •••• |
| Aq1905 | serA | D-3-phosphoglycerate dehydrogenase | 44.1% •••• |
| **Cell Envelope** | | | |
| Pili and fimbrae | | | |
| Aq1433 | fimZ | minor pilin | 34.9% •• |
| Aq1432 | ppdD1 | pilin | 40.6% •••• |
| Aq1434 | ppdD2 | pilin | 26.4% •••• |
| Aq1435 | ppdD3 | pilin | 28.2% ••• |
| Lipoproteins and porins | | | |
| Aq270 | lgt | prolipoprotein diacylglyceryl transferase | 30.1% •••• |
| Aq819 | lnt | apolipoprotein N-acyltransferase | 25.5% •••• |
| Aq652 | nlpD1 | lipoprotein | 25.4% •••• |
| Aq1753 | nlpD2 | lipoprotein NlpD fragment | 43.2% •••• |
| Aq529 | oprC | outer membrane protein c | 27.2% •••• |
| Aq2147 | pal | peptidoglycan associated lipoprotein | 35.1% •• |
| Aq1370 | rlpA1 | rare lipoprotein A | 61.1% •• |
| Aq1174 | rlpA2 | rare lipoprotein A | 40.6% ••• |
| Aq2166 | scbA | adhesion protein | 25.7% •••• |
| Aq619 | yfeA | adhesion B precursor | 28.5% •••• |
| Peptidoglycan | | | |
| Aq1827 | alr | alanine racemase | 33.2% •••• |
| Aq1681 | amiB | N-acetylmuramoyl-L-alanine amidase | 31.0% •••• |
| Aq2195 | bacA | undecaprenol kinase | 43.1% •••• |
| Aq1798 | cphA1 | beta lactamase precursor | 25.0% ••• |
| Aq974 | cphA2 | beta lactamase precursor | 29.4% ••• |
| Aq521 | ddlA | D-alanine:D-alanine ligase | 38.2% •••• |
| Aq301 | glmS | glucosamine-fructose-6-phosphate aminotransferase | 43.2% •••• |
| Aq607 | glmU | UDP-N-acetylglucosamine pyrophosphorylase | 37.6% ••• |
| Aq053 | mraY | phospho-N-acetylmuramoyl-pentapeptide-transferase | 47.5% •••• |
| Aq624 | mrcA | penicillin binding protein 1A | 33.2% •••• |
| Aq1281 | murA | UDP-N-acetylglucosamine 1-carboxyvinyltransferase | 45.7% •••• |
| Aq520 | murB1 | UDP-N-acetylenolpyruvoylglucosamine reductase | 35.6% •••• |
| Aq511 | murB2 | UDP-N-acetylenolpyruvoylglucosamine reductase | 38.9% •••• |
| Aq1360 | murC | UDP-N-acetylmuramate-alanine ligase | 46.1% •••• |
| Aq2075 | murD | UDP-N-acetylmuramoylalanine-D-glutamate ligase | 29.3% •••• |
| Aq1747 | murE | UDP-MurNac-tripeptide synthetase | 42.9% •••• |
| Aq821 | murF | UDP-MURNAC-pentapeptide sythetase | 32.3% •••• |
| Aq1177 | murG | phospho-N-acetylmuramoyl-pentapeptide-transferase | 30.5% •••• |
| Aq325 | murI | glutamate racemase | 43.4% •••• |
| Aq1189 | pbpA1 | penicillin binding protein 2 | 32.2% •••• |
| Aq556 | pbpA2 | penicillin binding protein 2 | 30.3% •••• |
| Aq185 | tagD1 | glycerol-3-phosphate cytidyltransferase | 52.0% •••• |
| Aq1368 | tagD2 | glycerol-3-phosphate cytidyltransferase | 67.2% ••• |
| Surface polysaccharides and lipopolysaccharides | | | |
| Aq1684 | alg | alginate synthesis-related protein | 37.2% •• |
| Aq1641 | cap | capsular polysaccharide biosynthesis protein | 30.8% •••• |
| Aq1899 | dmt | dolichol-phosphate mannosyltransferase | 40.2% •••• |
| Aq1772 | envA | UDP-3-0-acyl N-acetylglcosamine deacetylase | 36.5% •••• |
| Aq1757 | exbB | biopolymer transport exbB | 48.2% •••• |
| Aq1839 | exbD | biopolymer transport ExbD | 34.7% •••• |
| Aq1069 | galE | UDP-glucose-4-epimerase | 54.7% •••• |
| Aq1705 | galF | UDP-glucose pyrophosphorylase | 47.2% •••• |
| Aq908 | gmhA | phosphoheptose isomerase | 63.4% ••• |
| Aq085 | kdsA | 3-deoxy-d-manno-octulosonic acid 8-phosphate synthase | 52.0% •••• |
| Aq326 | kdtA | 3-deoxy-D-manno-2-octulosonic acid transferase | 28.9% •••• |
| Aq253 | kdtB | lipopolysaccharide core biosynthesis protein | 46.5% •••• |
| Aq1546 | kpsF | polysialic acid capsule expression protein | 45.9% •••• |
| Aq692 | kpsU | 3-deoxy-manno-octulosonate cytidylyltransferase | 41.3% •••• |
| Aq1742 | lgtF | beta 1,4 glucosyltransferase | 35.2% •••• |
| Aq604 | lpxA | acyl-[acyl-carrier-protein]-UDP-N-acetylglucosamine acyltransferase | 47.7% •••• |
| Aq1427 | lpxB | lipid A disaccharide synthetase | 31.6% •••• |
| Aq538 | lpxD | UDP-3-O-[3-hydroxymyristoyl] glucosamine N acyltransferase | 43.3% •••• |
| Aq718 | mpg | mannose-1-phosphate guanyltransferase | 34.1% •••• |
| Aq1096 | mtfA | mannosyltransferase A | 34.3% ••• |
| Aq515 | mtfB | mannosyltransferase B | 29.0% •••• |
| Aq516 | mtfC | mannosyltransferase C | 35.9% •••• |
| Aq1335 | nse | nucleotide sugar epimerase | 45.8% •••• |
| Aq505 | otnA | polysaccharide biosynthesis protein | 26.9% •••• |
| Aq504 | otnA' | polysaccharide biosynthesis protein (fragment) | 37.8% •• |
| Aq1543 | rfaC1 | ADP-heptose:LPS heptosyltransferase | 30.7% •••• |
| Aq145 | rfaC2 | ADP-heptose:LPS heptosyltransferase | 28.1% •••• |
| Aq344 | rfaD | ADP-L-glycero-D-manno-heptose-6-epimerase | 39.6% •••• |
| Aq565 | rfaE | ADP-heptose synthase | 44.0% •••• |
| Aq2115 | rfaG | glucosyl transferase I | 27.1% •••• |
| Aq1082 | rfbD | GDP-D-mannose dehydratase | 53.2% •••• |
| Aq519 | rfe | undecaprenyl-phosphate-alpha-N-acetylglucosaminyltransferase | 24.8% •••• |
| Aq1367 | spsI | glucose-1-phosphate thymidylyltransferase | 30.4% •• |
| Aq518 | spsK | spore coat polysaccharide biosynthesis protein SpsK | 49.5% ••• |
| Aq589 | xanB | mannose-6-phosphate isomerase/mannose-1-phosphate guanyl transferase | 40.9% •••• |
| **Cellular Processes** | | | |
| Cell division | | | |
| Aq698 | acrE | acriflavin resistance protein AcrE | 24.8% •••• |
| Aq1275 | cafA | cytoplasmic axial filament protein | 28.5% •••• |
| Aq523 | ftsA | cell division protein FtsA | 31.9% •••• |
| Aq936 | ftsH | cell division protein FtsH | 51.1% •••• |
| Aq1139 | ftsW | cell division protein FtsW | 30.8% ••• |
| Aq920 | ftsY | cell division protein FtsY | 35.2% ••• |
| Aq525 | ftsZ | cell division protein FtsZ | 48.6% •••• |
| Aq761 | gidA1 | glucose inhibited division protein A | 50.2% •••• |
| Aq691 | gidA2 | glucose inhibited division protein A | 57.5% •••• |
| Aq1582 | gidB | glucose inhibited division protein B | 39.4% •••• |
| Aq1718 | maf | MAF protein | 44.9% •••• |
| Aq1887 | mesJ | cell cycle protein MesJ | 27.7% •••• |
| Aq878 | minC | septum site-determining protein MinC | 39.4% •• |
| Aq1217 | minD1 | septum site-determining protein MinD | 33.1% •••• |
| Aq877 | minD2 | septum site-determining protein MinD | 54.5% •••• |
| Aq845 | mreB | rod shape determining protein MreB | 57.4% •••• |
| Aq025 | rodA | rod shape determining protein RodA | 37.6% •••• |
| Aq1130 | suflI | periplasmic cell division protein (SufI) | 28.1% •••• |
| Chaperones | | | |
| Aq154 | ctaB | cytochrome c oxidase assembly factor | 38.8% •••• |
| Aq1735 | dnaJ1 | chaperone DnaJ | 41.3% •••• |
| Aq703 | dnaJ2 | chaperone DnaJ | 45.1% •••• |
| Aq996 | dnaK | Hsp70 chaperone DnaK | 59.1% •••• |
| Aq433 | grpE | heat shock protein GrpE | 38.8% •••• |
| Aq192 | hslU | chaperone HslU | 57.5% •••• |
| Aq1283 | hspC | small heat shock protein (class I) | 31.0% •••• |
| Aq1991 | htpX | heat shock protein X | 51.1% •••• |
| Aq2200 | mopA | GroEL | 64.4% •••• |
| Aq2199 | mopB | GroES | 56.2% ••• |
| Detoxification | | | |
| Aq486 | ahpC1 | alkyl hydroperoxide reductase | 49.2% •••• |
| Aq858 | ahpC2 | alkyl hydroperoxide reductase | 53.4% •••• |
| Aq685 | arsC | arsenate reductase | 50.0% •••• |
| Aq136 | cpx | cytochrome c peroxidase | 48.9% •••• |
| Aq1005 | cutA | periplasmic divalent cation tolerance protein | 47.0% •••• |
| Aq1499 | sodA | superoxide dismutase (Fe/Mn family) | 34.2% •••• |
| Aq1050 | sodC1 | superoxide dismutase (Cu/Zn) | 39.5% •••• |
| Aq238 | sodC2 | superoxide dismutase (Cu/Zn) | 39.2% •••• |
| Aq488 | tpx | thiol peroxidase | 39.5% •••• |
| Motility | | | |
| Aq833 | flgA | flagellar protein FlgA | |
| Aq1184 | flgB | flagellar basal body rod protein FlgB | |
| Aq1183 | flgC | flagellar biosynthesis FlgC | 39.4% •••• |
| Aq1859 | flgE | flagellar hook protein FlgE | 30.8% •••• |
| Aq2051 | flgG1 | flagellar hook basal-body protein FlgG | 32.8% •••• |
| Aq834 | flgG2 | flagellar hook basal-body protein FlgG | 50.4% •••• |
| Aq1714 | flgH | flagellar L-ring protein FlgH | 31.9% •••• |
| Aq1713 | flgI | flagellar P-ring protein FlgI | 46.9% •••• |
| Aq1662 | flgK | flagellar hook associated protein FlgK | 21.9% •••• |
| Aq1663 | flgL | flagellar hook associated protein FlgL | 27.1% •• |
| Aq1212 | flhA | flagellar export protein | 44.0% •••• |
| Aq2014 | flhB | flagellar biosynthetic protein FlhB | 39.8% •••• |
| Aq1214 | flhF | flagellar biosynthesis FlhF | 28.7% •••• |
| Aq1998 | fliC | flagellin | 59.4% •••• |

Nature © Macmillan Publishers Ltd 1998

| | | | |
|---|---|---|---|
| Aq2001 | fliD | flagellar hook associated protein FliD | 24.3% •• |
| Aq1182 | fliF | Flagellar M-ring protein | 32.0% •••• |
| Aq653 | fliG | flagellar switch protein FliG | 35.9% •••• |
| Aq1595 | fliI | flagellar export protein | 44.6% •••• |
| Aq1860 | fliL | flagellar biosynthesis FliL | 30.6% •••• |
| Aq1539 | fliN | flagellar switch protein FliN | 42.9% ••• |
| Aq1920 | fliP | flagellar biosynthetic protein FliP | 47.7% •••• |
| Aq1962 | fliQ | flagellar biosynthesis protein FliQ | 45.5% •••• |
| Aq1961 | fliR | flagellar biosynthetic protein FliR | 29.7% •••• |
| Aq2002 | fliS | flagellar protein FliS | 30.8% •••• |
| Aq1003 | motA | flagellar motor protein MotA | 35.0% •••• |
| Aq1002 | motB1 | flagellar motor protein MotB | 36.8% •••• |
| Aq1001 | motB2 | flagellar motor protein MotB-like | 27.5% •••• |
| Secretion | | | |
| Aq1720 | ffh | signal recognition particle receptor protein | 49.1% •••• |
| Aq1288 | gspD | general secretion pathway protein D | 27.5% •••• |
| Aq1474 | gspE | general secretion pathway protein E | 48.8% •••• |
| Aq418 | gspG | general secretion pathway protein G | 50.7% •••• |
| Aq955 | lepB | type-I signal peptidase | 33.9% •••• |
| Aq1837 | lsp | lipoprotein signal peptidase | 37.4% •••• |
| Aq1271 | mpp | processing protease | 28.7% •••• |
| Aq747 | pilC1 | fimbrial assembly protein PilC | 37.4% •••• |
| Aq1285 | pilC2 | fimbrial assembly protein PilC | 28.9% •••• |
| Aq1601 | pilD | type 4 prepilin peptidase | 34.8% •••• |
| Aq745 | pilT | twitching motility protein PilT | 51.4% •••• |
| Aq2151 | pilU | twitching mobility protein | 41.6% •••• |
| Aq1870 | secA | preprotein translocase SecA subunit | 44.9% •••• |
| Aq973 | secD | protein export membrane protein SecD | 36.0% •••• |
| Aq1602 | secF | protein-export membrane protein | 41.4% •••• |
| Aq079 | secY | preprotein translocase SecY | 44.2% •••• |
| Aq2080 | sppA | proteinase IV | 43.4% •••• |
| Aq1971 | tapB | type IV pilus assembly protein TapB | 42.2% •••• |
| Aq1340 | tig | trigger factor | 27.4% •••• |
| **Central Intermediary Metabolism** | | | |
| One-carbon metabolism | | | |
| Aq1429 | metF | 5,10-methylenetetrahydrofolate reductase | 43.3% •••• |
| Aq1154 | metK | S-adenosylmethionine synthetase | 49.2% •••• |
| Aq1180 | sahH | S-adenosylhomocysteine hydrolase | 60.9% •••• |
| Cytoplasmic polysaccharides | | | |
| Aq1407 | bcsA | cellulose synthase catalytic subunit | 39.5% •••• |
| Aq1401 | celY | endoglucanase fragment | 33.0% •• |
| Aq721 | glgA | glycogen synthase | 38.1% •••• |
| Aq722 | glgB | 1,4-alpha-glucan branching enzyme | 56.5% •••• |
| Aq717 | glgP | glycogen phosphorylase | 37.0% •••• |
| Aq723 | malM | 4-alpha-glucanotransferase (amylomaltase) | 43.4% •••• |
| Tri-carboxylic acid cycle | | | |
| Aq1784 | aco | aconitase | 36.1% ••• |
| Aq1195 | forA1 | ferredoxin oxidoreductase alpha subunit | 31.5% ••• |
| Aq1167 | forA2 | ferredoxin oxidoreductase alpha subunit | 32.3% •••• |
| Aq1196 | forB1 | ferredoxin oxidoreductase beta subunit | 29.6% •••• |
| Aq1168 | forB2 | ferredoxin oxidoreductase beta subunit | 31.5% ••• |
| Aq1200 | forG1 | ferredoxin oxidoreductase gamma subunit | 34.5% •••• |
| Aq1169 | forG2 | ferredoxin oxidoreductase gamma subunit | 34.5% ••• |
| Aq594 | frdA | fumarate reductase flavoprotein subunit | 51.4% •••• |
| Aq553 | frdB1 | reductase iron-sulfur subunit | 35.2% •••• |
| Aq655 | frdB2 | fumarate reductase iron-sulfur subunit | 35.1% •••• |
| Aq1780 | fumB | fumarate hydratase (fumarase) | 46.4% •••• |
| Aq1679 | fumX | C-terminal fumarate hydratase, class I | 40.4% •••• |
| Aq150 | gltA | citrate synthase | 33.0% •••• |
| Aq1512 | icd | isocitrate dehydrogenase | 46.0% •••• |
| Aq1782 | mdh1 | malate dehydrogenase | 49.8% •••• |
| Aq1665 | mdh2 | malate dehydrogenase | 46.9% ••• |
| Aq1614 | oadA | oxaloacetate decarboxylase alpha chain | 50.1% •••• |
| Aq1306 | sucC1 | succinyl-CoA ligase beta subunit | 35.1% •••• |
| Aq1620 | sucC2 | succinyl-CoA ligase beta subunit | 52.9% •••• |
| Aq1888 | sucD1 | succinyl-CoA ligase alpha subunit | 41.7% ••• |
| Aq1622 | sucD2 | succinyl-CoA ligase alpha subunit | 65.7% •••• |
| Phosphate | | | |
| Aq1351 | phoH | phosphate starvation-inducible protein | 47.1% •••• |
| Aq1547 | ppa | inorganic pyrophosphatase | 56.5% •••• |
| Aq891 | ppx | exopolyphosphatase | 33.6% •••• |
| Polyamines | | | |
| Aq728 | speC | ornithine decarboxylase | 30.9% •••• |
| Aq062 | speE | spermidine synthase | 48.4% •••• |
| Sulfur | | | |
| Aq1081 | cysD | sulfate adenylyltransferase | 46.7% •••• |
| Aq1076 | rhdA | thiosulfate sulfurtransferase | 32.3% •••• |
| Aq1799 | rhdA | thiosulfate sulfurtransferase | 31.7% •••• |
| Aq455 | sor | sulfur oxygenase reductase | 36.7% •••• |
| Aq1803 | soxB | sulfur oxidation protein SoxB | 41.3% •••• |
| **Cofactor Biosynthesis** | | | |
| Lipoic acid biosynthesis | | | |
| Aq1355 | lipA | Lipoic acid synthetase | 48.9% ••• |
| Biotin | | | |
| Aq170 | bioA | DAPA aminotransferase | 51.7% •••• |
| Aq975 | bioB | biotin synthetase | 42.0% •••• |
| Aq557 | bioD | dethiobiotin synthetase | 41.5% •••• |
| Aq626 | bioF | 8-amino-7-oxononanoate synthase | 45.1% •••• |
| Aq1659 | bioW | 6-carboxyhexanoate-CoA ligase (pimeloyl CoA synthase) | 47.3% •••• |
| Aq566 | birA | biotin [acetyl-CoA-carboxylase] ligase | 37.5% •••• |
| Folic acid | | | |
| Aq2045 | folC | folylpolyglutamate synthetase | 31.8% •••• |
| Aq1898 | folD | methylenetetrahydrofolate dehydrogenase | 53.2% •••• |
| Aq239 | folE | GTP cyclohydrolase I | 57.1% •••• |
| Aq162 | folK | folate biosynthesis 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase | 43.7% •••• |
| Aq1468 | folP | dihydropteroate synthase | 45.8% •••• |
| Aq1144 | pabB | p-aminobenzoate synthetase | 41.5% ••• |
| Aq1606 | pabC | aminodeoxychorismate lyase | 29.0% •••• |
| Heme | | | |
| Aq207 | cobA | uroporphyrin-III c-methyltransferase | 52.1% •••• |
| Aq1237 | cysG | siroheme synthase | 36.9% •••• |
| Aq334 | dcuP | uroporphyrinogen decarboxylase | 41.4% •••• |
| Aq816 | gsa | glutamate-1-semialdehyde aminotransferase | 56.5% •••• |
| Aq1279 | hemA | glutamyl tRNA reductase (delta-aminolevulinate synthase) | 38.7% •••• |
| Aq2109 | hemB | porphobilinogen synthase | 64.5% •••• |
| Aq263 | hemC | porphobilinogen deaminase | 53.1% •••• |
| Aq1424 | hemF | oxygen-independent coproporphyrinogen III oxidase | 33.1% •••• |
| Aq2015 | hemG | protoporphyrinogen oxidase | 30.3% ••• |
| Aq948 | hemH | ferrochelatase | 46.4% •••• |
| Aq099 | hemK | protoporphyrinogen oxidase | 32.2% •••• |
| Aq2124 | hemN | oxygen-independent coproporphyrinogen II | 50.2% •••• |
| Molybdopterin | | | |
| Aq2183 | moaA2 | molybdenum cofactor biosynthesis protein A | 47.0% •••• |
| Aq527 | moaC | molybdenum cofactor biosynthesis moaC | 45.0% •• |
| Aq2181 | moaE | molybdopterin converting factor subunit 2 | 39.3% ••• |
| Aq1326 | mobB | molybdopterin-guainine dinucleotide biosynthesis protein B | 44.4% • |
| Aq030 | moeA1 | molybdenum cofactor biosynthesis protein A | 36.8% •••• |
| Aq1329 | moeB | molybdopterin biosynthesis protein MoeB | 54.1% •••• |
| Aq061 | mog | molybdenum cofactor biosynthesis MOG | 55.5% •••• |
| Aq049 | phhB | pterin-4a-carbinolamine dehydratase | 37.9% •••• |
| Panthenate | | | |
| Aq815 | dfp | pantothenate metabolism flavoprotein | 41.2% •••• |
| Aq1973 | panB | 3-methyl-2-oxobutanoate hydroxymethyltransferase | 45.5% •••• |
| Aq2132 | panC | pantothenate synthetase | 47.4% •••• |
| Aq476 | panD | aspartate 1-decarboxylase | 46.0% •••• |
| Pyridine nucleotides | | | |
| Aq1889 | nadA | quinolinate synthetase A | 44.3% •••• |
| Aq777 | nadB | L-aspartate oxidase | 36.7% ••• |
| Aq869 | nadC | quinolinate phosphoribosyl transferase | 47.0% •••• |
| Aq959 | nadE | NH(3)-dependent NAD+ synthetase | 39.6% •••• |
| Pyridoxal phosphate | | | |
| Aq852 | pdxA | pyridoxal phosphate biosynthetic protein PdxA | 36.8% •••• |
| Aq1423 | pdxJ | pyridoxal phosphate synthetase | 88.2% •••• |
| Quinones | | | |
| Aq895 | ispB | octoprenyl-diphosphate synthase | 35.7% •••• |
| Aq052 | ubiA | 4-hydroxybenzoate octaprenyltransferase | 41.4% •••• |
| Riboflavin | | | |
| Aq350 | ribA | GTP cyclohydrolase II | 61.7% •••• |
| Aq1707 | ribC | riboflavin synthase alpha chain | 45.3% •••• |
| Aq138 | ribD1 | riboflavin specific deaminase | 46.0% •••• |
| Aq436 | ribD2 | riboflavin specific deaminase | 42.9% •••• |
| Aq139 | ribF | riboflavin kinase | 38.4% •••• |
| Aq132 | ribH | riboflavin synthase beta subunit | 51.0% •••• |
| Thiamine | | | |
| Aq1204 | thiC | thiamine biosynthesis protein | 67.1% •••• |
| Aq1960 | thiD | HMP-P kinase | 40.5% •••• |
| Aq1366 | thiE1 | thiamine phosphate synthase | 36.3% •••• |
| Aq558 | thiE2 | thiamine phosphate synthase | 39.5% •••• |
| Aq2178 | thiG | thiamine biosynthesis, thiazole moiety | 52.5% •••• |
| Aq2119 | thiL | thiamine monophosphate kinase | 34.5% •••• |
| Thio- and glutaredoxin | | | |
| Aq443 | gua | glutaredoxin-like protein | 33.8% •••• |
| Aq1916 | trxA1 | thioredoxin | 58.9% ••• |
| Aq1811 | trxA2 | thioredoxin | 32.2% •• |
| Aq500 | trxB | thioredoxin reductase | 39.8% •••• |
| **Energy Metabolism** | | | |
| Aq1342 | gph | phosphoglycolate phosphatase | 33.9% •••• |
| ATP-Proton Motive Force | | | |
| Aq679 | atpA | ATP synthase F1 alpha subunit | 64.3% •••• |
| Aq179 | atpB | ATP synthase F0 subunit a | 36.4% •••• |
| Aq673 | atpC | ATP synthase F1 epsilon subunit | 37.4% •••• |
| Aq2038 | atpD | ATP synthase F1 beta subunit | 67.4% •••• |
| Aq177 | atpE | ATP synthase F0 subunit c | 53.8% ••• |
| Aq1586 | atpF1 | ATP synthase F0 subunit b | 26.3% •••• |
| Aq1587 | atpF2 | ATP synthase F0 subunit b | 25.5% •••• |
| Aq2041 | atpG | ATP synthase F1 gamma subunit | 39.9% •••• |
| Aq1588 | atpH | ATP synthase F1 delta chain | 28.1% •••• |
| Dehydrogenases | | | |
| Aq1362 | adh1 | alcohol dehydrogenase | 35.4% •••• |
| Aq1240 | adh2 | alcohol dehydrogenase | 28.8% •••• |
| Aq186 | aldH1 | aldehyde dehydrogenase | 41.9% •••• |
| Aq227 | aldH2 | aldehyde dehydrogenase | 28.0% •••• |
| Aq1145 | dhaT | 1,3 propanediol dehydrogenase | 36.6% •••• |
| Aq232 | dhsU | flavocytochrome C sulfide dehydrogenase | 33.6% •••• |
| Aq1769 | dld1 | D-lactate dehydrogenase | 45.3% •••• |
| Aq1234 | dmsA | DMSO reductase chain A | 25.0% •••• |
| Aq1232 | dmsB | DMSO reductase chain B | 38.4% ••• |
| Aq1231 | dmsC | DMSO reductase chain C | 29.5% • |
| Aq1691 | fdhE | formate dehydrogenase formation protein FdhE | 25.9% •••• |
| Aq1039 | fdoG | formate dehydrogenase alpha subunit | 50.0% •••• |
| Aq1046 | fdoH | formate dehydrogenase beta subunit | 45.7% •••• |
| Aq1049 | fdoI | formate dehydrogenase gamma subunit | 38.4% •••• |
| Aq1903 | gcsP1 | glycine dehydrogenase (decarboxylating) | 49.6% •••• |
| Aq1109 | gcsP2 | glycine dehydrogenase (decarboxylating) | 46.8% •••• |
| Aq1639 | glpC | oxido/reductase iron sulfur protein | 27.1% •••• |
| Aq395 | hdrA | heterodisulfide reductase subunit A | 39.7% •••• |
| Aq400 | hdrB | heterodisulfide reductase subunit B | 32.5% •••• |
| Aq398 | hdrC | heterodisulfide reductase subunit C | 35.7% • |
| Aq961 | hdrD | heterodisulfide reductase | 29.5% •••• |
| Aq038 | hibD | 3-hydroxyisobutyrate dehydrogenase | 34.6% •••• |
| Aq727 | ldhA | D-lactate dehydrogenase | 33.5% •••• |
| Aq736 | lpdA | dihydrolipoamide dehydrogenase | 37.0% •••• |
| Aq217 | narB | nitrate reductase narB | 39.1% •••• |
| Aq206 | nirB | nitrite reductase (NAD(P)H) large subunit | 35.3% ••• |
| Aq835 | nox | NADH oxidase | 33.1% •••• |
| Aq024 | nsd | nucleotide sugar dehydrogenase | 47.0% •••• |
| Aq135 | nueM | NADH dehydrogenase (ubiquinone) | 28.2% •••• |
| Aq1010 | udh | dehydrogenase | 29.7% •••• |
| Electron transport | | | |
| Aq2191 | coxA1 | cytochrome c oxidase subunit I | 42.4% •••• |
| Aq2192 | coxA2 | cytochrome c oxidase subunit I | 38.1% •••• |
| Aq2190 | coxB | cytochrome c oxidase subunit II | 27.4% •••• |
| Aq2188 | coxC | cytochrome c oxidase subunit III | 28.6% •••• |
| Aq153 | ctaA | heme O oxygenase | 28.1% •••• |
| Aq042 | cyc | cytochrome c | 25.8% ••• |
| Aq792 | cycB1 | cytochrome c552 | 29.9% •• |
| Aq1550 | cycB2 | cytochrome C552 | 38.7% •••• |
| Aq1357 | cydA | cytochrome oxidase d subunit I | 38.8% •••• |
| Aq1358 | cydB | cytochrome oxidase d subunit II | 31.2% •••• |
| Aq067 | dmsB | dimethylsulfoxide reductase chain B | 40.2% •••• |
| Aq235 | fccB' | sulfide dehydrogenase, flavoprotein subunit | 38.0% ••• |
| Aq919a | fdx1 | ferredoxin | 37.1% ••• |
| Aq1171a | fdx2 | ferredoxin | 43.9% •• |
| Aq1192a | fdx3 | ferredoxin | 35.0% ••• |
| Aq108a | fdx4 | ferredoxin | 56.6% ••• |
| Aq211 | fhp | flavohemoprotein | 43.4% •••• |
| Aq2096 | floX | flavodoxin | 32.5% •••• |
| Aq045 | petA | Rieske-I iron sulfur protein | 34.3% •••• |
| Aq044 | petB | cytochrome b | 38.3% ••• |
| Aq234 | soxF | Rieske-I iron sulfur protein | 29.0% •••• |
| Aq2186 | sqr | sulfide-quinone reductase | 41.0% •••• |
| Glycolysis and gluconeogenesis | | | |
| Aq484 | eno | enolase | 65.0% •••• |
| Aq1390 | fba | fructose-1,6-bisphosphate aldolase class II | 39.9% •••• |
| Aq1065 | gap | glyceraldehyde-3-phosphate dehydrogenase | 59.5% •••• |
| Aq434 | glpK | glycerol kinase | 51.0% •••• |
| Aq1744 | gpmA | phosphoglycerate mutase | 27.9% •••• |
| Aq1634 | gspA | glycerol-3-phosphate dehydrogenase (NAD+) | 40.5% •••• |

Nature © Macmillan Publishers Ltd 1998

| | | | |
|---|---|---|---|
| Aq1708 | pfkA | phosphofructokinase | 49.4% •••• |
| Aq750 | pgi | glucose-6-phosphate isomerase | 37.8% •••• |
| Aq118 | pgk | phosphoglycerate kinase | 54.5% •••• |
| Aq1990 | pgmA | phosphoglycerate mutase | 33.2% •••• |
| Aq501 | pmu | phosphoglucomutase/phosphomannomutase | 37.8% •••• |
| Aq2142 | ppsA | phosphoenolpyruvate synthase | 56.3% •••• |
| Aq1520 | pycA | pyruvate carboxylase c-terminal domain | 46.6% •••• |
| Aq1517 | pycB | pyruvate carboxylase n-terminal domain | 57.1% •••• |
| Aq360 | timA | triose phophate isomerase | 52.2% •••• |
| Hydrogenase | | | |
| Aq665 | hoxZ | Ni/Fe hydrogenase B-type cytochrome subunit | 40.4% •••• |
| Aq667 | hupD | HupD hydrogenase related function | 40.9% •••• |
| Aq666 | hupE | HupE hydrogenase related function | 38.3% •••• |
| Aq1021 | hypA | hydrogenase accessory protein HypA | 39.8% •••• |
| Aq671 | hypB | hydrogenase expression/formation protein B | 50.6% •••• |
| Aq1157 | hypD | hydrogenase expression/formation protein HypD | 56.1% •••• |
| Aq662 | mbhL1 | hydrogenase large subunit | 50.6% •••• |
| Aq960 | mbhL2 | hydrogenase large subunit | 44.3% •••• |
| Aq804 | mbhL3 | hydrogenase large subunit | 27.9% •••• |
| Aq660 | mbhS1 | hydrogenase small subunit | 66.6% •••• |
| Aq965 | mbhS2 | hydrogenase small subunit | 51.3% •••• |
| Aq802 | mbhS3 | hydrogenase small subunit | 36.7% •••• |
| Aq1591 | shyS | soluble hydrogenase small subunit | 41.6% •••• |
| Sugar metabolism | | | |
| Aq968 | cbbE2 | ribulose-5-phosphate 3-epimerase | 47.2% •••• |
| Aq1658 | fucA1 | fuculose-1-phosphate aldolase | 31.8% •••• |
| Aq1979 | fucA2 | fuculose-1-phosphate aldolase | 29.7% •••• |
| Aq498 | gnd | 6-phosphogluconate dehydrogenase | 45.2% •••• |
| Aq497 | gsdA | glucose-6-phosphate 1-dehydrogenase | 32.3% •••• |
| Aq1138 | rpiB | ribose 5-phosphate isomerase B | 54.5% •••• |
| Aq119 | talC | transaldolase | 71.1% •••• |
| Aq1765 | tktA | transketolase | 52.4% •••• |
| NADH dehydrogenase | | | |
| Aq1385 | nuoA1 | NADH dehydrogenase I chain A | 42.0% •••• |
| Aq1310 | nuoA2 | NADH dehydrogenase I chain A | 44.9% •••• |
| Aq1312 | nuoB | NADH dehydrogenase I chain B | 60.1% ••• |
| Aq551 | nuoD1 | NADH dehydrogenase I chain D | 37.7% •••• |
| Aq1314 | nuoD2 | NADH dehydrogenase I chain D | 42.2% •••• |
| Aq574 | nuoE | NADH dehydrogenase I chain E | 36.8% •••• |
| Aq573 | nuoF | NADH dehydrogenase I chain F | 20.5% •• |
| Aq437 | nuoG | NADH dehydrogenase I chain G | 35.4% •• |
| Aq1315 | nuoH1 | NADH dehydrogenase I chain H | 41.0% •••• |
| Aq1373 | nuoH2 | NADH dehydrogenase I chain H | 42.1% •••• |
| Aq1374 | nuoH3 | NADH dehydrogenase I chain H | 38.9% •••• |
| Aq1317 | nuoI1 | NADH dehydrogenase I chain I | 30.5% ••• |
| Aq1375 | nuoI2 | NADH dehydrogenase I chain I | 29.2% ••• |
| Aq1318 | nuoJ1 | NADH dehydrogenase I chain J | 35.4% •••• |
| Aq1377 | nuoJ2 | NADH dehydrogenase I chain J | 30.6% •••• |
| Aq1319 | nuoK1 | NADH dehydrogenase I chain K | 51.1% •••• |
| Aq1378 | nuoK2 | NADH dehydrogenase I chain K | 48.4% •••• |
| Aq1320 | nuoL1 | NADH dehydrogenase I chain L | 39.0% •••• |
| Aq866 | nuoL2 | NADH dehydrogenase I chain L | 30.2% ••• |
| Aq1379 | nuoL3 | NADH dehydrogenase I chain L | 43.1% •••• |
| Aq1321 | nuoM1 | NADH dehydrogenase I chain M | 43.6% •••• |
| Aq1382 | nuoM2 | NADH dehydrogenase I chain M | 36.9% •••• |
| Aq1322 | nuoN1 | NADH dehydrogenase I chain N | 34.1% •••• |
| Aq1383 | nuoN2 | NADH dehydrogenase I chain N | 32.8% •••• |
| **Lipid metabolism** | | | |
| Aq2058 | aas | 2-acylglycerophosphoethanolamine acyltransferase | 37.1% •••• |
| Aq1206 | accA | acetyl-CoA carboxylase alpha subunit | 57.1% ••• |
| Aq1363 | accB | biotin carboxyl carrier protein | 44.6% •••• |
| Aq1664 | accC1 | biotin carboxylase | 54.4% •••• |
| Aq1470 | accC2 | biotin carboxylase | 56.5% •••• |
| Aq445 | accD | acetyl-CoA carboxyltransferase beta subunit | 56.9% •••• |
| Aq1717a | acpP | acyl carrier protein | 71.2% •••• |
| Aq813 | acpS | holo-[acyl-carrier protein] synthase | 30.8% •••• |
| Aq2104 | acs | acetyl-coenzyme A synthetase | 54.0% •••• |
| Aq2103 | acs' | acetyl-coenzyme A synthetase c-terminal fragment | 61.2% •••• |
| Aq1249 | cds | phosphatidate cytidylyltransferase | 29.2% •••• |
| Aq1737 | cfa | cyclopropane-fatty-acyl-phospholipid synthase | 37.5% •••• |
| Aq892 | fabD | malonyl-CoA:Acyl carrier protein transacylase | 42.1% •••• |
| Aq1717 | fabF | 3-oxoacyl-[acyl-carrier-protein] synthase II | 58.4% •••• |
| Aq1716 | fabG | 3-oxoacyl-[acyl-carrier-protein] reductase | 52.9% •••• |
| Aq1099 | fabH | 3-oxoacyl-[acyl-carrier-protein] synthase III | 47.0% •••• |
| Aq1552 | fabI | enoyl-[acyl-carrier-protein] reductase (NADH) | 49.6% •••• |
| Aq056 | fabZ | (3R)-hydroxymyristoyl-(acyl carrier protein) dehydratase | 58.7% •••• |
| Aq999 | fadD | long-chain-fatty-acid CoA ligase | 30.0% ••• |
| Aq1638 | lplA | lipoate-protein ligase A | 28.1% •• |
| Aq958 | pgsA | phosphotidylglycerophosphate synthase | 37.3% •• |
| Aq2154 | pgsA | phosphotidylglycerophosphate synthase | 38.9% ••• |
| Aq1101 | plsX | PlsX protein | 43.7% •••• |
| **Purines, Pyrimidines, Nucleotides and Nucleosides** | | | |
| Aq094 | nrdA | ribonucleotide reductase alpha chain | 35.0% •••• |
| Aq1505 | nrdF | ribonucleotide reductase beta chain | 36.2% • |
| Purines | | | |
| Aq568 | deoD | purine nucleoside phosphorylase | 33.1% •••• |
| Aq236 | guaA | GMP synthase | 58.4% •••• |
| Aq2023 | guaB | inosine monophosphate dehydrogenase | 65.4% •••• |
| Aq544 | hpt | hypoxanthine-guanine phosphoribosyltransferase | 48.2% •••• |
| Aq078 | kad | adenylate kinase | 50.0% •••• |
| Aq1590 | ndk | nucleoside diphosphate kinase | 48.2% •••• |
| Aq1636 | prs | phosphoribosylpyrophosphate synthetase | 55.2% •••• |
| Aq1290 | purA | adenylosuccinate synthetase | 49.2% •••• |
| Aq597 | purB | adenylosuccinate lyase | 52.4% •••• |
| Aq2117 | purC | phosphoribosylaminoimidazole-succinocarboxamide synthase | 52.5% •••• |
| Aq742 | purD | phosphoribosylamine-glycine ligase | 54.2% •••• |
| Aq1178 | purE | phosphoribosylaminoimidazole carboxylase | 64.6% •••• |
| Aq1175 | purF | amidophosphoribosyltransferase | 42.7% •••• |
| Aq1963 | purH | phosphoribosylaminoimidazolecarboxamide formyltransferase | 48.2% •••• |
| Aq245 | purK | phosphoribosyl aminoimidazole carboxylase | 35.6% •••• |
| Aq1836 | purL | phosphoribosylformylglycinamidine synthase II | 49.3% ••• |
| Aq769 | purM | phosphoribosylformylglycinamidine cyclo-ligase | 50.0% •••• |
| Aq857 | purN | phosphoribosylglycinamide formyltransferase | 48.3% •••• |
| Aq1105 | purQ | phosphoribosyl formylglycinamidine synthase I | 51.1% •••• |
| Aq1818 | purU | formyltetrahydrofolate deformylase | 56.3% •••• |
| Pyrimidines | | | |
| Aq410 | carA | carbamoyl phosphate synthetase small subunit | 52.2% •••• |
| Aq1172 | carB | carbamoyl-phosphate synthase large subunit | 60.7% •••• |
| Aq2101 | carB | carbamoyl-phosphate synthase, large subunit | 63.1% •••• |
| Aq2153 | cmk | cytidylate kinase | 38.5% •••• |
| Aq1607 | dcd | deoxycytidine triphosphate deaminase | 39.5% •••• |
| Aq220 | dut | deoxyuridine 5'triphosphate nucleotidohydrolase | |
| Aq409 | pyrB | aspartate carbamoyltransferase catalytic chain | 42.0% •••• |
| Aq806 | pyrC | dihydroorotase | 37.3% ••• |
| Aq046 | pyrD | dihydroorotase dehydrogenase | 50.5% •••• |
| Aq1305 | pyrDB | dihydroorotate dehydrogenase electron transfer subunit | 34.7% •••• |
| Aq1580 | pyrF | orotidine-5'-phosphate decarboxylase | 37.2% •••• |
| Aq1334 | pyrG | CTP synthetase | 57.5% •••• |
| Aq713 | pyrH | UMP kinase | 62.1% •••• |
| Aq640 | thy | thymidylate synthase complementing protein | 30.5% ••• |
| Aq969 | tmk | thymidylate kinase | 35.1% •••• |
| Aq1907 | umpS | uridine 5-monophosphate synthase | 42.1% •••• |
| Aq2163 | uraP | uracil phosphoribosyltransferase | 42.0% •••• |
| **Regulation** | | | |
| Aq1058 | acrR1 | transcriptional regulator (TetR/AcrR family) | 34.1% ••• |
| Aq2179 | acrR2 | transcriptional regulator (TetR/AcrR family) | 31.0% ••• |
| Aq281 | acrR3 | transcriptional regulator (TetR/AcrR family) | 29.7% •••• |
| Aq1387 | arsR | transcriptional regulator (ArsR family) | 35.3% •••• |
| Aq1724 | degT | transcriptional regulator (DegT/DnrJ/Eryc1 family) | 34.1% •••• |
| Aq534 | draG | ADP-ribosylglycohydrolase | 32.1% •••• |
| Aq831 | exsB | trans-regulatory protein ExsB | 38.5% •••• |
| Aq490 | fnr | transcriptional regulator (Crp/Fnr family) | 29.5% •••• |
| Aq1207 | furR1 | transcriptional regulator (FurR family) | 37.9% •••• |
| Aq1418 | furR2 | transcriptional regulator (FurR family) | 34.6% •••• |
| Aq213 | glnBi | PII-like protein GlnBi | 48.0% •• |
| Aq1908 | hflX | GTP-binding protein HflX | 40.3% •••• |
| Aq1115 | hksP1 | histidine kinase sensor protein | 27.7% •• |
| Aq316 | hksP2 | histidine kinase sensor protein | 28.1% •••• |
| Aq905 | hksP3 | histidine kinase sensor protein | 23.6% ••• |
| Aq231 | hksP4 | histidine kinase sensor protein | 28.2% •••• |
| Aq1156 | hoxX | hydrogenase regulation HoxX | 46.7% •••• |
| Aq093 | hth | transcriptional regulator (H-T-H) | 50.2% •••• |
| Aq1019 | hypE | hydrogenase expression/formation protein | 44.3% •••• |
| Aq672 | hypF | transcriptional regulatory protein HypF | 44.8% •••• |
| Aq764 | iclR | transcriptional regulator (IclR family) | 30.4% •••• |
| Aq638 | lysR1 | transcriptional regulator (LysR family) | 32.8% •••• |
| Aq1038 | lysR2 | transcriptional regulator (LysR family) | 28.9% •••• |
| Aq702 | merR | transcriptional regulator (MerR family) | 32.8% •••• |
| Aq218 | nifA | transcriptional regulator (NifA family) | 42.8% •••• |
| Aq1117 | ntrC1 | transcriptional regulator (NtrC family) | 41.0% •••• |
| Aq1792 | ntrC2 | transcriptional regulator (NtrC family) | 40.2% •••• |
| Aq230 | ntrC3 | transcriptional regulator (NtrC family) | 40.0% •••• |
| Aq164 | ntrC4 | transcriptional regulator (NtrC family) | 38.3% •••• |
| Aq2069 | obg | GTP-binding protein | 54.9% •• |
| Aq319 | phoB | transcriptional regulator (PhoB-like) | 41.6% •••• |
| Aq906 | phoU | transcriptional regulator (PhoU-like) | 41.9% •••• |
| Aq844 | spoT | (p)ppGpp 3-pyrophosphohydrolase | 47.2% ••• |
| Aq1496 | xylR | transcriptional regulator (NagC/XylR family) | 29.3% •••• |
| **DNA Replication and Repair** | | | |
| Aq358 | dinG | ATP-dependent helicase (DinG family) | 27.9% •••• |
| Aq322 | dnaA | chromosome replication initiator protein DnaA | 36.5% •••• |
| Aq1472 | dnaB | replicative DNA helicase | 40.3% •••• |
| Aq910 | dnaC | DNA replication protein DnaC | 26.4% •••• |
| Aq1008 | dnaE | DNA polymerase III alpha subunit | 41.9% ••• |
| Aq1493 | dnaG | DNA primase | 39.8% ••• |
| Aq1882 | dnaN | DNA polymerase III beta chain | 32.1% •••• |
| Aq932 | dnaQ | DNA polymerase III epsilon subunit | 40.0% •••• |
| Aq1855 | dnaX | DNA polymerase III gamma subunit | 36.6% •••• |
| Aq1422 | dpbF | DNA polymerase beta family | 39.1% •••• |
| Aq1693 | dplF | N-terminus of phage SPO1 DNA polymerase | 37.3% •••• |
| Aq980 | gyrA | DNA gyrase A subunit | 43.6% •••• |
| Aq1026 | gyrB | gyrase B | 55.2% •• |
| Aq2057 | helX | DNA helicase | 49.7% •••• |
| Aq1484a | himA | DNA binding protein HU | 40.2% •••• |
| Aq2174 | ihfB | integration host factor beta subunit | 35.8% •••• |
| Aq1394 | lig | DNA ligase (ATP dependent) | 50.8% •••• |
| Aq633 | ligA | DNA ligase (NAD dependent) | 45.7% •••• |
| Aq1578 | mutL | DNA mismatch repair protein MutL | 72.3% •••• |
| Aq308 | mutS1 | DNA mismatch repair protein MutS | 77.5% •••• |
| Aq1242 | mutS2 | DNA mismatch repair protein MutS | 37.0% •••• |
| Aq1449 | mutT | 8-OXO-dGTPase domain (mutT domain) | 46.3% •••• |
| Aq282 | mutY1 | endonuclease III | 53.6% •••• |
| Aq172 | mutY2 | endonuclease III | 51.8% •••• |
| Aq496 | mutY3 | endonuclease III | 43.4% •••• |
| Aq1629 | nfo | deoxyribonuclease IV | 39.0% •••• |
| Aq710 | nucI | thermococcal nuclease homolog | 36.4% •••• |
| Aq1495 | ogt | O-6-methylguanine-DNA-alkyltransferase | 36.9% ••• |
| Aq1628 | pol | DNA polymerase I 3'-5' exo domain | 43.2% •••• |
| Aq1967 | polA | DNA polymerase I (PolI) | 30.5% •••• |
| Aq1610 | radC | DNA repair protein RadC | 39.0% •••• |
| Aq2150 | recA | recombination protein RecA | 88.5% •••• |
| Aq2053 | recG | ATP-dependent DNA helicase RecG | 38.9% •••• |
| Aq2155 | recJ | single-strand-DNA-specific exonuclease RecJ | 31.8% •••• |
| Aq561 | recN | recombination protein RecN | 27.7% •••• |
| Aq1478 | recR | recombination protein RecR | 38.3% •••• |
| Aq793 | rep | ATP-dependent DNA helicase REP | 33.4% •••• |
| Aq1886 | sbcD | ATP-dependent dsDNA exonuclease | 29.9% •••• |
| Aq064 | ssb | single stranded DNA-binding protein | 39.4% ••• |
| Aq657 | topA | topoisomerase I | 39.6% •••• |
| Aq1159 | topG1 | reverse gyrase | 41.6% •••• |
| Aq886 | topG2 | reverse gyrase | 35.1% •••• |
| Aq686 | uvrA | repair excision nuclease subunit A | 61.0% •••• |
| Aq1856 | uvrB | repair excision nuclease subunit B | 53.9% •••• |
| Aq2126 | uvrC | repair excision nuclease subunit C | 32.5% •••• |
| **Transcription** | | | |
| RNA polymerase and transcription factors | | | |
| Aq613 | deaD | ATP-dependent RNA helicase DeaD | 42.3% •••• |
| Aq357a | flgM | anti sigma factor FlgM | 20.6% •••• |
| Aq1218 | fliA | RNA polymerase sigma factorFliA | 37.2% •••• |
| Aq259 | nusA | transcription termination NusA | 45.4% ••• |
| Aq133 | nusB | transcription termination NusB | 32.3% •••• |
| Aq1931 | nusG | transcription antitermination protein NusG | 46.3% •• |
| Aq873 | rho | transcriptional terminator Rho | 59.6% •••• |
| Aq070 | rpoA | RNA polymerase alpha subunit | 40.4% •••• |
| Aq1939 | rpoB | RNA polymerase beta subunit | 44.0% •••• |
| Aq1945 | rpoC | RNA polymerase beta prime subunit | 46.9% •••• |
| Aq1490 | rpoD | RNA polymerase sigma factor RpoD | 41.6% •••• |
| Aq599 | rpoN | RNA polymerase sigma factor RpoN | 30.6% •••• |
| Aq1452 | rpoS | RNA polymerase sigma factor RpoS | 40.5% •••• |
| RNA modification | | | |
| Aq1816 | ksgA | dimethyladenosine transferase | 36.1% •••• |
| Aq1067 | miaA | tRNA delta-2-isopentenylpyrophosphate (IPP) transferase | 38.2% •••• |
| Aq411 | pcnB1 | poly A polymerase | 28.5% •••• |
| Aq2158 | pcnB2 | poly A polymerase | 33.9% •• |
| Aq221 | phpA | polyribonucleotide nucleotidyltransferase | 45.0% •••• |
| Aq894 | queA | queuosine biosynthesis protein | 46.9% •••• |
| Aq946 | rnc | RNase III | 35.8% •••• |
| Aq1955 | rnhB | RNase HII | 48.4% •••• |
| Aq924 | rnpH | RNase PH | 64.0% •••• |
| Aq1661 | spoU | rRNA methylase SpoU | 44.0% •••• |
| Aq1308 | tgt | queuine tRNA-ribosyltransferase | 52.6% •••• |
| Aq841 | trm1 | N2,N2-dimethylguanosine tRNA | |

Nature © Macmillan Publishers Ltd 1998

Nature © Macmillan Publishers Ltd 1998

| | | | |
|---|---|---|---|
| | | methyltransferase | 34.6% •••• |
| Aq1489 | trmD | tRNA guanine-N1 methyltransferase | 42.9% •••• |
| Aq749 | truA | pseudouridine synthase I | 33.1% •••• |
| Aq705 | truB | tRNA pseudouridine 55 synthase | 38.2% •••• |
| Aq1890 | tsnR | rRNA methylase | 36.4% •••• |
| Aq2046 | vacB | VacB protein (ribonuclease II family) | 37.9% •••• |
| Aq257 | ygcA | RNA methyltransferase (TrmA-family) | 28.8% •••• |
| **Translation** | | | |
| Aq2131 | fmt | methionyl-tRNA formyltransferase | 45.7% •••• |
| Aq247 | gatA | glutamyl-tRNA(Gln) amidotransferase subunit A | 53.6% •••• |
| Aq461 | gatB | glutamyl-tRNA(Gln) amidotransferase subunit B | 48.8% •••• |
| Aq2147a | gatC | glutamyl-tRNA(Gln) amidotransferase subunit C | 41.1% •••• |
| Aq346 | pth | peptidyl-tRNA hydrolase | 48.8% •••• |
| Aminoacyl tRNA synthetases | | | |
| Aq1293 | alaS | alanyl-tRNA synthetase | 46.6% •••• |
| Aq923 | argS | arginyl-tRNA synthetase | 39.4% •••• |
| Aq1677 | aspS | aspartyl-tRNA synthetase | 51.3% •••• |
| Aq1068 | cysS | cysteinyl-tRNA synthetase | 45.0% •••• |
| Aq763 | genX | lysyl-tRNA synthetase (genX) homolog | 38.6% •••• |
| Aq1221 | gltX | glutamyl-tRNA synthetase | 48.5% •••• |
| Aq945 | glyQ | glycyl-tRNA synthetase alpha subunit | 61.9% •••• |
| Aq2141 | glyS | glycyl-tRNA synthetase beta subunit | 37.1% •••• |
| Aq122 | hisS1 | histidyl-tRNA synthetase | 43.3% •••• |
| Aq1155 | hisS2 | histidyl-tRNA synthetase | 34.9% •••• |
| Aq305 | ileS | isoleucyl-tRNA synthetase | 82.1% •• |
| Aq351 | leuS | leucyl-tRNA synthetase alpha subunit | 50.7% •••• |
| Aq1770 | leuS' | leucyl-tRNA synthetase beta subunit | 47.2% •••• |
| Aq1202 | lysU | lysyl-tRNA synthetase | 53.2% ••• |
| Aq1257 | metG | methionyl-tRNA synthetase alpha subunit | 45.0% •••• |
| Aq422 | metG' | methionyl-tRNA synthetase beta subunit | 64.2% •••• |
| Aq953 | pheS | phenylalanyl-tRNA synthetase alpha subunit | 51.9% •••• |
| Aq1730 | pheT | phenylalanyl-tRNA synthetase beta subunit | 35.4% •••• |
| Aq365 | proS | proline-tRNA synthetase | 44.1% •••• |
| Aq298 | serS | seryl-tRNA synthetase | 59.4% •••• |
| Aq1667 | thrS | threonyl-tRNA synthetase | 48.5% •••• |
| Aq992 | trpS | tryptophanyl-tRNA synthetase | 38.4% •••• |
| Aq1751 | tyrS | tyrosyl tRNA synthetase | 56.2% •••• |
| Aq1413 | valS | valyl-tRNA synthetase | 33.2% •••• |
| Ribosomal Proteins | | | |
| Aq1935 | rplA | ribosomal protein L01 | 57.9% •••• |
| Aq013 | rplB | ribosomal protein L02 | 46.9% •••• |
| Aq009 | rplC | ribosomal protein L03 | 53.8% •••• |
| Aq011 | rplD | ribosomal protein L04 | 51.3% •••• |
| Aq1652 | rplE | ribosomal protein L05 | 67.0% •••• |
| Aq1649 | rplF | ribosomal protein L06 | 46.2% •••• |
| Aq2042 | rplI | ribosomal protein L09 | 35.6% •••• |
| Aq1936 | rplJ | ribosomal protein L10 | 36.5% •••• |
| Aq1933 | rplK | ribosomal protein L11 | 71.4% •••• |
| Aq1937 | rplL | ribosomal protein L7/L12 | 75.4% •••• |
| Aq1877 | rplM | ribosomal protein L13 | 60.6% •••• |
| Aq1654 | rplN | ribosomal protein L14 | 59.5% •••• |
| Aq1642 | rplO | ribosomal protein L15 | 57.4% •••• |
| Aq018 | rplP | ribosomal protein L16 | 59.3% •••• |
| Aq069 | rplQ | ribosomal protein L17 | 48.7% •••• |
| Aq1648 | rplR | ribosomal protein L18 | 62.7% •••• |
| Aq1954 | rplS | ribosomal protein L19 | 59.8% ••• |
| Aq952 | rplT | ribosomal protein L20 | 63.5% •••• |
| Aq016a | rplV | ribosomal protein L22 | 47.3% •••• |
| Aq012 | rplW | ribosomal protein L23 | 52.2% •••• |
| Aq1653 | rplX | ribosomal protein L24 | 50.8% •••• |
| Aq1644 | rpmD | ribosomal protein L30 | 46.4% •• |
| Aq1930a | rpmG | ribosomal protein L33 | 67.9% •••• |
| Aq792a | rpmI | ribosomal protein L35 | 48.3% •••• |
| Aq1485 | rpsA | ribosomal protein S01 | 32.6% •••• |
| Aq2007 | rpsB | ribosomal protein S02 | 60.3% •••• |
| Aq017 | rpsC | ribosomal protein S03 | 54.0% •••• |
| Aq072 | rpsD | ribosomal protein S04 | 51.9% •••• |
| Aq1645 | rpsE | ribosomal protein S05 | 60.6% ••• |
| Aq063 | rpsF | ribosomal protein S06 | 32.7% •••• |
| Aq1832 | rpsG1 | ribosomal protein S07 | 52.5% •••• |
| Aq734 | rpsG2 | ribosomal protein S07 | 51.9% •••• |
| Aq1651 | rpsH | ribosomal protein S08 | 39.9% •••• |
| Aq1878 | rpsI | ribosomal protein S09 | 50.5% ••• |
| Aq008 | rpsJ | ribosomal protein S10 | 55.9% •••• |
| Aq073 | rpsK | ribosomal protein S11 | 60.7% •••• |
| Aq735 | rpsL1 | ribosomal protein S12 | 78.9% •••• |
| Aq1834 | rpsL2 | ribosomal protein S12 | 78.9% •••• |
| Aq074 | rpsM | ribosomal protein S13 | 61.9% •••• |
| Aq1651a | rpsN | ribosomal protein S14 | 51.6% •••• |
| Aq226a | rpsO | ribosomal protein S15 | 61.6% •••• |
| Aq123 | rpsP | ribosomal protein S16 | 36.6% ••• |
| Aq020 | rpsQ | ribosomal protein S17 | 59.6% •••• |
| Aq064a | rpsR | ribosomal protein S18 | 48.5% •••• |
| Aq015 | rpsS | ribosomal protein S19 | 63.1% •• |
| Aq1767 | rpsT | ribosomal protein S20 | 40.0% ••• |
| Aq867a | rpsU | ribosomal protein S21 | 38.2% •••• |
| Translation factors | | | |
| Aq1364 | efp | elongation factor P | 48.6% •••• |
| Aq2114 | eif | initiation factor eIF-2B alpha subunit | 58.4% ••• |
| Aq712 | frr | ribosome recycling factor | 43.0% •••• |
| Aq001 | fusA | elongation factor EF-G | 91.9% •••• |
| Aq075a | infA | initiation factor IF-1 | 69.1% •••• |
| Aq2032 | infB | initiation factor IF-2 | 48.5% •••• |
| Aq1777 | infC | initiation factor IF-3 | 53.6% •••• |
| Aq876 | prfA | peptide chain release factor RF-1 | 54.8% •••• |
| Aq1840 | prfB | peptide chain release factor RF-2 | 49.9% •••• |
| Aq1033 | selB | elongation factor SelB | 30.4% •••• |
| Aq715 | tsf | elongation factor EF-Ts | 35.8% •••• |
| Aq005 | tufA1 | elongation factor EF-Tu | 74.4% •••• |
| Aq1928 | tufA2 | elongation Factor EF-Tu | 73.9% •••• |
| Protein modification | | | |
| Aq731 | ccdA | cytochrome c-type biogenesis protein | 32.0% •••• |
| Aq579 | def | polypeptide deformylase | 41.4% •••• |
| Aq2093 | dsbC | thiol:disulfide interchange protein | 27.6% •••• |
| Aq055 | hemX1 | cytochrome c biogenesis protein | 26.2% •••• |
| Aq2043 | hemX2 | cytochrome c biogenesis protein | 36.2% •••• |
| Aq1053 | nifS1 | FeS cluster formation protein NifS | 38.5% •••• |
| Aq739 | nifS2 | FeS cluster formation protein NifS | 45.5% •••• |
| Aq1871 | pmbA | peptide maturation | 25.6% •••• |
| Aq2102 | prmA | ribosomal protein L11 methyltransferase | 35.1% •••• |
| Aq567 | rimI | ribosomal-protein-alanine acetyltransferase | 37.9% •••• |
| Aq576 | stpK | ser/thr protein kinase | 30.8% •••• |
| Aq152 | tlpA | thiol disulfide interchange protein | 37.6% ••• |
| **Proteases** | | | |
| Aq1950 | aprV | serine protease | 26.5% ••• |
| Aq1672 | clpB | ATPase subunit of ATP-dependent protease | 46.8% •••• |
| Aq1296 | clpC | ATP-dependent Clp protease | 54.9% •••• |
| Aq1339 | clpP | ATP-dependent Clp protease proteolytic subunit | 65.4% •••• |
| Aq1337 | clpX | ATP-dependent protease ATPase subunit clpX | 66.1% •••• |
| Aq1015 | col | collagenase | 41.3% •••• |
| Aq801 | gcp | sialoglycoprotease | 45.5% •••• |
| Aq1671 | hslV | heat shock protein HsLV | 57.6% •••• |
| Aq1450 | htrA | periplasmic serine protease | 38.3% •••• |
| Aq242 | lon | Lon protease | 50.6% •••• |
| Aq076 | map | methionyl aminopeptidase | 44.1% •••• |
| Aq1459 | npr | neutral protease | 27.7% •• |
| Aq2099 | pepA | leucine aminopeptidase | 39.5% •••• |
| Aq1535 | pepQ | xaa-pro dipeptidase | 31.9% •••• |
| Aq618 | pfpI | protease I | 41.8% •••• |
| Aq797 | prc | carboxyl-terminal protease | 41.8% •••• |
| Aq552 | sms | ATP-dependent protease sms | 46.2% •••• |
| Aq2204 | ymxG | processing protease | 28.3% •••• |
| **Transport** | | | |
| Aq1222 | abcT1 | ABC transporter | 34.7% •••• |
| Aq620 | abcT2 | ABC transporter | 36.8% •••• |
| Aq1095 | abcT3 | ABC transporter (ABC-2 subfamily) | 34.4% •••• |
| Aq1094 | abcT4 | ABC transporter | 37.7% •••• |
| Aq1097 | abcT5 | ABC transporter (hlyB subfamily) | 45.5% •••• |
| Aq417 | abcT6 | ABC transporter | 51.8% •••• |
| Aq413 | abcT7 | ABC transporter | 51.5% •••• |
| Aq297 | abcT8 | ABC transporter | 49.3% •••• |
| Aq2160 | abcT9 | ABC transporter | 45.3% •••• |
| Aq1531 | abcT10 | ABC transporter | 36.4% •••• |
| Aq2122 | abcT11 | ABC transporter | 42.5% •••• |
| Aq2137 | abcT12 | ABC transporter | 38.2% •••• |
| Aq1563 | abcT13 | ABC transporter (MsbA subfamily) | 30.5% •••• |
| Aq695 | acrD1 | cation efflux system (AcrB/AcrD/AcrF family) | 22.7% •••• |
| Aq1122 | acrD2 | cation efflux system (AcrB/AcrD/AcrF family) | 32.0% •••• |
| Aq469 | acrD3 | cation efflux system (AcrB/AcrD/AcrF family) | 34.2% •••• |
| Aq786 | acrD4 | cation efflux (AcrB/AcrD/AcrF family) | 27.7% •••• |
| Aq112 | amtB | ammonium transporter | 49.0% •••• |
| Aq682 | arsA1 | anion transporting ATPase | 41.5% •••• |
| Aq343 | arsA2 | anion transporting ATPase | 33.9% •••• |
| Aq851 | corA | Mg(2+) and Co(2+) transport protein | 31.1% •••• |
| Aq724 | ctrA1 | cation transporting ATPase (E1-E2 family) | 30.7% •••• |
| Aq1445 | ctrA2 | cation transporting ATPase (E1-E2 family) | 28.1% •••• |
| Aq1125 | ctrA3 | cation transporting ATPase (E1-E2 family) | 43.8% •••• |
| Aq1132 | czcB1 | cation efflux system (czcB-like) | 23.7% •••• |
| Aq1331 | czcB2 | cation efflux system (czcB-like) | 26.9% •••• |
| Aq468 | czcB2 | cation efflux system (czcB-like) | 28.5% •••• |
| Aq1073 | czcD | cation efflux system (CzcD-like) | 43.4% •••• |
| Aq911 | ebs | erythrocyte band 7 homolog | 50.2% •••• |
| Aq1062 | emrB | major facilitator family transporter | 28.3% •••• |
| Aq1255 | feoB | ferrous iron transport protein B | 32.6% •••• |
| Aq1330 | gltP | proton/sodium-glutamate symport protein | 35.6% •••• |
| Aq1268 | hvsT | high affinity sulfate transporter | 29.4% •••• |
| Aq1863 | kch | potassium channel protein | 30.1% ••• |
| Aq1725 | lepA | G-protein LepA | 59.8% •••• |
| Aq1229 | mffT | transporter (major facilitator family) | 37.2% •••• |
| Aq447 | mgtC | Mg(2+) transport ATPase | 36.2% •••• |
| Aq1609 | modA | molybdate periplasmic binding protein | 38.2% •••• |
| Aq086 | modC | Molybdenum transport system permease | 44.8% •••• |
| Aq415 | napA1 | Na(+)/H(+) antiporter | 27.6% •••• |
| Aq929 | napA2 | Na(+)/H(+) antiporter | 32.7% •••• |
| Aq2030 | napA3 | Na(+)/H(+) antiporter | 26.8% •••• |
| Aq215 | nasA | nitrate transporter | 35.8% ••• |
| Aq1441 | oppA | transporter (extracellular solute binding protein family 5) | 37.0% •••• |
| Aq481 | oppB | transporter (OppBC family) | 46.2% •••• |
| Aq1509 | oppC | oligopeptide transport system permease | 46.2% •••• |
| Aq2019 | pstA | phosphate transport system permease PstA | 43.5% •••• |
| Aq1055 | pstB | phosphate transport ATP binding protein | 68.1% •••• |
| Aq2018 | pstC | phosphate transport system permease protein C | 45.2% •••• |
| Aq2016 | pstS | phosphate-binding periplasmic protein | 52.4% •••• |
| Aq2129 | sbf | Na(+) dependent transporter (Sbf family) | 34.9% •••• |
| Aq098 | secG | protein export membrane protein SecG | 35.7% •••• |
| Aq2077 | snf | Na(+):neurotransmitter symporter (Snf family) | 25.7% •••• |
| Aq2106 | ssf | Na(+):solute symporter (Ssf family) | 47.4% • |
| Aq1988 | tolQ | TolQ homolog | 32.5% • |
| Aq1504 | trk1 | K+ transport protein homolog | 40.6% •••• |
| Aq031 | trnS | transporter (Pho87 family) | 46.8% •••• |
| **Uncategorized** | | | |
| Aq1023 | acuC1 | acetoin utilization protein | 36.9% •••• |
| Aq2110 | acuC2 | acetoin utilization protein | 38.6% •••• |
| Aq158 | apfA | AP4A hydrolase | 36.6% •••• |
| Aq458 | bcp | bacterioferritin comigratory protein | 40.6% •••• |
| Aq542 | bcpC | phosphonopyruvate decarboxylase | 37.4% •••• |
| Aq147 | cobW | cobalamin synthesis related protein CobW | 29.5% ••• |
| Aq1303a | cspC | cold shock protein | 67.2% •••• |
| Aq1265 | cstA | carbon starvation protein A | 33.0% •••• |
| Aq348 | ctc | general stress protein Ctc | 34.7% •••• |
| Aq212 | cynS | cyanate hydrolase | 39.5% •••• |
| Aq337 | cysQ | CysQ protein | 47.4% •••• |
| Aq528 | delF | phenylacrylic acid decarboxylase | 52.4% •••• |
| Aq148 | deoC | deoxyribose-phosphate aldolase | 46.6% •••• |
| Aq2095 | dksA | dnaK suppressor protein | 35.1% •••• |
| Aq1994 | era1 | GTP-binding protein Era | 49.7% •••• |
| Aq1919 | era2 | GTP binding protein Era | 43.0% •••• |
| Aq1540 | gcpE | GcpE protein | 50.1% •••• |
| Aq1052 | gcsH1 | glycine cleavage system protein H | 28.6% •••• |
| Aq1657 | gcsH2 | glycine cleavage system protein H | 39.8% ••• |
| Aq944 | gcsH3 | glycine cleavage system protein H | 36.7% ••• |
| Aq1108 | gcsH4 | glycine cleavage system protein H | 44.8% ••• |
| Aq1458 | gcvT | aminomethyltransferase (glycine cleavage system T protein) | 42.2% •••• |
| Aq108b | hfq | host factor I | 53.5% •••• |
| Aq101 | hly | hemolysin | 33.7% ••• |
| Aq2120 | hlyC | hemolysin homolog protein | 29.3% •••• |
| Aq1091 | hylA | hemolysin | 33.5% ••• |
| Aq708 | hyuA | N-methylhydantoinase A | 39.8% •••• |
| Aq1925 | hyuB | N-methylhydantoinase B | 43.1% •••• |
| Aq1579 | iagB | invasion protein IagB | 38.3% ••• |
| Aq1983 | imp2 | myo-inositol-1(or 4)-monophosphatase | 36.0% •••• |
| Aq748 | ispA | geranylgeranyl pyrophosphate synthase | 40.7% •••• |
| Aq1739 | lytB | LytB protein | 43.9% •••• |
| Aq1977 | masA | enolase-phosphatase E-1 | 42.3% •••• |
| Aq1560 | mglA1 | gliding motility protein | 42.4% •••• |
| Aq1823 | mglA2 | gliding motility protein MglA | 34.1% •••• |
| Aq1789 | mviB | 'virulence factor' homolog MviB | 29.7% •••• |
| Aq587 | neaC | N-ethylammeline chlorohydrolase | 42.8% •••• |
| Aq1820 | nfeD | nodulation competitiveness protein NfeD | 37.9% •••• |
| Aq896 | nifU | NifU protein | 48.3% •••• |
| Aq1300 | omp | outer membrane protein | 25.5% •••• |
| Aq1507 | omt | O-methyltransferase | 39.5% •••• |
| Aq967 | ostA | organic solvent tolerance protein | 22.0% •••• |
| Aq141 | pkcI | protein kinase C inhibitor (HIT family) | 59.0% •••• |
| Aq994 | pncA | pyrazinamidase/nicotinamidase | 39.1% •••• |
| Aq057 | sfsA | sugar fermentation stimulation protein | 27.3% •••• |
| Aq287 | smb | small protein B | 52.0% •••• |
| Aq832 | surE | stationary phase survival protein SurE | 44.1% •••• |
| Aq871 | thdF | thiophene and furan oxidation protein | 45.4% •••• |
| Aq2021 | tldD | TldD protein | 40.9% •••• |
| Aq773 | tly | hemolysin | 43.8% •••• |
| Aq629 | xcpC | chromosome assembly protein homolog | 33.3% •••• |

Nature © Macmillan Publishers Ltd 1998

Nature © Macmillan Publishers Ltd 1998