Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*

Karen E. Nelson, Rebecca A. Clayton, Steven R. Gill, Michelle L. Gwinn, Robert J. Dodson, Daniel H. Haft, Erin K. Hickey, Jeremy D. Peterson, William C. Nelson, Karen A. Ketchum, Lisa McDonald, Teresa R. Utterback, Joel A. Malek, Katja D. Linher, Mina M. Garrett, Ashley M. Stewart, Matthew D. Cotton, Matthew S. Pratt, Cheryl A. Phillips, Delwood Richardson, John Heidelberg, Granger G. Sutton, Robert D. Fleischmann, Jonathan A. Eisen, Owen White, Steven L. Salzberg, Hamilton O. Smith, J. Craig Venter & Claire M. Fraser

The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA

The 1,860,725-base-pair genome of *Thermotoga maritima* MSB8 contains 1,877 predicted coding regions, 1,014 (54%) of which have functional assignments and 863 (46%) of which are of unknown function. Genome analysis reveals numerous pathways involved in degradation of sugars and plant polysaccharides, and 108 genes that have orthologues only in the genomes of other thermophilic Eubacteria and Archaea. Of the Eubacteria sequenced to date, *T. maritima* has the highest percentage (24%) of genes that are most similar to archaeal genes. Eighty-one archaeal-like genes are clustered in 15 regions of the *T. maritima* genome that range in size from 4 to 20 kilobases. Conservation of gene order between *T. maritima* and Archaea in many of the clustered regions suggests that lateral gene transfer may have occurred between thermophilic Eubacteria and Archaea.

Thermotoga maritima, a non-spore-forming, rod-shaped bacterium belonging to the order Thermotogales, was originally isolated from geothermal heated marine sediment at Vulcano, Italy[1], and has an optimum growth temperature of 80 °C. *T. maritima* metabolizes many simple and complex carbohydrates including glucose, sucrose, starch, cellulose and xylan[1,2]. Both cellulose and xylan, through conversion to fuels (such as H_2), have great potential as renewable carbon and energy sources.

T. maritima is also of evolutionary significance, because small-subunit ribosomal RNA (SSU rRNA) phylogeny has placed this bacterium as one of the deepest and most slowly evolving lineages in the Eubacteria[3]. To elucidate further its unique metabolic properties and evolutionary relationship to other microbial species, we sequenced the genome of the type strain *T. maritima* MSB8 using the whole-genome random-sequencing method previously described[4,5].

General features of the genome

The genome of *T. maritima* is a single circular chromosome consisting of 1,860,725 base pairs (bp) (Fig. 1) with an average G + C content of 46%. A single rRNA operon (16S–23S–5S), containing an isoleucine transfer RNA and an alanyl tRNA in the spacer region between the small- and large-subunit genes, corresponds to the one region of the chromosome with a significantly higher G + C content (62%). A region of significantly lower G + C content (34%) encodes lipopolysaccharide biosynthesis (LPS) proteins.

On the basis of analysis of G + C ratio, G–C skew[6] (G − C/G + C) and asymmetric distribution of oligomers[7] in the *T. maritima* genome, we could not identify a characteristic bacterial origin of replication, a situation similar to that observed in the genomes of the Archaea *Methanococcus jannaschii*[8] and *Archaeoglobus fulgidus*[5]. We assigned base-pair one of the genome at the beginning of the longest stretch (2.6-kb) of 30-bp repeats (Fig. 2).

Open reading frames. We identified 1,877 open reading frames (ORFs) (Figs 1, 2; Tables 1, 2), with an average size of 947 nucleotides, using the coding-analysis program GLIMMER[9] (see Methods). Coding sequences cover 95% of the chromosome. Predicted protein sequences were searched against a non-redundant protein database and biological roles were assigned to 1,014 (54%)

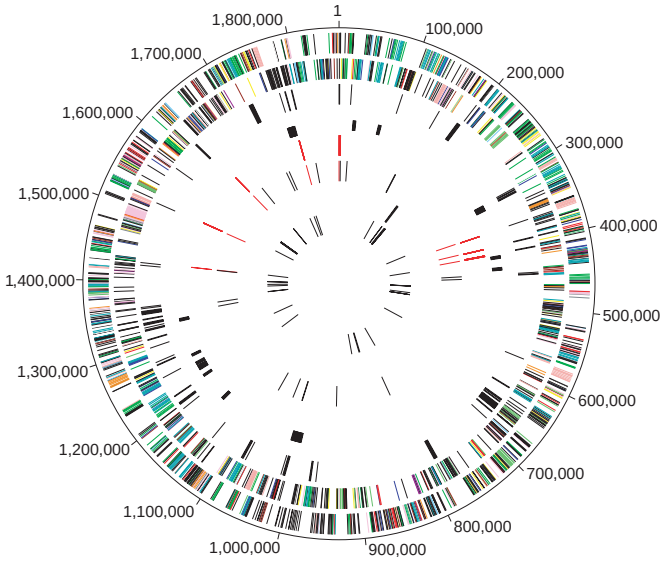


Figure 1 Circular representation of the *T. maritima* MSB8 genome showing predicted-coding regions and other features. Outer circle, predicted protein-coding regions on the plus strand classified by role according to the colour code in Fig. 2 (unknowns and hypotheticals are in black). Second circle, predicted protein-coding regions on the minus strand. Third circle, χ^2 composition in 2000-bp windows (see Methods); bands correspond to χ^2 values with $P \leq 1.9 \times 10^{-9}$. Fourth circle, Archaea-like islands on the genome. Fifth circle, small repeats. Sixth circle, large repeats (black), large repeats associated with small repeats (red). Seventh and eighth circles, rRNAs and tRNAs, respectively.

Table 1 General features of the *T. maritima* MSB8 genome

General features			
Length of sequence	1,860,725		
G + C ratio	46%		
Total no. of sequences	30,140		
Average read length (bp)	531		
Open reading frames	1,877		
Protein coding regions	95%		
Ribosomals	1 5S–16S–23S		
tRNAs	46 (10 clusters/19 single genes)		
Chromosomal coding sequences			
No. similar to known proteins	1,014		
No. of conserved hypotheticals	407		
No. similar to proteins of unknown function	83		
No. without a database match	373		
Total	1,877		
Repeats			
Class	Length	Copies	Database match
SR-01	30	143	tttccatacctctaaggaattattgaaaca
LR-01	1,897	2	hypothetical protein
LR-02	1,403	2	α-glucosidase
LR-03	1,137	4	putative transposase
LR-04	1,082	2	methyl-accepting chemotaxis protein
LR-05	858	2	putative transposase
LR-06	555	2	helicase
LR-07	252	2	excinuclease
LR-08	241	2	putative transposase

of them using the classification scheme adapted from ref. 10. Four-hundred-and-seven (22%) predicted coding sequences matched hypothetical coding sequences from other species, and 373 (20%) had no database match. Forty-six stable tRNAs with specificity for all 20 amino acids were identified.

Repeats. The *T. maritima* genome has 143 copies of a 30-bp repeat found in eight distinct clusters on the chromosome (Figs 1, 2). The 30-bp repeats are interspersed with a unique 39–40-bp sequence and are followed by a 452-bp sequence to form a structure identical to that described for the genomes of *M. jannaschii*[8] and *A. fulgidus*[5]. Our examination of the *Aquifex aeolicus* genome[11] reveals that similar repeats are present, but they are fewer in number, and shorter, than those in *T. maritima*. These small/large repeat structures have only been identified in the genome sequences of thermophiles.

In addition to the 30-bp repeat structures, eight classes of large repeats, more than 200 bp in length and with >95% identity to each other, are present in the *T. maritima* genome (Fig. 1, Table 1).

Multigene families. We identified 214 gene families ($P \leq 10^{-5}$ over 60% of the length of the sequence—see Methods) in *T. maritima*. Of these families, 126 consist of two members, and the largest gene family (of ATP-binding subunits of ABC transporters) contains 67 members. Two other large families (one with 22 members and one with 47 members) consist exclusively of proteins involved in transport. Fifteen families, unique to *T. maritima*, consist of proteins with no database match. Eleven of these families have two members, and the largest group has seven members (http://www.tigr.org/tdb/mdb/mdb.html).

Solute uptake and metabolism

Several transporters reflecting the heterotrophic metabolism of this species are present, including those for importing maltose (*malE*), ribose (*rbsB*) and spermidine and/or putrescine (*potD*), as well as several carbohydrate transporters whose specificity is unknown. Carriers for the uptake of amino acids and oligopeptides, and ion-transport systems for the acquisition of K^+ (*trkA/H*), Mg^{2+} (*mgtE*), NH_4^+ (*amt*), PO_4^{2-} (*pstA/C/B/S*) and both oxidized and reduced forms of iron (*feoA/B*) are present. The large number of transporters for carbohydrates and amino acids (Figs 3, 4) suggests that the environment in which *T. maritima* is found is rich in organic material.

The predominant mechanism of transport in *T. maritima* is ATP-coupled solute flux. Eighty-four percent of the proteins in the transport category are subunits of ATP-binding cassette (ABC) transporters. Phylogenetic comparisons of the periplasmic solute-binding protein (SBP) component (Fig. 4) roughly parallels the families defined in other Eubacteria[12] with a marked expansion of proteins specific for oligopeptides. Nine of the eleven oligopeptide systems appear to be in operons with genes essential for sugar metabolism (Fig. 5). Of the published genomes, only *Pyrococcus horikoshii*[13] has an oligopeptide transporter associated with a sugar degrading enzyme (β-galactosidase). This operonic structure suggests that there is coordinate regulation of peptide import with sugar degradation in these two organisms, although it is far more extensive in *T. maritima*. This contrasts with classical regulatory networks where the transported substrate affects transcription of the ABC transporter genes[14].

Almost 7% of the predicted coding sequences in the *T. maritima* genome are involved in metabolism of simple and complex sugars, more than twice the percentage seen in other eubacterial and archaeal species sequenced to date. Several genes encoding proteins involved in the sequential degradation of xylan are present. Genes encoding endoglucanases (*celA/B*) and β-glucosidases (*bglB*) that are involved in cellulose degradation were identified, confirming the presence of the cellulolytic systems predicted by biochemical studies of this organism[15]. *T. maritima* does not have a complex system for the degradation of plant cellulosic materials, as described for the thermophilic bacterium *Clostridium thermocellum*, in which the degradation of cellulose depends on a multienzyme complex known as the cellulosome, composed of between 14 and 26 subunits[16].

Glucose catabolism in *T. maritima* involves the Embden-Meyerhof and Entner-Doudoroff glycolytic pathways. In addition, the non-oxidative branch of the pentose-phosphate pathway appears to be involved in glucose breakdown. Genome analysis indicates that *T. maritima* can metabolize glycerol, gluconate and numerous sugars including amylose, maltose and galactose, as well as the amino acids aspartate, threonine and glycine (Fig. 3). CoA-SH-dependent ferredoxin oxidoreductases specific for pyruvate, as well as partial and complete operons for ferredoxin oxidoreductases of unknown specificity, are also present on the genome.

Biosynthetic pathways for nine amino acids were identified in *T. maritima* (Fig. 3). In addition, genes that encode proteins for the biosynthesis of biotin, folic acid, haem, porphyrin, lipoate, menaquinone, ubiquinone, pantothenate and pyridoxine were identified. *T. maritima* may also synthesize glycogen as a storage polysaccharide.

Along with an ability to gain energy through a fermentative metabolism, *T. maritima* can grow as a respiratory organism, generating energy in the presence of Fe(III) (ref. 17). Growth with sulphur as the terminal electron acceptor does not produce ATP[18],

Figure 2 Linear representation of the *T. maritima* MSB8 genome. The locations of each predicted protein-coding region (colour-coded by biological role), RNA genes, tRNAs and repeat elements are indicated. Arrows represent the direction of transcription for each predicted coding region. Numbers next to the tRNA symbols represent the number of tRNAs at a locus. Numbers next to GES represent the number of membrane-spanning domains predicted by the Goldman, Engelman, and Steitz scale as calculated by TopPred[45] for that protein. Only proteins with five or more GES domains are shown. Presumed transporter specificity is indicated above predicted coding regions identified as transporters. Transporter abbreviations are as follows: +, cations; H^+, protons; K^+, potassium; Pi, phosphate; Zn, zinc; aa, amino acids; Na^+, sodium; COH, sugar; aaX, oligopeptides; mal, maltose; rib, ribose; s/p, spermidine/putrescine; ura, uracil; ant, antibiotics; Fe^{2+}, iron(II); Fe^{3+}, iron (III); NH_4^+, ammonium; bcaa, branched chain amino acids; g3Pi, glycerol-3-phosphate; glyc, glycerol; chro, chromate; Mg^{2+}, magnesium; question marks (?) indicate where substrate specificity is uncertain or unknown. Members of paralogous gene families are identified by family number in a box above the predicted coding region.

but this pathway allows for the elimination of growth inhibitory H_2 which is produced during fermentative growth. Various flavoproteins and iron-sulphur proteins have been identified as potential electron carriers.

Response to environmental stimuli

T. maritima demonstrates a carbohydrate-dependent thermotactic response to temperature gradients between 50 and 105 °C (ref. 19).

Table 2 *T. maritima* MSB8 gene list

Gene identification numbers that correspond to those in Fig. 2 are listed here with the prefix TM followed by the common name assigned to each protein, the three- or four-letter gene name in parentheses, the organism with the most significant match in braces and the percent similarity to the best match. Each gene identified is listed in its functional role category (adopted from ref. 10). In cases where the substrate specificity of a protein could not be unambiguously determined, a more general common name was used and no gene name was assigned. In some cases a gene without known substrate specificity could be confidently assigned to a particular family as found in PROSITE (*http://expasy.hcuge.ch/sprot/prosite.html/*) or SWISS-PROT (*http://expasy.hcuge.ch/sprot/*). The term 'related' is used in two ways in the common names: (1) when the TM protein is a partial but significant match to a database protein: the TM protein is assigned to the Unknown role category; or (2) when the TM protein is a very good match to a database protein whose function is not found in *T. maritima*: the TM protein may be assigned to a role category appropriate to the known function of the database match. Abbreviations are as follows: Common names: AA, amino acid; NH3, ammonia; NH4+, ammonium; AFS, authentic frameshift; APM, authentic point mutation; Bprt, binding protein; BRAA, branched chain AA; Cl–, chloride; CoA, coenzyme A; DHase, dehydrogenase; dep, dependent; elong, elongation; fam, family; flgr, flagellar; init, initiation; Fe, iron; Fe3+, iron(III); Fe2+, iron(II); LPS, lipopolysaccharide; Mg2+, magnesium; Mn2+, manganese; OP, oligopeptide; PPase, phosphatase; P, phosphate; PPR, phosphoribosyl; K+, potassium; prt, protein; put, putative; RDase, reductase; reg, regulation, regulator, regulatory; rel, related; ssDNA, single stranded DNA; Na+, sodium; S, sulfur; sub, subunit; Sase, synthase, synthetase; term, termination; Tase, transferase; transp, transporter; Zn, zinc. Organism of best match: Ac, *Acinetobacter calcoaceticus*; Ab, *Agaricus bisporus*; Ar, *Agrobacterium radiobacter*; At, *Agrobacterium tumefaciens*; Ae, *Alcaligenes eutrophus*; Alc, *Alcaligenes* sp.; Alt, *Alteromonas* sp.; Amy, *Amycolata* sp.; Ana, *Anabaena* sp.; Ath, *Anaerocellum thermophilum*; Aa, *Aquifex aeolicus*; Atl, *Arabidopsis thaliana*; Af, *Archaeoglobus fulgidus*; Art, *Arthrobacter* sp.; Aca, *Azorhizobium caulinodans*; Av, *Azotobacter vinelandii*; Bbv, *Bacillus brevis*; Bca, *Bacillus caldolyticus*; Bce, *Bacillus cereus*; Bci, *Bacillus circulans*; Bf, *Bacillus firmus*; Bl, *Bacillus licheniformis*; Bm, *Bacillus megaterium*; Bac, *Bacillus* sp.; Bsp, *Bacillus sphaericus*; Bst, *Bacillus stearothermophilus*; Bs, *Bacillus subtilis*; T4, Bacteriophage T4; Bn, *Bacteroides nodosus*; Bt, *Bacteroides thetaiotaomicron*; Bv, *Beta vulgaris*; Bp, *Bordetella pertussis*; Bb, *Borrelia burgdorferi*; Bbr, *Brevibacillus brevis*; Bli, *Brevibacterium linens*; Ba, *Brucella abortus*; Ce, *Caenorhabditis elegans*; Cj, *Campylobacter jejuni*; Ca, *Candida albicans*; Ch, *Capra hircus*; Cc, *Caulobacter crescentus*; Cp, *Chlamydia psittaci*; Cr, *Chlamydomonas reinhardtii*; Cau, *Chloroflexus aurantiacus*; Cac, *Clostridium acetobutylicum*; Cla, *Clostridium acidiurici*; Chi, *Clostridium histolyticum*; Cpa, *Clostridium pasteurianum*; Cpe, *Clostridium perfringens*; Clo, *Clostridium* sp.; Ct, *Clostridium thermocellum*; Cam, *Cornyebacterium ammoniagenes*; Dd, *Desulfovibrio desulfuricans*; Df, *Desulfovibrio fructosovorans; Dg, Desulfovibrio gigas*; Dv, *Desulfovibrio vulgaris*; Dt, *Dictyoglomus thermophilum*; Eco, *Eikenella corrodens*; Ea, *Enterobacter aerogenes*; Ef, *Enterococcus faecalis*; Ech, *Erwinia chrysanthemi*; Ec, *Escherichia coli*; Eac, *Eubacterium acidaminophilum*; Eg, *Euglena gracilis*; Fi, *Fervidobacterium islandicum*; Hi, *Haemophilus influenzae*; Hp, *Helicobacter pylori*; Hs, *Homo sapiens*; Kp, *Klebsiella pneumoniae*; Lf, *Lactobacillus fermentum*; Lp, *Lactobacillus pentosus*; Ll, *Lactococcus lactis*; Lpn, *Legionella pneumophila*; Lm, *Leptospira meyeri*; Lmo, *Listeria monocytogenes*; Mc, *Mesembryanthemum crystallinum*; Mta, *Methanobacterium thermoautotrophicum*; Mtf, *Methanobacterium thermoformicicum*; Mj, *Methanococcus jannaschii*; Mk, *Methylobacterium extorquens*; Mlu, *Micrococcus luteus*; Ma, *Microcystis aeruginosa*; Mo, *Micromonospora olivasterospora*; Mmu, *Mus musculus*; Ml, *Mycobacterium leprae*; Ms, *Mycobacterium smegmatis*; Mt, *Mycobacterium tuberculosis*; Mg, *Mycoplasma genitalium*; Mp, *Mycoplasma pneumoniae*; Mx, *Myxococcus xanthus*; Nf, *Naegleria fowleri*; Ng, *Neisseria gonorrhoeae*; Nos, *Nostoc* sp.; Pd, *Paracoccus denitrificans*; Pha, *Pasteurella haemolytica*; Pan, *Podospora anserina*; Pg, *Porphyromonas gingivalis*; Pm, *Propionigenium modestum*; Pa, *Pseudomonas aeruginosa*; Pc, *Pseudomonas cichorii*; Pf, *Pseudomonas fluorescens*; Pp, *Pseudomonas putida*; Pse, *Pseudomonas* sp.; Ps, *Pseudomonas stutzeri*; Psy, *Pseudomonas syringae*; Pfu, *Pyrococcus furiosus*; Ph, *Pyrococcus horikoshii*; Pyr, *Pyrococcus* sp.; Rn, *Rattus norvegicus*; Ra, *Reclinomonas americana*; Rc, *Rhodobacter capsulatus*; Sc, *Saccharomyces cerevisiae*; Se, *Saccharopolyspora erythraea*; Sch, *Salmonella choleraesuis*; Sd, *Shigella dysenteriae*; Sa, *Staphylococcus aureus*; Sx, *Staphylococcus xylosus*; Sau, *Stigmatella aurantiaca*; Sb, *Streptococcus bovis*; Sm, *Streptococcus mutans*; Sp, *Streptococcus pneumoniae*; St, *Streptococcus thermophilus*; Sco, *Streptomyces coelicolor*; Ss, *Sulfolobus solfataricus*; Ssc, *Sus scrofa*; Syn, *Synechococcus* sp.; SPCC, *Synechocystis* PCC6803; Scy, *Synechocystis* sp.; Tb, *Thermoanaerobacter brockii*; Tth, *Thermoanaerobacterium thermosaccharolyticum*; Tl, *Thermococcus litoralis*; Tm, *Thermotoga maritima*; Tn, *Thermotoga neapolitana*; Ta, *Thermus aquaticus*; Tt, *Thermus thermophilus*; Td, *Treponema denticola*; Tp, *Treponema pallidum*; Va, *Vibrio alginolyticus*; Vc, *Vibrio cholerae*; Vp, *Vibrio parachaemolyticus*; Ws, *Wolinella succinogenes*; Xl, *Xenopus laevis*; Ye, *Yersinia enterocolitica*.

Motility is regulated by a two-component histidine kinase signal-transduction pathway that is assembled from products of the Che genes (*cheA/B/C/D/R/W/Y*) and seven methyl-accepting chemotactic transducer proteins (MCPs). Genes encoding *T. maritima* MCPs are most closely related to homologues in *Bacillus subtilis*, with specificity for both amino acids and carbohydrates. Although *T. maritima* demonstrates no chemotactic response to serine[19], the presence of amino-acid-specific MCPs indicates that *T. maritima* may respond to aspartate, threonine and glycine, which appear to be catabolized by this bacterium.

In addition to the chemotactic-response kinases, other members of the two-component signalling family identified in *T. maritima* are likely to have a role in monitoring and responding to environmental stimuli such as temperature and nutrients. This group of six signalling partners includes the previously identified histidine-kinase (*hpkA*)/response-regulator (*drrA*) pair belonging to the OmpR–PhoB subfamily of transcriptional regulators[20].

Although no heat- or cold-shock response has been experimentally demonstrated in *T. maritima*, the genome contains genes encoding heat-shock proteins (*dnaJ/K*, *groEL/ES*, *grpE*, *hslU/V* and *hsp*) and cold-shock proteins (encoded by *cspB/L*), which probably help regulate responses to changes in ambient temperature and growth conditions. *T. maritima* also contains genes encoding a general stress protein (*ctc*) and a stationary-phase-survival protein (*surE*), both of which are presumably involved in stress survival.

Cellular activities

Protein secretion, competence and transformation. In addition to the Sec-A-dependent secretory pathway, *T. maritima* has two specialized export systems. The first, which is assembled from homologues of FliH/P/R and FlhA/B[21], is probably associated with the secretion and assembly of the single *T. maritima* flagellum[19]. The second is a type-II secretion pathway that is assembled from the general secretion pathway proteins D, E and F (homologues of the PulD transport family of *Klebsiella oxytoca*[22]). The *T. maritima* type-II secretion pathway probably serves as the primary mechanism for secretion of degradative enzymes required for the utilization of polysaccharides.

Competence has not been demonstrated in *T. maritima*, but it is possible that the type-II general secretion pathway proteins D, E and F, and the type IV pilin leader peptidase, type IV pilin-related protein and *pilT* identified in *T. maritima* may function in natural competence and transformation, as described for other organisms[23]. In addition, *T. maritima* has homologues of the competence genes *dprA*, *comM* and *comE* of *Haemophilus influenzae*, and *comEA* and *comFC* of *B. subtilis*. There are other protein-coding sequences with weak homology to competence genes from *B. subtilis*. This suggests that there may be an inherent system for the uptake of exogenous DNA, facilitating genetic exchange between *T. maritima* and other organisms.

Transcription and translation. Genes encoding the three subunits (α, β, β') of the core RNA polymerase were identified in *T. maritima* (*rpoA/B/C*, respectively) along with four σ factors; σ^A (also named σ^{70}) (*rpoD*), σ^E (*rpoE*), σ^H (*fliA*) and σ^{28} (*spoOH*). Although σ^A has been previously identified in *T. maritima*[24], the roles and specificity of the remaining σ factors in transcription regulation are unknown. The *nusA/B/G* transcription antitermination genes and a member of the *greA/B* transcription-elongation-factor family were identified along with the *rho* transcription-termination factor.

The genome of *T. maritima* is similar to other sequenced bacterial genomes in that the gene for glutaminyl tRNA-synthetase is missing. An alternative synthesis mechanism is the transamidation of Glu-tRNA^{Gln} to Gln-tRNA^{Gln} by Glu-tRNA^{Gln} amidotransferase, a heterotrimeric enzyme found in both Eubacteria and Archaea[25]. Like *Helicobacter pylori*[26], *T. maritima* lacks an asparaginyl-tRNA synthetase. Transamidation of Asp-tRNA^{Asn} is presumably involved in generation of Asn-tRNA^{Asn}, similar to that reported in the archaeon *Haloferax volcanii*[27].

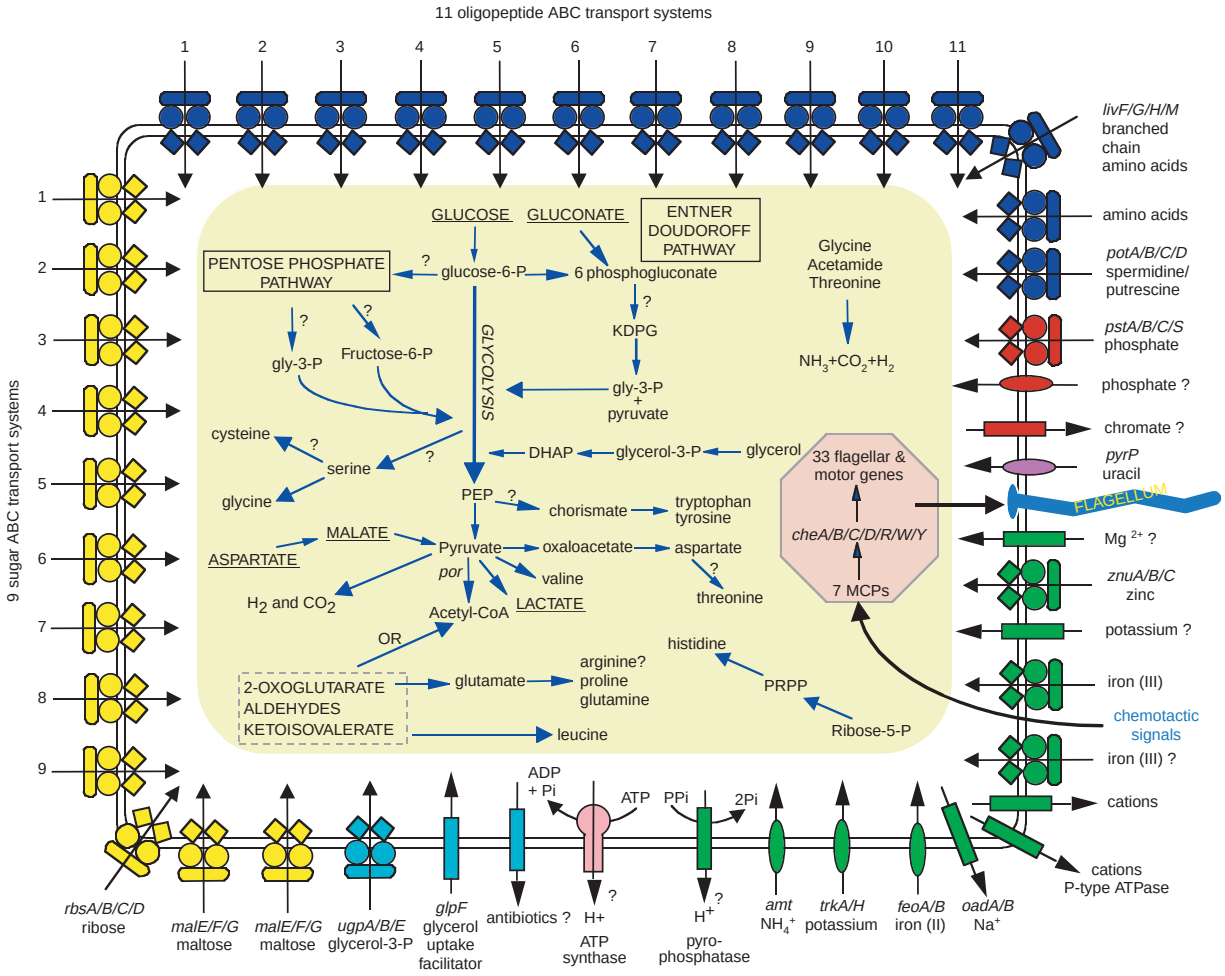


Figure 3 Overview of metabolism and transport in *T. maritima* MSB8. Pathways for energy production and the metabolism of organic compounds, acids and aldehydes are shown. Each gene product with a predicted function in ion or solute transport is illustrated. Transporters are grouped by substrate specificity according to role category: cations (green), anions (red), carbohydrates (yellow), purines (purple), amino acids/peptides/amines (dark blue) and other (light blue). Question marks associated with transporters indicate uncertainties in substrate specificity or direction of transport. Question marks associated with metabolic pathways indicate where an expected activity was not found. Permeases are represented by ovals. ABC transport systems are shown as composite figures of ovals, diamonds and circles. All other transporters are drawn as rectangles. Export or import of solutes is designated by the direction of the arrows through the transporter. If precise substrate specificity could not be determined for a transporter, no gene name was assigned and a more general common name reflecting the type of substrate being transported was used. Abbreviations: PRPP, phosphoribosyl-pyrophosphate; gly, glyceraldehyde; PEP, phosphoenolpyruvate; ATP, adenosine triphosphate; ADP, adenosine diphosphate; DHAP, dihydroxyacetone phosphate; OR, oxidoreductases; MCP, methyl-accepting chemotaxis protein; KDPG, 2-keto-3-deoxy-6-phosphogluconate; *por*, pyruvate oxidoreductase.

Cell division. The gene content of *T. maritima* demonstrates that the basic mechanism of cell division is similar to that found in other Eubacteria. The *ftsA/H/Y/Z* genes were identified along with two genes from the min locus, *minC/D*. These genes function together to specify the position and formation of the constricting ring that leads to final division.

Detoxification. Proteins for detoxification in this strict anaerobe include two NADH oxidases (*nox*), a putative alkyl hydroperoxide reductase, a heavy-metal-resistance transcriptional regulator and the periplasmic divalent cation-tolerance protein (*cutA*).

Phylogenetics and comparative genomics

The availability of the complete genome sequence of *T. maritima* is important for evolutionary studies, as *T. maritima* has been suggested to be one of the deepest branching eubacterial species, on the basis of phylogenetic analysis of SSU rRNA genes[3]. Although SSU rRNA analysis has been useful in identifying and characterizing bacterial strains and species, many questions remain regarding the validity of evolutionary conclusions based on SSU rRNA analysis (see, for example, ref. 28).

To capitalize on the information contained in completed genome sequences, and to reduce complications caused by different species sets in phylogenetic analyses of different genes[29], we identified a subset of 33 genes (see Methods) for which homologues were conserved in all species sequenced to date[4,5,8,11,13,26,30–39] (Table 3). This subset was used to generate multiple sequence alignments and phylogenetic trees using both parsimony and distance methods (see Methods).

A few significant patterns arise from the analysis of these phylogenies. First, for the majority of genes, the Archaea constitute a distinct monophyletic group separate from the Eubacteria, a phylogenetic pattern also found in SSU rRNA trees. However, this finding does not relate to whether the Archaea as a whole are monophyletic, as the species of Archaea represented here are all *Euryarchaeota*. Second, most yeast genes group with the archaeal genes as found for SSU rRNA. In addition, in almost all trees, species that are part of the same bacterial phyla group together (for example, *Escherichia coli* and *H. influenzae*, *Borrelia burgdorferi* and *Treponema pallidum*). Beyond this, there are significant differences in the topologies of different genes, and there is little

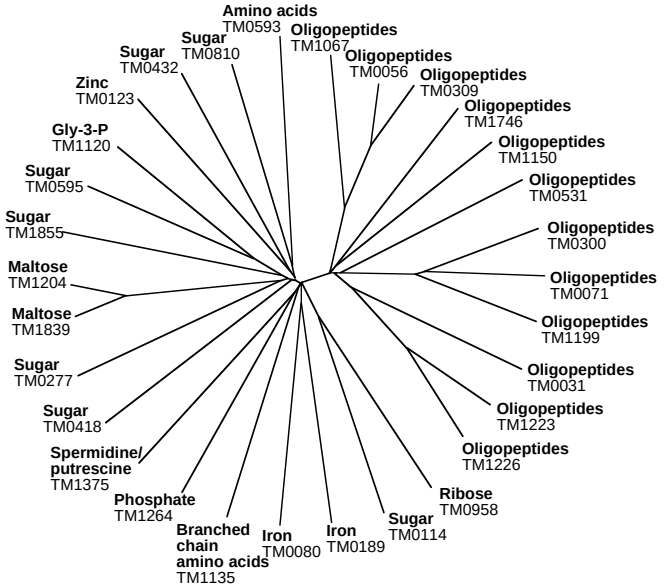


Figure 4 Phylogenetic pattern of the periplasmic SBP component of the oligopeptide transporters and other transporters present on the *T. maritima* MSB8 genome. Sperm/putre, spermidine/putrescine; Gly-3-P, glycerol-3-phosphate; sugar, unsure of specificity of sugar transporter.

agreement between these gene trees and the rRNA tree. In particular, we find little support for the rRNA-based positions of *Aquifex* and *Thermotoga*.

The lack of congruence of trees for different genes could result from the limited number of species represented by the completed genomes and/or the small size of some of these genes. However, we believe that the differences are real both because of high bootstrap support and because differences in topology have been found by others in the analysis of other genes[28]. Mechanisms that could lead to these differences include gene duplication, gene loss and horizontal gene transfer. Thus we conclude that, based on single-gene analysis, the phylogenetic position of *Aquifex* and *Thermotoga*, and the nature of the deepest branching eubacterial species, should be considered to be ambiguous.

Phylogenetic analysis of individual genes has limited evolutionary studies to a small percentage of the genes in any given genome. As an alternative to single-gene phylogenetic analysis, we compared *T. maritima*'s genome sequence to those of the completely sequenced microbial species by observing patterns of similarity (see Methods). Of the 1,877 predicted coding sequences in the *T. maritima* genome, 52% are most similar to proteins in eubacterial species, most to the Gram-positive bacterium *B. subtilis* (21%) and to *A. aeolicus* (15%). In addition, 24% of the predicted coding sequences are most similar to proteins in archaeal species (Table 4), almost half to *P. horikoshii*. This similarity of *T. maritima* to the Archaea contrasts with the other Eubacteria in which no more than 16% of the coding sequences in *A. aeolicus*, and 7% of the coding sequences in *B. subtilis*, are most similar to archaeal proteins. By whole-genome similarity comparison, *T. maritima* appears to be the most Archaea-like of all sequenced Eubacteria[4,11,26,31–38].

The Archaea-like nature of the *T. maritima* genome does not necessarily reflect a closely shared common ancestor between *T. maritima* and the Archaea, as the extensive similarity between these species could have arisen in many ways. These include loss of these genes in other lineages, extensive sequence divergence in mesophilic Eubacteria (coupled with a retention of such genes in the thermophilic Archaea and *T. maritima*) and, alternatively, that a few genes in the *T. maritima* genome with strong similarity to archaeal genes have expanded into large gene families, resulting in more genes with a higher level of similarity to archaeal genes. Such mechanisms could lead to similarity between Archaea and *T. maritima* regardless of the evolutionary history of these species.

We believe, however, that much of the similarity between *T. maritima* and the Archaea is due to shared ancestry of portions of the genome as a result of extensive lateral gene transfer between these lineages (Tables 3, 4). One line of evidence consistent with previous studies which have argued in support of lateral transfer[40] is that the 451 Archaea-like genes in *T. maritima* are not uniformly distributed among the biological role categories (Table 4). The majority of genes involved in housekeeping functions such as transcription, translation, DNA replication and cell division are most similar to orthologues in eubacterial species. In contrast, 49% of transporters (92), 60% of electron transport proteins (28) and 42% of conserved hypothetical proteins (173) are most similar to archaeal genes. Another observation which would support lateral gene transfer is that 81 of the Archaea-like genes in the *T. maritima* genome are clustered in 15 regions of the chromosome that range in size from ~4 to 20 kb. The genes, and conservation of gene order in seven of these regions, have only been described in the genome sequences of thermophilic Archaea. In addition, two of the clustered regions are associated with the 30-bp repeat elements found among Archaea and *T. maritima*, lending support to the idea that these repeat elements may be involved in gene transfer.

χ^2-analysis of the genome sequence (see Methods) lends additional support to the theory of lateral gene transfer in *T. maritima*. Based on the assumption that the DNA composition is relatively uniform throughout the genome, there are at least 51 regions in the chromosome (Fig. 1) that have a significantly different composition ($P \leq 1.9 \times 10^{-9}$). Forty-two of these regions include genes and repeat structures that have highest levels of similarity to regions on the chromosomes of other thermophiles, including the thermophilic Archaea and *Aquifex*. All of the 30-bp small repeat areas have a χ^2 composition that is substantially different from the rest of the

Table 3 Genes shared by the *T. maritima* MSB8 genome

33 homologues in the completed genomes *apt*, *argS*, *eno*, *ftsZ*, *gcp*, *gltX*, *groES*, *hisT*, *pheS*, *pgk*, *prs rplA*, *rplB*, *rplC*, *rplE*, *rplF*, *rplK*, *rplN*, *rplR*, *rplV*, *rpsB*, *rpsC*, *rpsD*, *rpsE*, *rpsG*, *rpsH*, *rpsJ*, *rpsK*, *rpsL*, *rpsM*, *rpsQ*, *rpS*, *secY*.
T. maritima genes matching only hyperthermophiles: 108 total; 93 hypothetical proteins; *flgA* putative; s-layer-related protein; rubrerythrin putative; 2 ABC transporters; glutamate synthase-related protein; glutaredoxin putative; putative glycerate kinase; putative hydrogenase; putative NADH dehydrogenase; putative LPS biosynthesis protein; putative phosphonopyruvate decarboxylase; putative pyruvate formate lyase activating enzyme; alanine acetyltransferase-related protein; sensory box protein.
T. maritima genes matching only Archaea 71 total; 64 hypothetical proteins; 2 ABC transporters; glutamate synthase-related protein; putative glycerate kinase; putative hydrogenase; rubrerythrin; sensory box protein.

Table 4 Top eubacterial or archaeal match in *T. maritima* by role ID

Role category	Eubacteria	Archaea	Eukaryotes	None
Amino acid biosynthesis	49	20	2	1
Purines, pyrimidines, etc.	32	11	2	0
Fatty acid and phospholipid metab.	12	3	0	0
Biosynthesis of cofactors etc.	22	10	0	0
Central intermediary metabolism	27	12	1	3
Autotrophic metabolism	0	0	0	0
Energy metabolism	117	56	1	18
Transport	89	92	0	7
DNA metabolism	45	6	0	2
Transcription	17	0	0	0
Translation	118	6	0	6
Regulatory functions	61	9	0	0
Cell envelope	57	9	1	7
Cellular processes	55	10	0	0
Other	9	8	0	1
Hypotheticals	213	173	2	19
Unknown	52	26	0	5
TOTAL	975	451	9	69

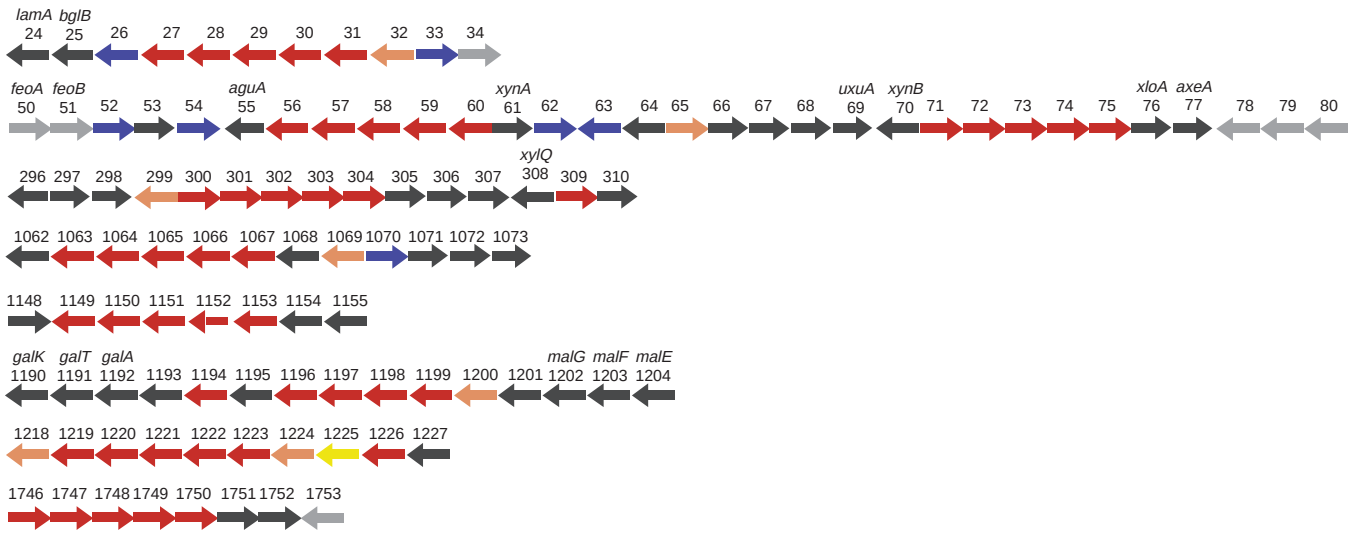


Figure 5 Linear representation of the location of nine oligopeptide transporter operons on the *T. maritima* MSB8 genome. Arrows represent functional categories as follows: black, sugar metabolism; red, transporters; orange, regulators; blue, hypothetical; yellow, conserved hypothetical; grey, other. *aguA*, α-glucuronidase; *axeA*, acetyl xylan esterase; *bglB*, β-glucosidase; *feoA*, iron(II) transport protein A; *feoB*, iron(II) transport protein B; *galA*, α-galactosidase; *galK*, galactokinase; *galT*, galactose-l-phosphate uridylyltransferase; *lamA*, laminarinase; *malE/F/G*, maltose ABC transporter; *uxuA*, D-mannonate hydrolase; *xloA*, xylosidase; *xylQ*, α-xylosidase; *xynA*, endo-1,4-β-xylanase A; *xynB*, endo-1,4-β-xylanase B. Arrows are not proportional to the size of predicted-coding regions.

genome. Compositional bias as seen in these χ^2 distributions has previously been used in support of lateral gene transfer[41].

In summary, completion of the sequence of the *T. maritima* genome has revealed a degree of similarity with the Archaea in terms of gene content and overall genome organization that was not previously appreciated. Although the core of *T. maritima* may be eubacterial, almost one quarter of the genome is archaeal in nature. The mosaic nature of *T. maritima* appears to be the result of extensive lateral gene transfer with the Archaea. The accumulating evidence for the high frequency of lateral transfer events, and the lack of congruence among the phylogenies of different genes, indicates that the emphasis of phylogenetic studies on individual genes as indicators of organismal evolution is probably not accurate. Undoubtedly, our understanding of the complex relationships among prokaryotes will continue to increase as other microbial genome sequencing projects are completed. □

Methods

Whole-genome random sequencing procedure. The type strain, *T. maritima* MSB8, was grown from a culture derived from a single cell isolated by optical tweezers and provided by R. Huber (University of Regensburg). Cloning, sequencing and assembly were as described for genomes sequenced by TIGR[4,5]. One small-insert plasmid library (1.5–2.5 kb) was generated by random mechanical shearing of genomic DNA. One large-insert lambda library was generated by partial *Tsp*5091 digestion and ligation to λ-DASHII/*Eco*RI vector (Stratagene). In the initial random sequencing phase, ~7-fold sequence coverage was achieved with 27,789 sequences from plasmid clones (average read length, 531 bases). The plasmid and λ-sequences were jointly assembled using TIGR Assembler[42]. Sequences from both ends of 546 λ-clones served as a genome scaffold, verifying the orientation, order and integrity of the contigs. Sequence gaps were closed by editing the ends of sequence traces and/or primer walking on plasmid clones. Physical gaps were closed by direct sequencing off genomic DNA, or combinatorial PCR followed by sequencing of the PCR product. The final genome sequence is based on 30,140 sequences.

Paralogous gene families were constructed by searching the ORFs against themselves using BLASTX[43], identifying pairwise matches above $P \leq 10^{-5}$ over 60% of the query search length, and subsequently clustering these matches into multigene families. Multiple sequence alignments for these protein families were generated using MSA (G. G. Sutton & T. Bussey, personal communication), an annealing algorithm, and the alignments were scrutinized.

For the comparative genomics, the *T. maritima* ORFs were added to a set of all ORFs from 16 published microbial genomes[4,5,8,11,13,26,30–39]. This dataset was searched against itself using BLASTX[43], and pair-wise matches were identified, clustered and converted to multiple sequence alignments as above. From this set of alignments a subset of 33 homologous gene families ($P \leq 10^{-5}$ over 60% of the query search length) were identified. These alignments were enhanced using the CLUSTAL X program, and regions of the alignments that were ambiguous, hypervariable or that contained large alignment gaps were excluded from subsequent phylogenetic analysis. Phylogenetic trees were generated from the curated alignments using the PAUP 4.0.0d64 and PHYLIP computer programs. Parsimony analysis was conducted using the *heuristic search* algorithm of PAUP and *protpars* of PHYLIP. Distance-based trees were generated using the neighbour-joining algorithm of ref. 44. One hundred bootstrap replicates were conducted for all trees. For all PAUP analyses, multiple step matrices were used in calculation of distances and in parsimony analysis. The whole genome set of pairwise BLASTX[43] search results was also used to determine 'best hits' of *T. maritima* and *A. aeolicus* to other genomes.

For the χ^2 analysis we computed the distribution of all 64 trinucleotides (3mers) for the complete genome, and then computed the 3mer distribution in 2,000 bp windows across the genome. We used windows that overlapped by half their length; that is, 1,000 bp. For each window, we computed the χ^2 statistic on the difference between its 3mer content and that of the whole genome. A large value of this statistic means that the composition within the window is different from the rest of the genome. Figure 1 illustrates those regions of the genome with χ^2 values whose probability is less than 1.9×10^{-9}; the probability values for these regions are based on the assumption that the DNA composition is relatively uniform throughout the genome. Because this assumption may be incorrect, we prefer to interpret high χ^2 values merely as indicators of regions on the chromosome that appear unusual and that demand further scrutiny.

ORF prediction and gene family identification. An initial set of ORFs likely to encode proteins was identified by GLIMMER[9], and those shorter than 30 codons were eliminated. ORFs that overlapped were visually inspected, and in some cases removed. ORFs were searched against a non-redundant protein database as previously described[4]. Frameshifts and point mutations were detected and corrected where appropriate as described previously[26]. Remaining frameshifts and point mutations are considered to be authentic and corresponding regions were annotated as 'authentic frameshift' or 'authentic point mutation' respectively. Two sets of hidden Markov models (HMMs) were used to determine ORF membership in families and superfamilies. These included 527 HMMs from pfam v2.0, and 199 HMMs from the TIGR orthologue

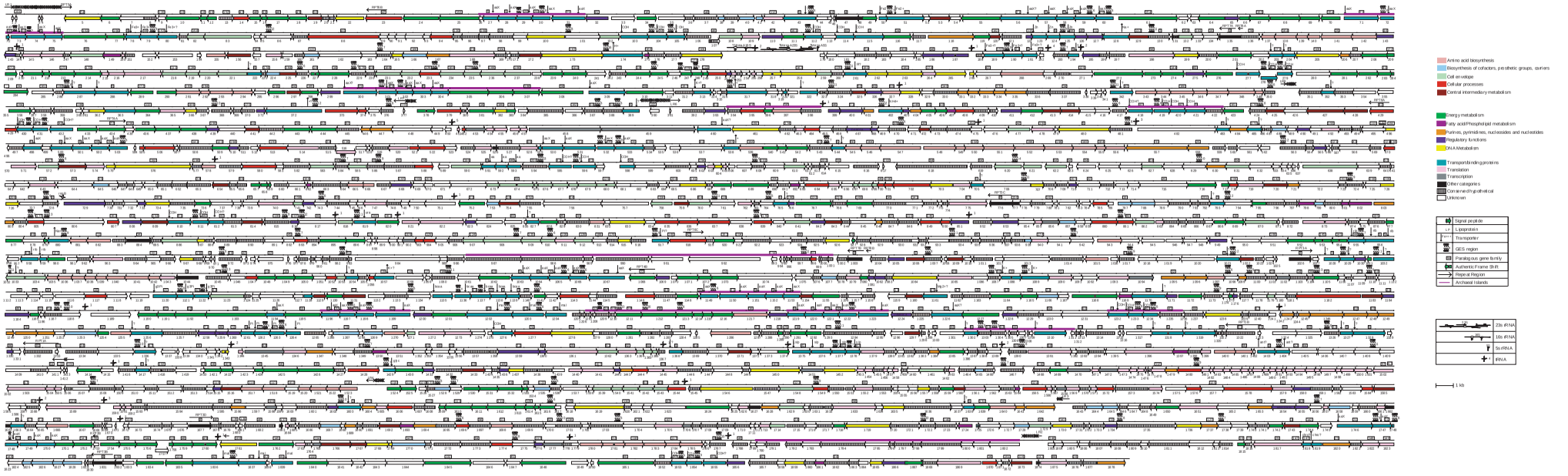
resource. TopPred[45] was used to identify membrane-spanning domains (MSDs) in proteins.

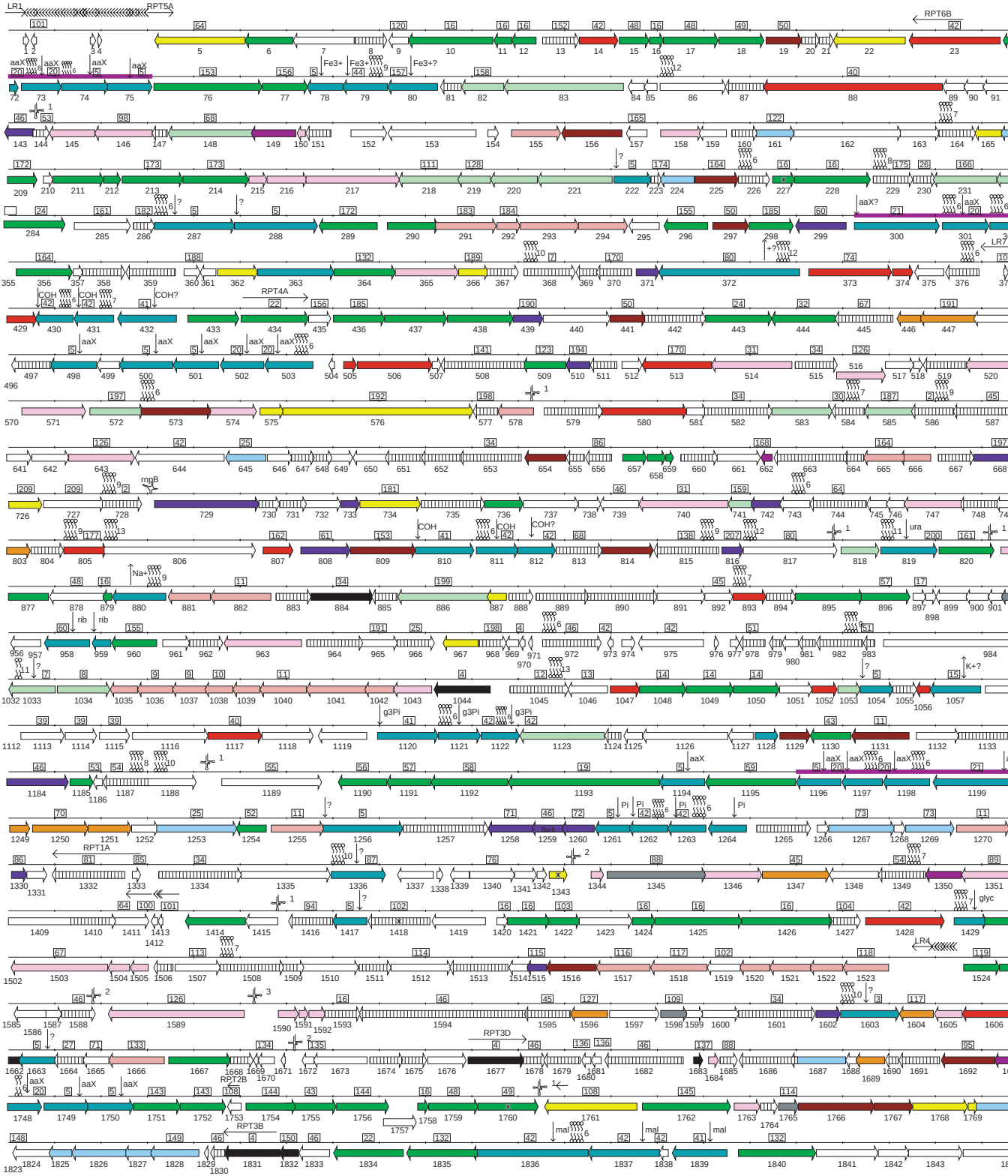
Received 9 November 1998; accepted 31 March 1999.

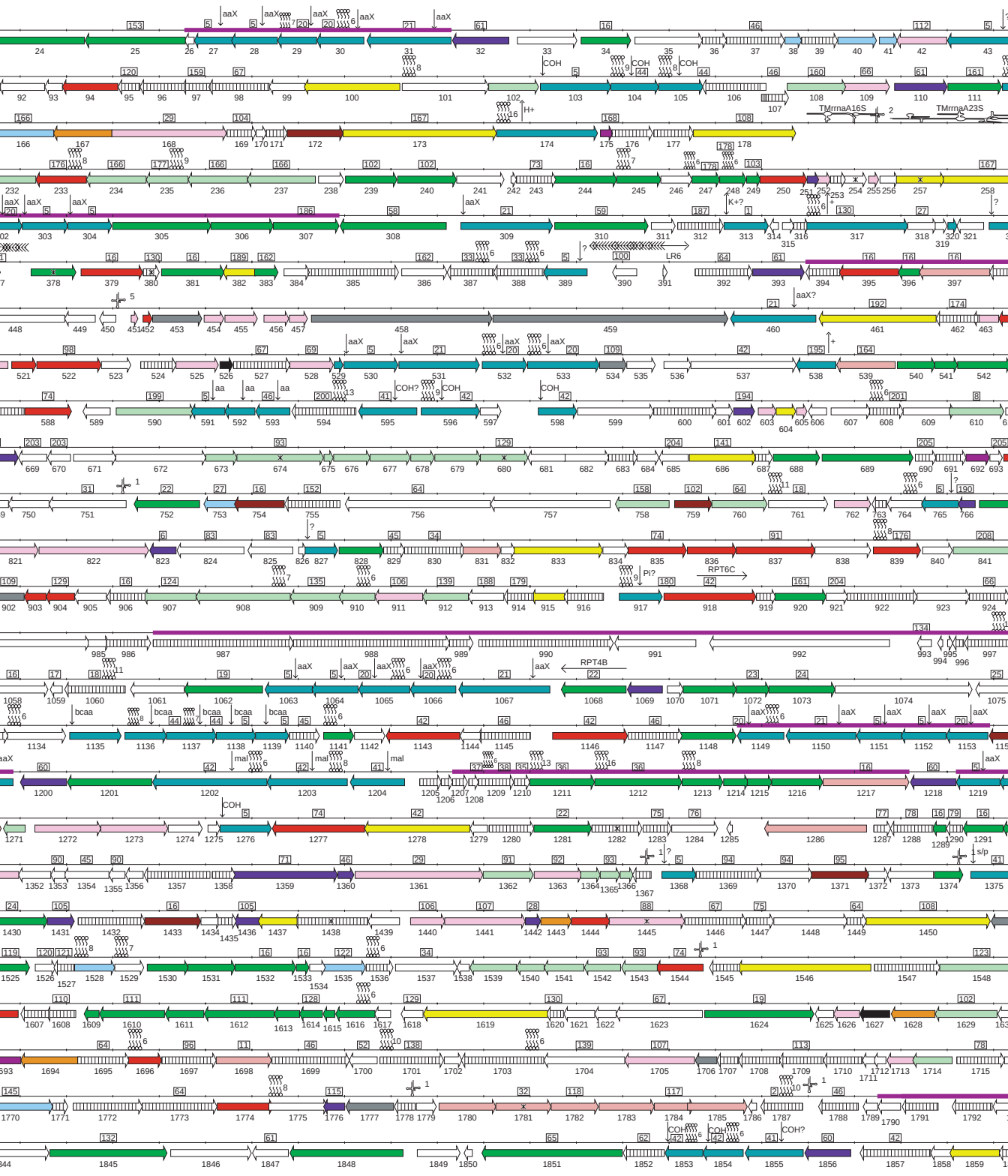
1. Huber, R. *et al*. *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90 °C. *Arch. Microbiol.* **144,** 324–333 (1986).
2. Huber, R. & Stetter, K. O. in *The Prokaryotes* (eds Balows, A. *et al.*) 3809–3815 (Springer, Berlin, Heidelberg, New York, 1992).
3. Achenbach-Richter, L., Gupta, R., Stetter, K. O. & Woese, C. R. Were the original eubacteria thermophiles? *Syst. Appl. Microbiol.* **9,** 34–39 (1987).
4. Fleischmann, R. D. *et al*. Whole-genome random sequence of *Haemophilus influenzae* Rd. *Science* **269,** 496–512 (1995).
5. Klenk, H.-P. *et al*. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390,** 364–370 (1997).
6. Lobry, J. R. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* **13,** 660–665 (1996).
7. Salzberg, S., Salzberg, A., Kerlavage, A. & Tomb, J.-F. Skewed oligomers and origins of replication. *Gene* **217,** 57–67 (1998).
8. Bult, C. J. *et al*. Complete genome sequence of the methanogenic archaeon *Methanococcus jannaschii*. *Science* **273,** 1058–1073 (1996).
9. Salzberg, S. L., Delcher, A. L., Kasif, S. & White, O. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* **15,** 544–548 (1998).
10. Riley, M. Functions of gene products of *Escherichia coli*. *Microbiol. Rev.* **57,** 862–952 (1993).
11. Deckert, G. *et al*. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392,** 353–358 (1998).
12. Saier, M. H. Jr Computer-aided analyses of transport protein sequences: gleaning evidence concerning function, structure, biogenesis, and evolution. *Microbiol. Rev.* **58,** 71–93 (1994).
13. Kawarabayasi, Y. *et al*. Complete sequence and gene organization of the genome of a hyperthermophilic archaebacterium, *Pyrococcus horikoshii* OT3. *DNA Res.* **5,** 55–76 (1998).
14. Boos, W. & Lucht, J. M. in Escherichia coli *and* Salmonella *Cellular and Molecular Biology* (eds Neidhardt, F. C. *et al.*) 1175–1209 (ASM, Washington, 1996).
15. Bronnenmeier, K., Kern, A., Liebl, W. & Staudenbauer, W. L. Purification of *Thermotoga maritima* enzymes for the degradation of cellulosic materials. *Appl. Environ. Microbiol.* **61,** 1399–1407 (1995).
16. Felix, C. R. & Ljungdahl, L. O. The cellulosome; the exocellular organelle of *Clostridium*. *Annu. Rev. Microbiol.* **47,** 791–819 (1993).
17. Vargas, M., Kashefi, K., Blunt-Harris, E. L. & Lovley, D. R. Microbiological evidence for Fe(III) reduction on early Earth. *Nature* **395,** 65–67 (1998).
18. Janssen, P. H. & Morgan, H. W. Heterotrophic sulfur reduction by *Thermotoga* sp. strain FjSS3.B1. *FEMS Microbiol. Lett.* **75,** 213–217 (1992).
19. Gluch, M. F., Typke, D. & Baumeister, W. Motility and thermotactic responses of *Thermotoga maritima*. *J. Bacteriol.* **177,** 5473–5479 (1995).
20. Lee, P. J. & Stock, A. M. Characterization of the genes and proteins of a two-component system from the hyperthermophilic bacterium *Thermotoga maritima*. *J. Bacteriol.* **178,** 5579–5585 (1996).
21. Macnab, R. M. in Escherichia coli *and* Salmonella *Cellular and Molecular Biology* (eds Neidhardt, F. C. *et al.*) 123–145 (ASM, Washington, 1996).
22. Hueck, C. J. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol. Mol. Biol. Rev.* **62,** 379–433 (1998).
23. Dubnau, D. Binding and transport of transforming DNA by *Bacillus subtilis*: the role of type IV pilin-like proteins—a review. *Gene* **11,** 191–198 (1997).
24. Gruber, T. M. & Bryant, D. A. Molecular systematic studies of eubacteria, using sigma70-type sigma factors of group 1 and group 2. *J. Bacteriol.* **179,** 1734–1747 (1997).
25. Zhang, J., Hardham, J., Barbour, A. & Norris, S. Antigenic variation in Lyme disease *Borreliae* by promiscuous recombination of VMP-like sequence cassettes. *Cell* **89,** 275–285 (1997).
26. Tomb, J.-F. *et al*. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Science* **388,** 539–547 (1997).
27. Curnow, A. W., Ibba, M. & Soll, D. tRNA-dependent asparagine formation. *Nature* **382,** 589–590 (1996).
28. Brown, J. R. & Doolittle, W. F. *Archaea* and the prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev.* **61,** 456–502 (1997).
29. Eisen, J. A. The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. *J. Mol. Evol.* **41,** 1105–1123 (1995).
30. Smith, D. R. *et al*. Complete genome sequence of *Methanobacterium thermoautotrophicum*ΔH: Functional analysis and comparative genomics. *J. Bacteriol.* **179,** 7135–7155 (1997).
31. Kunst, F. *et al*. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390,** 249–256 (1997).
32. Fraser, C. M. *et al*. Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi*. *Nature* **390,** 580–586 (1997).
33. Blattner, F. R. *et al*. The complete genome sequence of *Escherichia coli* K-12. *Science* **277,** 1453–1462 (1997).
34. Cole, S. T. *et al*. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **11,** 537–544 (1998).
35. Fraser, C. M. *et al*. The minimal gene complement of *Mycoplasma genitalium*. *Science* **270,** 397–403 (1995).
36. Himmelreich, R. *et al*. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* **24,** 4420–4449 (1996).
37. Kaneko, T. *et al*. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis sp.* strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* (suppl.) **3,** 185–209 (1996).
38. Fraser, C. M. *et al*. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* **281,** 375–388 (1998).
39. Goffeau, A. *et al*. Life with 6000 genes. *Science* **274,** 563–567 (1996).
40. Huang, Y.-P. & Ito, J. The hyperthermophilic bacterium *Thermotoga maritima* has two different classes of family C DNA polymerases: evolutionary implications. *Nucleic Acids Res.* **26,** 5300–5309 (1998).
41. Lawrence, J. G. & Ochman, H. Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* **44,** 383–397 (1997).
42. Sutton, G. G., White, O., Adams, M. D. & Kerlavage, A. R. TIGR Assembler: A new tool for assembling large shotgun sequencing projects. *Genome Seq. Technol.* **1,** 9–19 (1995).
43. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
44. Saitou, N. & Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4,** 406–425 (1987).
45. Claros, M. G. & von Heijne, G. TopPred II: an improved software for membrane protein structure predictions. *Comput. Appl. Biosci.* **10,** 685–686 (1994).

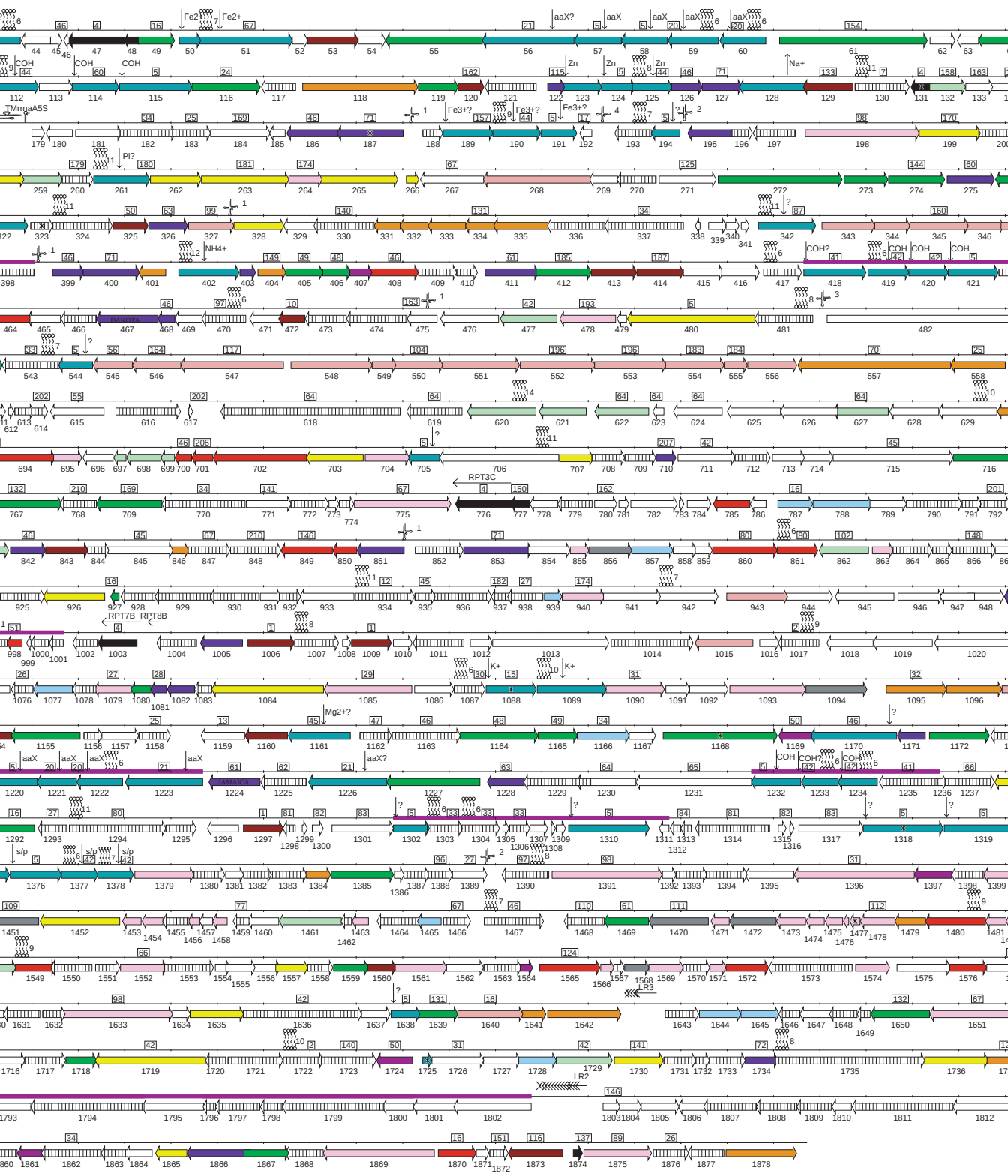
Acknowledgements. This work was supported by the US Department of Energy, Office of Biological and Environmental Research. We thank M. Heaney, J. Scott, D. Maas and B. Vincent for software and database support; R. Roberts, F. Kunst, and M. Simon for useful discussions; and R. Huber for providing *T. maritima* MSB8 cells.

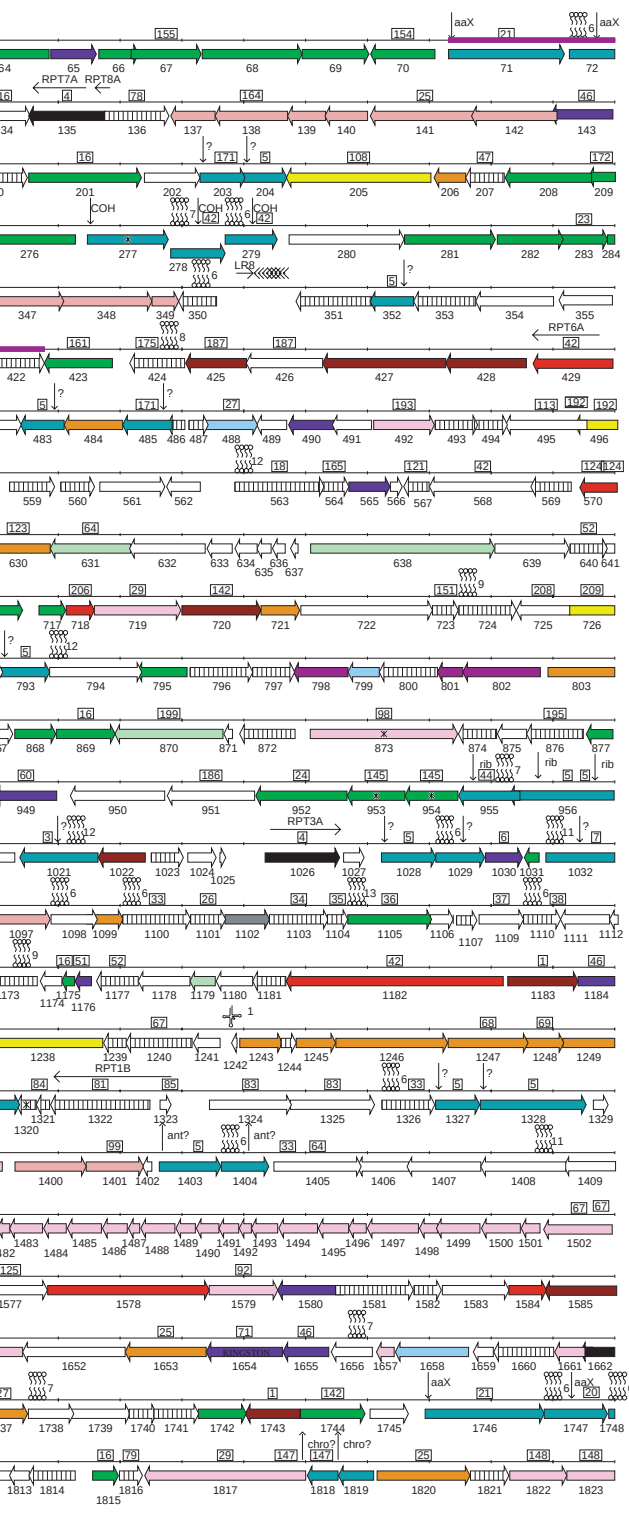
Correspondence and requests for materials should be addressed to C.M.F. (e-mail: btm@tigr.org). The annotated genome sequence and the gene family alignments are available on the World-Wide Web at http://www.tigr.org/tdb/mdb/. The sequence has been deposited in GenBank with accession number AE000512.


Amino acid biosynthesis
Biosynthesis of cofactors, prosthetic groups, carriers
Cell envelope
Cellular processes
Central intermediary metabolism
Energy metabolism
Fatty acid/Phospholipid metabolism
Purines, pyrimidines, nucleosides and nucleotides
Regulatory functions
DNA Metabolism
Transport/binding proteins
Translation
Transcription
Other categories
Conserved hypothetical
Unknown
Signal peptide
Lipoprotein
Transporter
GIES region
Paralogous gene family
Authentic Frame Shift
Repeat Region
Archaeal Islands
23s rRNA
16s rRNA
5s rRNA
tRNA
1 kb


LR1
RPT5A
RPT6B
aaX
Fe3+
Fe3+?
aaX?
LR7
RPT4A
COH
COH?
rnpB
Na+
rib
g3Pi
Pi
RPT1A
RPT2B
RPT3D
RPT3B
LR4
ura
glyc
K+?
mal


aaX
COH
H+
TMrrnaA16S
TMrrnaA23S
K+?
LR6
aaX?
RPT6C
Pi?
RPT4B
bcaa
mal
s/p


Fe2+
Fe2+
aaX?
aaX
aaX
aaX
aaX
COH
COH
Zn
Zn
Zn
Na+
TMrrnaA5S
Fe3+?
Fe3+?
Fe3+?
Pi?
NH4+
COH?
COH
COH
COH
DAKOTA
RPT3C
RPT7B
RPT8B
K+
K+
Mg2+?
aaX
aaX
aaX
aaX
COH
COH
COH
JAMAICA
s/p
s/p
s/p
LR3
LR2



- Amino acid biosynthesis
- Biosynthesis of cofactors, prosthetic groups, carriers
- Cell envelope
- Cellular processes
- Central intermediary metabolism

- Energy metabolism
- Fatty acid/Phospholipid metabolism
- Purines, pyrimidines, nucleosides and nucleotides
- Regulatory functions
- DNA Metabolism

- Transport/binding proteins
- Translation
- Transcription
- Other categories
- Conserved hypothetical
- Unknown

	Signal peptide
LP	Lipoprotein
Fe+++	Transporter
9	GES region
94	Paralogous gene family
	Authentic Frame Shift
	Repeat Region
	Archaeal Islands

23S	23s rRNA
16S	16s rRNA
5S	5s rRNA
4	tRNA

1 kb

Table 2. *Thermotoga maritima* Gene List

Amino acid biosynthesis

Aromatic amino acid family

TM0349 3-dehydroquinase dehydratase (aroD) {Hi} 70.7
TM0345 3-phosphoshikimate-1-carboxyvinylTase (aroA) {Bs} 66.3
TM0142 anthranilate Sase component I (trpE) {Tm} 99.8
TM0141 anthranilate Sase component II (trpGD) {Tm} 99.8
TM0343 chorismate mutase, put {Sx} 73.2
TM0155 chorismate mutase/prephenate dehydratase {Af} 80.5
TM0347 chorismate Sase (aroC) {Aa} 61.9
TM0140 indole-3-glycerol P Sase (trpC) {Tm} 99.6
TM0139 PPRanthranilate isomerase (trpF) {Tm} 100
TM0344 prephenate DHase (tyrA) {SPCC} 49.4
TM0346 shikimate 5-DHase (aroE) {Aa} 66.9
TM0348 shikimate kinase/3-dehydroquinate Sase (aroBK) {SPCC} 54.8
TM0137 tryptophan Sase, alpha sub (trpA) {Tm} 100
TM0138 tryptophan Sase, beta sub (trpB-1) {Tm} 100
TM0539 tryptophan Sase, beta sub (trpB-2) {Af} 77.0

Aspartate family

TM0268 5-methyltetrahydrofolate S-homocysteine methylTase {Ml} 54.3
TM1286 5-methyltetrahydropteroyltriglutamate—homocysteine methylTase {Aa} 82.5
TM1255 aspartate aminoTase (aspC-1) {Ta} 66.2
TM1698 aspartate aminoTase (aspC-2) {Ta} 53.3
TM1400 aspartate aminoTase, put {Mtf} 66.3
TM1523 aspartate-semialdehyde DHase {Aa} 69.5
TM1518 aspartokinase II (lysC-1) {Bst} 65.0
TM0547 aspartokinase II (lysC-2) {Bst} 67.0
TM1270 cystathionine gamma-Sase (metB) {Hp} 59.5
TM1517 diaminopimelate decarboxylase (lysA) {Aa} 66.6
TM1522 diaminopimelate epimerase (dapF) {Aa} 59.5
TM1520 dihydrodipicolinate RDase (dapB) {Psy} 62.7
TM1521 dihydrodipicolinate Sase (dapA) {Mj} 72.5
TM0545 homoserine kinase, put {Aa} 61.8
TM0881 homoserine O-succinylTase (metA) {Ec} 69.8
TM1666 succinyl-diaminopimelate desuccinylase, put {Af} 71.2
TM0546 threonine Sase (thrC) {Aa} 82.6

Glutamate family

TM1784 acetylglutamate kinase (argB) {Mj} 74.8
TM1785 acetylornithine aminoTase (argD) {Aa} 71.6
TM1781 argininosuccinate lyase (argH), AFS {Aa} 58.6
TM1780 argininosuccinate Sase (argG) {Mmu} 76.8
TM0293 gamma-glutamyl P RDase (proA) {Aa} 72.7
TM0294 glutamate 5-kinase (proB) {Aa} 64.9
TM1015 glutamate DHase {Tm} 100
TM1783 glutamate N-acetylTase (argJ) {Bst} 64.3
TM0397 glutamate Sase, alpha sub {Af} 81.2
TM1217 glutamate Sase, beta sub (gltD-1) {Pyr} 58.5
TM1640 glutamate Sase, beta sub (gltD-2) {Pyr} 74.8
TM0943 glutamine Sase (glnA) {Tm} 99.5
TM1782 N-acetyl-gamma-glutamyl-P RDase (argC) {SPCC} 67.4
TM1097 ornithine carbamoylTase, anabolic (argF) {Tm} 100
TM0578 pyrroline-5-carboxylate RDase (proC) {Aa} 60.2

Pyruvate family

TM0553 2-isopropylmalate Sase (leuA) {Aa} 72.8
TM0552 2-isopropylmalate Sase, put {Aa} 81.1
TM0554 3-isopropylmalate dehydratase, large sub (leuC) {Aa} 76.9
TM0291 3-isopropylmalate dehydratase, large sub, put {Af} 77.2
TM0555 3-isopropylmalate dehydratase, small sub (leuD) {Af} 76.5
TM0292 3-isopropylmalate dehydratase, small sub, put {Af} 80.0
TM0556 3-isopropylmalate DHase (leuB) {Aa} 74.4
TM0548 acetolactate Sase, large sub (ilvB) {Af} 71.8
TM0549 acetolactate Sase, small sub (ilvN) {Af} 73.0
TM0831 branched-chain AA aminoTase, put {Tm} 99.0
TM0551 dihydroxy-acid dehydratase (ilvD) {Aa} 74.0
TM0550 ketol-acid reductoisomerase (ilvC) {SPCC} 81.8

Serine family

TM0665 cysteine Sase (cysK) {Bs} 73.2
TM1401 D-3-phosphoglycerate DHase (serA) {Mj} 72.7
TM0882 O-acetylhomoserine sulfhydrylase (cysD) {Lm} 74.2
TM0327 phosphoglycerate DHase, put {Ph} 68.8
TM0666 serine acetylTase (cysE) {Sx} 59.9

Histidine family

TM1038 amidoTase (hisH) {Mj} 64.2
TM1042 ATP PPRTase (hisG) {Aa} 60.8
TM1036 cyclase (hisF) {Aa} 74.0
TM1041 histidinol DHase (hisD) {Bs} 68.5
TM1040 histidinol-P aminoTase (hisC) {Aa} 59.0
TM1039 imidazoleglycerol-P dehydratase (hisB) {Ca} 56.1
TM1035 PPR-AMP cyclohydrolase / PPR-ATP pyrophospho-hydrolase {Aa} 70.4
TM1037 PPRformimino-5-aminoimidazole carboxamide ribotide isomerase (hisA) {Ll} 56.3

Purines, pyrimidines, nucleosides, and nucleotides

2'-Deoxyribonucleotide metabolism

TM0118 ribonucleotide RDase, B12-dep (nrdJ) {Tm} 100

Nucleotide and nucleoside interconversions

TM1479 adenylate kinase (adk) {Af} 74.9

Purine ribonucleotide biosynthesis

TM1095 adenylosuccinate lyase (purB) {Aa} 73.6
TM1096 adenylosuccinate Sase (purA) {Aa} 66.5
TM1247 amidoPPRTase (purF) {Bs} 66.1
TM1641 dihydrofolate RDase (dyrA) {Tm} 100
TM1820 GMP Sase (guaA) {Aa} 76.5
TM1347 inosine-5'-monoP DHase (guaB) {Aa} 85.7
TM1628 PPR pyroP Sase (prs) {Cam} 75.9
TM1250 PPRamine—glycine ligase (purD) {Aa} 62.5
TM0447 PPRaminoimidazole carboxylase, ATPase sub (purK) {Bs} 63.3
TM0446 PPRaminoimidazole carboxylase, catalytic sub (purE) {Mt} 73.6
TM1243 PPRaminoimidazole-succinocarboxamide Sase (purC) {Mj} 68.0
TM1249 PPRaminoimidazolecarboxamide formylTase/IMP cyclohydrolase (purH) {Aa} 61.4
TM1251 PPRformylglycinamidine cyclo-ligase (purM) {Hs} 58.5
TM1245 PPRformylglycinamidine Sase I (purQ) {Bs} 71.8
TM1246 PPRformylglycinamidine Sase II (purL) {Aa} 62.4
TM1248 PPRglycinamide formylTase (purN) {Bs} 70.6

Pyrimidine ribonucleotide biosynthesis

TM1642 aspartate carbamoylTase, catalytic and reg chains (pyrBI) {Tm} 100
TM0557 carbamoyl-P Sase, large chain (carB) {Af} 78.5
TM0558 carbamoyl-P Sase, small chain (carA) {Aa} 75.3
TM0803 CTP Sase (pyrG) {Aa} 71.9
TM0404 deoxycytidylate deaminase, put {Af} 74.0
TM0335 dihydroorotase (pyrC) {Bs} 55.2
TM0333 dihydroorotate DHase (pyrD) {Bs} 61.4
TM0334 dihydroorotate DHase electron transfer prt {Bs} 53.6
TM0331 orotate PPRTase (pyrE) {Ta} 72.2
TM0332 orotidine 5'-P decarboxylase, put {Ec} 52.2
TM0484 pyrimidine precursor biosynth enzyme, put {Sc} 52.2
TM1694 thiamin biosynth prt Thil (thil) {Bs} 68.7
TM1099 thymidylate kinase (tmk) {Aa} 71.7
TM1604 uridylate kinase (pyrH) {Aa} 68.9

Salvage of nucleosides and nucleotides

TM1384 adenine PPRTase (apt) {Bs} 75.3
TM0846 cytidine/deoxycytidine deaminase (cdd) {Bs} 73.6
TM1443 cytidylate kinase (cmk) {Bs} 68.2
TM1689 guanylate kinase (gmk) {Hp} 64.8
TM0206 hypoxanthine PPRTase (hpt) {Bs} 66.7
TM0167 phosphopentomutase (deoB) {Bs} 72.1
TM1596 purine nucleoside phosphorylase (deoD-1) {Bs} 68.5
TM1737 purine nucleoside phosphorylase (deoD-2) {Bs} 74.7
TM1653 pyrimidine-nucleoside phosphorylase (deoA) {Bst} 74.5
TM0401 thymidine kinase (tdk) {Mp} 57.8
TM0721 uracil PPRTase (upp) {Bca} 77.0

Sugar-nucleotide biosynthesis and conversions

TM0630 nucleotide sugar epimerase, put {Mta} 74.9
TM1878 UDP-sugar hydrolase (ushA) {Ec} 63.0

Fatty acid and phospholipid metabolism

Biosynthesis

TM0801 (3R)-hydroxymyristoyl-(acyl carrier prt) dehydratase (fabZ) {Aa} 72.5
TM1693 1-acyl-sn-glycerol-3-P acetylTase, put {Bb} 57.8
TM1169 3-oxoacyl-(acyl carrier prt) RDase (fabG-1) {Aa} 61.7
TM1724 3-oxoacyl-(acyl carrier prt) RDase (fabG-2) {Bs} 71.8
TM0802 3-oxoacyl-(acyl carrier prt) Sase II (fabF) {Aa} 68.7
TM0175 acyl carrier prt (acpP-1) {Aa} 69.1
TM0662 acyl carrier prt (acpP-2) {Aa} 86.3
TM1861 CDP-diacylglycerol—glycerol-3-P 3-phosphatidyl-Tase (pgsA) {Bs} 60.9
TM0407 diacylglycerol kinase, put {SPCC} 58.3
TM0149 fatty acid/phospholipid synthesis prt (plsX) {Aa} 66.3
TM0692 holo-(acyl carrier prt) Sase (acpS) {Ec} 63.5
TM0798 malonyl CoA-acyl carrier prt transacylase (fabD) {Bs} 64.2
TM1397 phosphatidate cytidylylTase, put {Bs} 55.0

Degradation

TM1350 lipase, put {Af} 62.6

Other

TM1564 acylPPase, put {Af} 69.7

Biosynthesis of cofactors, prosthetic groups, and carriers

Biotin

TM1269 biotin Sase, put {SPCC} 49.6
TM0224 biotin—(acetyl-CoA carboxylase) Sase {Mta} 57.3
TM0799 bioY prt (bioY) {Bsp} 59.0

Folic acid

TM0041 2-amino-4-hydroxy-6-hydroxymethyl-dihydropteridine pyrophosphokinase (folK) {Bs} 65.9
TM0040 dihydropteroate Sase (folP) {Bs} 68.6
TM0166 folylpolyglutamate Sase/dihydrofolate Sase (folC) {Bs} 57.8

Heme and porphyrin

TM1465 cob(I)alamin adenosylTase (cobA) {Ph} 72.8
TM0488 hemK prt (hemK) {Hp} 62.4
TM1166 oxygen-indep coproporphyrinogen III oxidase, put {Aa} 62.1

Menaquinone and ubiquinone

TM1528 1,4-dihydroxy-2-naphthoate octaprenylTase, put {Bs} 57.7
TM1535 octoprenyl-diP Sase, put {Aa} 63.8
TM0753 ubiquinone/menaquinone biosynth methylTase, put {Bs} 54.8

Pantothenate

TM1728 3-methyl-2-oxobutanoate hydroxymethylTase (panB) {Ph} 75.7
TM0939 aspartate 1-decarboxylase (panD) {Aa} 65.5
TM1687 DNA/pantothenate metabolism flavoprt (dfp) {Aa} 65.0
TM1077 pantoate—beta-alanine ligase (panC) {Aa} 74.2

Riboflavin

TM1825 6,7-dimethyl-8-ribityllumazine Sase (ribH) {Bs} 69.5
TM1826 GTP cyclohydrolase II/3,4-dihydroxy-2-butanone 4-P Sase (ribA) {Bs} 63.1
TM0857 riboflavin kinase/FMN adenylylTase (ribF) {Aa} 59.9
TM1827 riboflavin Sase, alpha chain (ribE) {Aa} 60.6
TM1828 riboflavin-specific deaminase (ribD) {Aa} 65.5

Thiamine

TM1770 1-deoxyxylulose-5-P Sase (dxs) {Ec} 65.5
TM0788 thiamine biosynth prt, put {Mj} 75.1
TM0787 thiamine biosynth enzyme {Mta} 67.8
TM1267 thiH prt, put {Ec} 53.9

Pyridine nucleotides

TM1253 NH(3)-dep NAD(+) Sase (nadE) {Aa} 70.0
TM0645 NH(3)-dep NAD(+) Sase, put {Mj} 61.4
TM1645 nicotinate-nucleotide pyrophosphorylase (nadC) {Mta} 62.1
TM1644 quinolinate Sase A (nadA) {Mj} 70.1

Other

TM0038 6-pyruvoyl tetrahydrobiopterin Sase, put {Ec} 62.2
TM0161 geranyltransTase (ispA) {Bst} 62.1
TM1658 S-adenosylmethionine Sase (metK) {Bs} 81.8

Central intermediary metabolism

Amino sugars

TM0814 N-acetylglucosamine-6-P deacetylase (nagA) {Bs} 56.6

Polyamine biosynthesis

TM1873 ornithine decarboxylase (odc) {Aa} 64.2
TM0654 spermidine Sase (speE) {Bs} 69.7

Other

TM0225 1-aminocyclopropane-1-carboxylate deaminase, put {Ec} 55.1
TM1129 5-methylthioadenosine/S-adenosylhomocysteine nucleosidase (pfs) {Bs} 55.8
TM0573 acylTase, put {Aa} 50.6
TM0759 acylTase, put {Mta} 63.4
TM0172 adenosylhomocysteinase (ahcY) {Tm} 99.8
TM0156 alkaline PPase (phoB) {Bs} 58.5
TM0472 amidoTase, put {Ml} 66.7
TM1371 aminoTase, class V (nifS) {SPCC} 61.4
TM1692 aminoTase, class V {Ana} 62.0
TM1131 aminoTase, put {Ss} 64.2
TM0129 carboxypeptidase G2, put {Pse} 53.1
TM0413 creatinine amidohydrolase, put {Pse} 49.2
TM0414 DHase {Bs} 67.4
TM1022 esterase (estA) {Tm} 100
TM1160 esterase (estB) {Tm} 100
TM0053 esterase, put {Pa} 57.4
TM1766 formate—tetrahydrofolate ligase (fhs) {Cla} 74.1
TM0843 formiminoTase-cyclodeaminase/formimino-tetrahydrofolate cyclodeaminase, put {Ssc} 63.7
TM1585 glycerate kinase, put {Me} 55.6
TM1516 hydrolase, ama/hipO/hyuC fam {Bs} 57.6
TM0809 hydrolase, put {Alt} 59.6
TM1767 methylenetetrahydrofolate DHase/methenyltetra-hydrofolate cyclohydrolase (folD) {Bs} 66.7
TM0754 oxidoRDase {Ph} 63.4
TM1743 oxidoRDase, aldo/keto RDase fam {Bs} 59.8
TM1009 oxidoRDase, aldo/keto RDase fam {Bs} 69.2
TM1006 oxidoRDase, aldo/keto RDase fam {Hp} 71.7
TM1183 oxidoRDase, aldo/keto RDase fam {Mta} 61.1
TM0427 oxidoRDase, put {Aa} 45.1
TM1433 oxidoRDase, put {Af} 56.7
TM0428 oxidoRDase, put {Af} 49.3
TM1297 oxidoRDase, put {Bs} 55.8
TM0425 oxidoRDase, put {Bs} 48.1
TM0120 oxidoRDase, put {Ce} 56.3
TM0441 oxidoRDase, short chain DHase/RDase fam {Af} 62.9
TM0019 oxidoRDase, short chain DHase/RDase fam {Bs} 59.4
TM0325 oxidoRDase, short chain DHase/RDase fam {Bs} 57.9
TM0297 oxidoRDase, short chain DHase/RDase fam {Mt} 59.9
TM1154 oxidoRDase, sol/devB fam {Ana} 57.7
TM1560 serine cycle enzyme, put {Me} 55.3
TM0720 serine hydroxymethylTase (glyA) {Aa} 82.3

Energy metabolism

Amino acids and amines

TM0119 acetamidase, put {Ms} 49.6
TM0211 aminomethylTase (gcvT) {Bs} 67.0
TM0444 aspartate NH3-lyase (aspA) {Bs} 64.1
TM0212 glycine cleavage system H prt (gcvH) {Ec} 71.1
TM0213 glycine DHase (decarboxylating) sub 1 (yqhJ) {Aa} 69.3
TM0214 glycine DHase (decarboxylating) sub 2 (yqhK) {Ph} 78.9
TM1744 L-allo-threonine aldolase {Ec} 67.6
TM0356 threonine dehydratase catabolic {Ec} 74.4

Anaerobic

TM1867 L-lactate DHase (ldh) {Tm} 100

ATP-proton motive force interconversion

TM1616 ATP Sase F0, sub a (atpB) {Bm} 60.9
TM1614 ATP Sase F0, sub b (atpF) {Bf} 68.7
TM1615 ATP Sase F0, sub c (atpE) {Eg} 67.2
TM1612 ATP Sase F1, sub alpha (atpA) {Aa} 82.2
TM1610 ATP Sase F1, sub beta (atpD) {Tm} 100
TM1613 ATP Sase F1, sub delta (atpH) {Cr} 55.3
TM1609 ATP Sase F1, sub epsilon (atpC) {Bf} 60.2
TM1611 ATP Sase F1, sub gamma (atpG) {Aa} 58.5

Electron transport

TM1141 cytochrome C-type biogenesis prt, put {Hi} 56.1
TM1531 electron transfer flavoprt, alpha sub (etfA) {Cac} 68.2
TM1530 electron transfer flavoprt, beta sub (etfB) {Cac} 66.5
TM1639 electron transfer prt, put {Mj} 57.6
TM0244 electron transport complex prt, put {Rc} 60.9
TM1426 Fe-hydrogenase, sub alpha (hydA) {Tm} 100
TM1425 Fe-hydrogenase, sub beta (hydB) {Tm} 100
TM1424 Fe-hydrogenase, sub gamma (hydC) {Tm} 100
TM0927 ferredoxin (fdx) {Tm} 100
TM1289 ferredoxin {Dd} 66.0
TM1175 ferredoxin {Mj} 52.5
TM1815 ferredoxin {Mj} 60.6
TM0879 ferredoxin {Mtp} 61.8
TM1533 ferrodoxin (fixX) {Av} 68.8
TM1532 fixC prt (fixC) {Aca} 66.8
TM1031 glutaredoxin {Cpa} 73.6
TM0868 glutaredoxin-rel prt {Pfu} 68.3
TM1421 hydrogenase, put {Dv} 52.2
TM1291 Fe-S cluster-Bprt {Af} 66.1
TM0396 Fe-S cluster-Bprt {Af} 65.3
TM0034 Fe-S cluster-Bprt {Mta} 62.4
TM0049 Fe-S cluster-Bprt {Mta} 56.5
TM1292 Fe-S cluster-Bprt, put {Af} 63.1

TM0245	Na-translocating NADH-quinone Rdase (nqr2) {Va}	58.8
TM0247	Na-translocating NADH-quinone Rdase (nqr4) {Va}	64.1
TM0248	Na-translocating NADH-quinone Rdase (nqr5) {Va}	68.1
TM1215	NADH DHase, 30 kDa sub, put {Ra}	56.7
TM1216	NADH DHase, 49 kDa sub, put {Ph}	70.2
TM1214	NADH DHase, put {Hp}	63.8
TM1211	NADH DHase, put {Ph}	56.3
TM1212	NADH DHase, put {Ph}	57.7
TM1213	NADH DHase, put {Ph}	64.4
TM1105	NADH DHase, put {SPCC}	60.2
TM0383	NADH oxidoRDase, put {Mta}	74.4
TM0012	NADP-reducing hydrogenase, sub A (hndA-1) {Df}	73.7
TM0227	NADP-reducing hydrogenase, sub A (hndA-2), AFS {Df}	63.6
TM0011	NADP-reducing hydrogenase, sub B (hndB) {Df}	61.9
TM0010	NADP-reducing hydrogenase, sub C (hndC-1) {Df}	78.7
TM0228	NADP-reducing hydrogenase, sub C (hndC-2) {Df}	72.3
TM0201	NADP-reducing hydrogenase, sub D, put {Df}	69.3
TM0658	neelaredoxin {Dg}	71.5
TM1172	prismane prt {Mta}	79.8
TM1422	rnfB-rel prt {Mta}	56.3
TM0249	rnfB-rel prt {Rc}	64.9
TM0659	rubredoxin {Cpa}	83.0
TM0657	rubrerythrin {Af}	74.5
TM0869	thioredoxin RDase (trxB) {Ph}	67.6
Entner-Doudoroff		
TM0066	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase {Bs}	67.4
TM0443	gluconate kinase {Bs}	59.3
Fermentation		
TM1164	2-oxoacid ferredoxin oxidoRDase, alpha sub {Mta}	69.0
TM1165	2-oxoacid ferredoxin oxidoRDase, beta sub {Mta}	77.8
TM1760	2-oxoisovalerate oxidoRDase, alpha sub, put, AFS {Mta}	62.8
TM1759	2-oxoisovalerate oxidoRDase, beta sub, put {Mta}	68.1
TM1758	2-oxoisovalerate oxidoRDase, gamma sub, put {Mta}	61.7
TM0274	acetate kinase (ackA) {Alc}	80.3
TM0920	alcohol DHase, Fe-containing {Aa}	57.7
TM0111	alcohol DHase, Fe-containing {Bs}	54.1
TM0412	alcohol DHase, Zn-containing {Bs}	54.9
TM0298	alcohol DHase, Zn-containing {Ec}	51.2
TM0436	alcohol DHase, Zn-containing {Pp}	51.8
TM1756	branched-chain-fatty-acid kinase, put {Bs}	73.6
TM1754	butyrate kinase, put {Cac}	64.6
TM0405	keto/oxoacid ferredoxin oxidoRDase, beta sub, put {Mj}	69.1
TM0406	keto/oxoacid ferredoxin oxidoRDase, gamma sub, put {Mj}	59.8
TM0820	NADH-dep butanol DHase, put {Bs}	68.2
TM1130	P butyrylTase (ptb-1) {Cac}	66.7
TM1755	P butyrylTase (ptb-2) {Cac}	63.0
TM0017	pyruvate ferredoxin oxidoRDase, alpha sub (porA) {Tm}	100
TM0018	pyruvate ferredoxin oxidoRDase, beta sub (porB) {Tm}	100
TM0016	pyruvate ferredoxin oxidoRDase, delta sub (porD) {Tm}	100
TM0015	pyruvate ferredoxin oxidoRDase, gamma sub (porG) {Tm}	99.5
Glycolysis/gluconeogenesis		
TM0209	6-phosphofructokinase (pfkA) {Bst}	72.1
TM0289	6-phosphofructokinase, pyroP-dep {Nf}	69.0
TM0877	enolase (eno) {Bs}	83.1
TM0273	fructose-bisP aldolase {Aa}	78.3
TM1469	glucokinase {Bm}	58.5
TM1385	glucose-6-P isomerase (pgi) {Mj}	61.5
TM0688	glyceraldehyde-3-P DHase (gap) {Tm}	100
TM0689	phosphoglycerate kinase/triose-P isomerase {Tm}	99.7
TM1374	phosphoglycerate mutase {Aa}	62.4
TM0208	pyruvate kinase {Bs}	64.5
Pentose phosphate pathway		
TM0438	6-phosphogluconate DHase, decarboxylating (gnd) {Ec}	69.5
TM1155	glucose-6-P 1-DHase {SPCC}	68.5
TM1718	ribulose-P 3-epimerase {Mj}	66.4
TM0953	transketolase, C-terminal sub {Mj}	62.7
TM0954	transketolase, N-terminal sub {Mj}	71.4
TM1762	transketolase, put {Mmu}	53.5
Pyruvate dehydrogenase		
TM0381	dihydrolipoamide DHase (pdhD) {Bst}	61.4
Sugars		
TM1281	6-phospho-beta-glucosidase {Bs}	59.4
TM0077	acetyl xylan esterase (axeA) {Tn}	95.4
TM0282	aldose 1-epimerase (galM) {Ac}	63.8
TM1192	alpha-galactosidase (galA) {Tm}	100
TM1834	alpha-glucosidase (aglA) {Tm}	100
TM1068	alpha-glucosidase, put {Tm}	76.0
TM0434	alpha-glucosidase, put {Tm}	75.8
TM0752	alpha-glucosidase, put {Tm}	74.0
TM0055	alpha-glucuronidase (aguA) {Tm}	100
TM0306	alpha-L-fucosidase, put {Hs}	56.9
TM0308	alpha-xylosidase (xylQ) {Lp}	63.8
TM1201	arabinogalactan endo-1,4-beta-galactosidase, put {Bci}	66.8
TM0310	beta-D-galactosidase {Bci}	68.3
TM1414	beta-fructosidase (bfrA) {Tm}	100
TM1193	beta-galactosidase {Tm}	99.6
TM1195	beta-galactosidase {Tm}	99.5
TM0025	beta-glucosidase (bglB) {Tn}	92.8
TM1062	beta-glucuronidase {Ec}	57.7
TM1624	beta-mannosidase, put {Ch}	53.8
TM1254	beta-phosphoglucomutase, put {Bs}	57.2
TM1848	cellobiose-phosphorylase (cepA) {Tn}	97.7
TM1835	cyclomaltodextrinase, put {Bac}	60.2
TM0069	D-mannonate hydrolase (uxuA) {Tn}	99.4
TM0068	D-mannonate oxidoRDase, put {Ec}	50.9
TM0437	exo-poly-alpha-D-galacturonosidase, put {Ech}	49.5
TM0296	fructokinase {Bv}	57.1
TM1190	galactokinase (galK) {Tm}	100
TM1191	galactose-1-P uridylylTase (galT) {Tm}	100
TM0896	galactose-1-P uridylylTase, put {Hi}	52.0
TM0276	L-arabinose isomerase (araA) {Sch}	73.1
TM0307	L-fucose isomerase, put {Hi}	46.6
TM0767	maltodextrin glycosylTase (mmtA) {Tm}	100
TM0736	mannose-6-P isomerase {Bs}	56.2
TM1742	nagD prt, put {Ec}	56.0
TM0769	phosphomannomutase {Bs}	59.5
TM0960	ribokinase (rbsK) {Hi}	64.1
TM1071	sugar isomerase {Art}	50.2
TM0283	sugar isomerase {Ec}	60.8
TM1073	sugar kinase {Ec}	59.4
TM0284	sugar kinase, FGGY fam {Bs}	54.4
TM0116	sugar kinase, FGGY fam {Sx}	59.5
TM0795	sugar kinase, pfkB fam {Bs}	45.3
TM0828	sugar kinase, pfkB fam {Sa}	54.1
TM1072	sugar-P aldolase {Ec}	51.8
TM1080	sugar-P isomerase {Aa}	77.6
TM0509	UDP-glucose 4-epimerase, put {Mj}	68.9
TM0064	uronate isomerase, put {Ec}	58.3
TM1667	xylose isomerase (xylA) {Tn}	98.9
TM0076	xylosidase (xloA) {Tn}	98.1
TCA cycle		
TM0290	citrate Sase {Af}	66.4
TM0541	fumarate hydratase, C-terminal sub {Aa}	67.3
TM0540	fumarate hydratase, N-terminal sub {Aa}	62.9
TM1148	isocitrate DHase (NADP) {Sc}	74.6
Biosynthesis and degradation of polysaccharides		
TM0067	2-keto-3-deoxygluconate kinase {Bs}	47.4
TM0364	4-alpha-glucanoTase {Tm}	99.8
TM1840	alpha-amylase (amyA) {Tm}	100
TM1650	alpha-amylase, put {Dt}	52.1
TM1168	alpha-glucan phosphorylase (agpA), AFS {Tm}	99.3
TM0281	alpha-L-arabinofuranosidase {Bs}	58.7
TM1851	alpha-mannosidase, put {SPCC}	57.8
TM1227	endo-1,4-beta-mannosidase {Ab}	54.4
TM0061	endo-1,4-beta-xylanase A (xynA) {Tm}	100
TM0070	endo-1,4-beta-xylanase B (xynB) {Tn}	93.9
TM1524	endoglucanase (celA) {Tm}	100
TM1525	endoglucanase (celB) {Tm}	100
TM1751	endoglucanase {Clo}	52.3
TM1752	endoglucanase {Clo}	51.5
TM1050	endoglucanase {Ct}	64.7
TM1048	endoglucanase {Mj}	57.6
TM1049	endoglucanase {Mj}	60.2
TM0305	endoglucanase, put {Sco}	55.1
TM0239	glucose-1-P adenylylTase (glgC-1) {Bs}	61.0
TM0240	glucose-1-P adenylylTase (glgC-2) {Bst}	75.3
TM0895	glycogen Sase (glgA) {Bst}	63.6
TM0024	laminarinase (lamA) {Tn}	95.5
TM0433	pectate lyase {Amy}	54.7
TM1845	pullulanase (pulA) {Tm}	100
Other		
TM1559	deoxyribose-P aldolase (deoC) {Mta}	71.5
TM0423	glycerol DHase (gldA) {Ec}	68.5
TM0952	glycerol kinase (glpK-1) {Bs}	68.2
TM1430	glycerol kinase (glpK-2) {Bs}	81.9
TM0378	glycerol-3-P DHase (gpsA), AFS {Bs}	65.5
TM0542	malate oxidoRDase {Sb}	70.9
TM1185	methylglyoxal Sase {Ec}	71.0
TM0006	muconate cycloisomerase {Pp}	47.8
TM0716	propionyl-CoA carboxylase, beta chain (pccB) {Se}	76.3
TM0717	propionyl-CoA carboxylase, gamma sub {Pm}	51.9
TM0272	pyruvate,orthoP dikinase {Mc}	75.2

Transport and binding proteins

Amino acids, peptides and amines		
TM0591	AA ABC transp, ATP-Bprt {Af}	80.1
TM0593	AA ABC transp, periplasmic AA-Bprt {Af}	64.7
TM0592	AA ABC transp, permease prt {Af}	72.9
TM1139	BRAA ABC transp, ATP-Bprt (livF) {Af}	74.5
TM1138	BRAA ABC transp, ATP-Bprt (livG) {Sch}	74.0
TM1135	BRAA ABC transp, periplasmic AA-Bprt (livJ) {Ec}	49.9
TM1136	BRAA ABC transp, permease prt (livH) {Sch}	68.1
TM1137	BRAA ABC transp, permease prt (livM) {Pa}	62.7
TM1151	OP ABC transp, ATP-Bprt {Af}	74.6
TM1152	OP ABC transp, ATP-Bprt {Af}	75.1
TM1063	OP ABC transp, ATP-Bprt {Bs}	75.3
TM1064	OP ABC transp, ATP-Bprt {Bs}	72.9
TM1750	OP ABC transp, ATP-Bprt {Bs}	76.4
TM0057	OP ABC transp, ATP-Bprt {Bs}	74.8
TM0058	OP ABC transp, ATP-Bprt {Bs}	69.1
TM0500	OP ABC transp, ATP-Bprt {Bs}	79.6
TM0501	OP ABC transp, ATP-Bprt {Bs}	74.8
TM0530	OP ABC transp, ATP-Bprt {Bs}	73.6
TM0498	OP ABC transp, ATP-Bprt {Hi}	72.4
TM0303	OP ABC transp, ATP-Bprt {Ph}	76.0
TM1196	OP ABC transp, ATP-Bprt {Ph}	66.0
TM1219	OP ABC transp, ATP-Bprt {Ph}	74.9
TM1220	OP ABC transp, ATP-Bprt {Ph}	79.1
TM0027	OP ABC transp, ATP-Bprt {Ph}	71.1
TM0028	OP ABC transp, ATP-Bprt {Ph}	69.0
TM0075	OP ABC transp, ATP-Bprt {Ph}	68.9
TM0304	OP ABC transp, ATP-Bprt {Ph}	65.5
TM1749	OP ABC transp, ATP-Bprt {Sch}	69.6
TM1194	OP ABC transp, ATP-Bprt {Tm}	100
TM0074	OP ABC transp, ATP-Bprt {Tm}	81.3
TM0460	OP ABC transp, periplasmic OP Bprt, put {Ec}	48.2
TM1067	OP ABC transp, periplasmic OP-Bprt {Aa}	51.3
TM0309	OP ABC transp, periplasmic OP-Bprt {Aa}	53.4
TM0531	OP ABC transp, periplasmic OP-Bprt {Aa}	60.1
TM1150	OP ABC transp, periplasmic OP-Bprt {Af}	66.8
TM1746	OP ABC transp, periplasmic OP-Bprt {Bb}	56.5
TM0031	OP ABC transp, periplasmic OP-Bprt {Hp}	45.2
TM1223	OP ABC transp, periplasmic OP-Bprt {Ph}	59.0
TM1199	OP ABC transp, periplasmic OP-Bprt {SPCC}	53.4
TM0071	OP ABC transp, periplasmic OP-Bprt {Tn}	84.6
TM0300	OP ABC transp, periplasmic OP-Bprt, put {Aa}	48.2
TM0056	OP ABC transp, periplasmic OP-Bprt, put {Ec}	50.2
TM1226	OP ABC transp, periplasmic OP-Bprt, put {Ph}	57.1
TM1065	OP ABC transp, permease prt {Aa}	67.8
TM0072	OP ABC transp, permease prt {Aa}	53.9
TM1149	OP ABC transp, permease prt {Af}	78.2
TM1153	OP ABC transp, permease prt {Af}	78.4
TM1066	OP ABC transp, permease prt {Bs}	62.9
TM1748	OP ABC transp, permease prt {Bs}	64.7
TM0060	OP ABC transp, permease prt {Bs}	61.3
TM0502	OP ABC transp, permease prt {Bs}	63.9
TM0503	OP ABC transp, permease prt {Bs}	62.5
TM1747	OP ABC transp, permease prt {Hi}	65.9
TM0073	OP ABC transp, permease prt {Mm}	56.6
TM1197	OP ABC transp, permease prt {Ph}	62.5
TM1198	OP ABC transp, permease prt {Ph}	58.5
TM1221	OP ABC transp, permease prt {Ph}	72.4
TM1222	OP ABC transp, permease prt {Ph}	79.6
TM0029	OP ABC transp, permease prt {Ph}	70.2
TM0030	OP ABC transp, permease prt {Ph}	68.0
TM0301	OP ABC transp, permease prt {Ph}	59.9
TM0302	OP ABC transp, permease prt {Ph}	57.1
TM0059	OP ABC transp, permease prt {SPCC}	65.9
TM0532	OP ABC transp, permease prt {SPCC}	69.2
TM0533	OP ABC transp, permease prt {SPCC}	71.7
TM1376	spermidine/putrescine ABC transp, ATP-Bprt (potA) {Ec}	67.4
TM1375	spermidine/putrescine ABC transp, periplasmic spermidine/putrescine-Bprt (potD) {Ec}	61.0
TM1377	spermidine/putrescine ABC transp, permease prt (potB) {Hi}	64.0
TM1378	spermidine/putrescine ABC transp, permease prt (potC) {Ec}	71.2
Anions		
TM1818	chromate transport prt, put {Bb}	61.1
TM1819	chromate transport prt, put {Syn}	55.1
TM1261	P ABC transp, ATP-Bprt (pstB) {Af}	80.9
TM1264	P ABC transp, periplasmic P-Bprt (pstS) {Af}	58.8
TM1262	P ABC transp, permease prt (pstAC) {Mta}	59.6
TM1263	P ABC transp, permease prt (pstC) {Mta}	60.2
TM0261	P permease, put {Af}	76.3
TM0917	P permease, put {Af}	56.5
Carbohydrates, organic alcohols, and acids		
TM1204	maltose ABC transp, periplasmic maltose-Bprt (malE) {Tm}	91.6
TM1839	maltose ABC transp, periplasmic maltose-Bprt (malE) {Tm}	100
TM1203	maltose ABC transp, permease prt (malF) {Ea}	66.6
TM1837	maltose ABC transp, permease prt (malF) {Ea}	70.6
TM1202	maltose ABC transp, permease prt (malG) {Ec}	65.9
TM1836	maltose ABC transp, permease prt (malG) {Ec}	67.0
TM0956	ribose ABC transp, ATP-Bprt (rbsA) {Hi}	68.5
TM0959	ribose ABC transp, membrane-associated prt (rbsD) {Bs}	65.2
TM0958	ribose ABC transp, periplasmic ribose-Bprt (rbsB) {Ec}	59.0
TM0955	ribose ABC transp, permease prt (rbsC) {Bs}	70.6
TM0103	sugar ABC transp, ATP-Bprt {Af}	70.3
TM0115	sugar ABC transp, ATP-Bprt {Bs}	66.1
TM1276	sugar ABC transp, ATP-Bprt {Bs}	75.1
TM1232	sugar ABC transp, ATP-Bprt {Ph}	78.0
TM0421	sugar ABC transp, ATP-Bprt {Ph}	71.0
TM0114	sugar ABC transp, periplasmic sugar-Bprt {Bs}	51.0
TM0810	sugar ABC transp, periplasmic sugar-Bprt {SPCC}	55.0
TM0277	sugar ABC transp, periplasmic sugar-Bprt, AFS {Pf}	44.3
TM0595	sugar ABC transp, periplasmic sugar-Bprt, put {Ec}	48.6
TM1855	sugar ABC transp, periplasmic sugar-Bprt, put {Pf}	47.8
TM0432	sugar ABC transp, periplasmic sugar-Bprt, put {Sm}	55.2
TM0418	sugar ABC transp, periplasmic sugar-Bprt, put {Tm}	47.1
TM0104	sugar ABC transp, permease prt {Af}	59.0
TM0105	sugar ABC transp, permease prt {Af}	63.8
TM0112	sugar ABC transp, permease prt {Bs}	65.3
TM0419	sugar ABC transp, permease prt {Bs}	54.0
TM1853	sugar ABC transp, permease prt {Cpe}	66.0
TM0430	sugar ABC transp, permease prt {Ml}	60.1
TM0596	sugar ABC transp, permease prt {Ml}	61.0
TM0278	sugar ABC transp, permease prt {Pf}	56.0
TM0279	sugar ABC transp, permease prt {Pf}	61.1
TM1234	sugar ABC transp, permease prt {Ph}	74.9
TM0420	sugar ABC transp, permease prt {Ph}	64.7
TM1854	sugar ABC transp, permease prt {SPCC}	60.9
TM0431	sugar ABC transp, permease prt {SPCC}	56.2
TM0598	sugar ABC transp, permease prt {SPCC}	64.2
TM0811	sugar ABC transp, permease prt {SPCC}	68.2
TM1233	sugar ABC transp, permease prt, put {Ph}	75.3
TM0812	sugar ABC transp, permease prt, put {SPCC}	63.3
Cations		
TM0538	cation efflux system prt {Ph}	81.1
TM0372	cation efflux system prt, put {Aa}	53.8
TM0317	cation-transporting ATPase, P-type {Af}	64.3
TM1128	ferritin {Af}	72.0
TM0320	heavy metal Bprt {Hp}	68.2
TM0529	heavy metal Bprt {Pf}	56.5
TM0050	iron(II) transport prt A (feoA) {Ec}	63.8
TM0051	iron(II) transport prt B (feoB) {Mj}	64.7
TM0078	iron(III) ABC transp, ATP-Bprt {Af}	63.3
TM0191	iron(III) ABC transp, ATP-Bprt, put {Af}	73.8
TM0189	iron(III) ABC transp, periplasmic Fe-Bprt, put {Mj}	79.5
TM0080	iron(III) ABC transp, periplasmic-Bprt, put {Bs}	56.8
TM0079	iron(III) ABC transp, permease prt {Af}	60.7
TM0190	iron(III) ABC transp, permease prt, put {Mj}	75.1
TM0313	K+ channel, beta sub {Atl}	73.8
TM1057	K+ channel, put {SPCC}	
TM1161	Mg2+ transp MgtE, put {SPCC}	61.3
TM0402	NH4+ transp (amt) {Af}	
TM0128	oxaloacetate decarboxylase, alpha sub (oadA) {Kp}	67.4

TM0880 oxaloacetate decarboxylase, beta sub (oadB) {Af} 74.0
TM0174 pyroPPase, proton-translocating {Bv} 65.1
TM1088 TRK system K+ uptake prt TrkA (trkA), AFS {Ec} 51.7
TM1089 TRK system K+ uptake prt TrkH (trkH) {Mta} 56.9
TM1725 vacuolar ATP Sase sub D-rel prt, APM {Ph} 83.7
TM0124 Zn ABC transp, ATP-Bprt (znuC) {Ec} 63.8
TM0123 Zn ABC transp, periplasmic Zn-Bprt (znuA) {Ec} 49.6
TM0125 Zn ABC transp, permease prt (znuB) {Ec} 53.9
Nucleosides, purines and pyrimidines
TM0819 uracil permease (pyrP) {Bca} 69.4
Other
TM1403 antibiotic ABC transp, ATP-Bprt, put {Ph} 85.8
TM1404 antibiotic ABC transp, transmembrane prt, put {Af} 77.9
TM0363 fibronectin-Bprt, put {Bs} 54.0
TM1429 glycerol uptake facilitator prt (glpF) {Bs} 79.9
TM1120 glycerol-3-P ABC transp, periplasmic glycerol-3-P-Bprt (ugpB) {Ec} 51.4
TM1121 glycerol-3-P ABC transp, permease prt (ugpA) {Ec} 61.3
TM1122 glycerol-3-P ABC transp, permease prt (ugpE) {Ec} 59.5
Unknown substrate
TM1028 ABC transp, ATP-Bprt {Af} 60.5
TM0194 ABC transp, ATP-Bprt {Af} 61.2
TM0204 ABC transp, ATP-Bprt {Af} 68.3
TM0352 ABC transp, ATP-Bprt {Af} 79.5
TM0389 ABC transp, ATP-Bprt {Af} 65.9
TM0483 ABC transp, ATP-Bprt {Af} 66.3
TM0705 ABC transp, ATP-Bprt {Af} 73.6
TM1327 ABC transp, ATP-Bprt {Bs} 64.2
TM0287 ABC transp, ATP-Bprt {Bs} 68.4
TM0288 ABC transp, ATP-Bprt {Bs} 71.7
TM1256 ABC transp, ATP-Bprt {Bs} 61.4
TM1319 ABC transp, ATP-Bprt {Bs} 52.6
TM1328 ABC transp, ATP-Bprt {Bs} 53.4
TM0043 ABC transp, ATP-Bprt {Bs} 59.4
TM1054 ABC transp, ATP-Bprt {Bb} 70.9
TM1638 ABC transp, ATP-Bprt {Bb} 62.0
TM1310 ABC transp, ATP-Bprt {Bbr} 50.8
TM0222 ABC transp, ATP-Bprt {Mta} 65.0
TM1417 ABC transp, ATP-Bprt {Mj} 75.0
TM0765 ABC transp, ATP-Bprt {Mj} 61.4
TM1663 ABC transp, ATP-Bprt {Mg} 63.5
TM1302 ABC transp, ATP-Bprt {Ph} 63.2
TM1368 ABC transp, ATP-Bprt {Ph} 70.8
TM0544 ABC transp, ATP-Bprt {Ph} 64.4
TM0793 ABC transp, ATP-Bprt {Ss} 64.9
TM1318 ABC transp, ATP-Bprt, AFS {Bs} 53.9
TM0827 ABC transp, ATP-Bprt, put {Lf} 57.1
TM0322 ABC transp, periplasmic substrate-Bprt, put {Rc} 56.1
TM1170 ABC transp, periplasmic substrate-Bprt/conserved hypothetical prt {Ec} 46.3
TM0485 ABC transp, permease prt, cysTW fam {Hi} 63.4
TM0203 ABC transp, permease prt, cysTW fam {Pa} 50.2
TM1029 ABC transp, permease prt, put {Bl} 50.3
TM1032 permease, put {Bs} 48.5
TM1603 permease, put {Bs} 59.9
TM0342 permease, put {Ph} 66.0
TM1336 permease, put {Sp} 51.2
TM1021 permease, put {Tm} 100

DNA metabolism
DNA replication, recombination, and repair
TM0205 ATP-dep DNA helicase (recG) {Bs}_ 63.6
TM1238 ATP-dep DNA helicase (uvrd) {Tt} 58.6
TM0926 chromosomal replication initiator prt (dnaA) {Tm} 100
TM0703 competence-damage inducible prt, put {Bs} 61.6
TM0575 crossover junction endodeoxyribonuclease (ruvC) {Hi} 63.3
TM0362 deoxyribonuclease IV (nfo) {Aa} 58.7
TM1437 dimethyladenosine Tase (ksgA) {Bs} 64.0
TM1084 DNA gyrase, sub A (gyrA) {Tm} 99.9
TM0833 DNA gyrase, sub B (gyrB) {Tm} 100
TM0005 DNA helicase, put {Af} 77.6
TM0100 DNA ligase (ligA) {Aa} 68.8
TM0022 DNA mismatch repair prt (mutL) {Tm} 100
TM1719 DNA mismatch repair prt (mutS) {Tm} 99.5
TM1278 DNA mismatch repair prt, put {Bs} 60.3
TM0576 DNA polymerase III, alpha sub (dnaE) {Bs} 68.9
TM0461 DNA polymerase III, alpha sub, put {Aa} 59.8
TM0262 DNA polymerase III, beta sub (dnaN) {Hi} 54.7
TM0496 DNA polymerase III, epsilon sub, put {Aa} 55.6
TM0686 DNA polymerase III, gamma and tau sub (dnaZX) {Bs} 57.9
TM1452 DNA primase (dnaG) {Aa} 58.5
TM0199 DNA repair prt (radA) {Ec} 63.8
TM1557 DNA repair prt (radC) {Bs} 66.8
TM1859 DNA repair prt (recA) {Tm} 100
TM0257 DNA replication enhancer, put, AFS {Fi} 76.9
TM0258 DNA topoisomerase (topA) {Tm} 100
TM1619 DNA-directed DNA polymerase I (polA) {Cau} 65.5
TM0366 endonuclease III (nth) {Mta} 74.7
TM1865 endonuclease V (nfi) {Ec} 65.9
TM0480 excinuclease ABC, sub A (uvrA) {Aa} 77.8
TM1761 excinuclease ABC, sub B (uvrB) {Mta} 74.7
TM0265 excinuclease ABC, sub C (uvrC) {Mta} 60.9
TM0734 glucose-inhibited division prt (gid) {Bs} 72.1
TM0263 glucose-inhibited division prt A (gidA) {Bs} 72.0
TM0707 glucose-inhibited division prt B (gidB) {Pp} 50.8
TM0165 Holliday junction DNA helicase (ruvA) {Ec} 58.9
TM1730 Holliday junction DNA helicase (ruvB) {Tm} 100
TM0967 integrase-recombinase prt (xerC) {Mta} 63.5
TM0887 methylated-DNA-prt-cysteine methylTase (ogt) {Af} 64.7
TM0178 primosomal prt N' (priA) {Bs} 54.7
TM1343 pyrimidine dimer DNA glycosylase (uveA), APM {Mlu} 62.5
TM0382 repair endonuclease, put {Mj} 69.7
TM1736 replicative DNA helicase (dnaB) {Bs} 70.3
TM0173 reverse gyrase (rgy) {Pfu} 63.1
TM0915 ribonuclease HII (rnhB) {Hi} 68.3
TM0604 ssDNA-Bprt, put {Aa} 57.1
TM1546 ssDNA-specific exonuclease, put {Bs} 59.1
TM0726 tldD prt (tldD) {Ec} 60.2
TM1450 transcription-repair coupling factor, put {Bs} 64.7
Restriction/modification
TM0328 m4C-methylTase {Hp} 58.4
Degradation of DNA
TM1768 exodeoxyribonuclease VII, large sub (xseA) {Hi} 59.5
TM1769 exodeoxyribonuclease, small sub (xseB) {Ec} 70.3
TM1635 exonuclease, put {Aa} 55.1
Chromosome-associated proteins
TM0266 DNA-Bprt, HU (hupB) {Tm} 98.9

Transcription
Degradation of RNA
TM1345 polynucleotide phosphorylase (pnp) {Bs} 70.0
TM1102 ribonuclease III (rnc) {Bb} 65.2
DNA-dependent RNA polymerase
TM1472 DNA-directed RNA polymerase, alpha sub (rpoA) {Mt} 65.2
TM0458 DNA-directed RNA polymerase, beta sub (rpoB) {Tm} 100
TM0459 DNA-directed RNA polymerase, beta' sub (rpoC) {Tm} 100
Transcription factors
TM1777 N utilization substance prt A (nusA) {Bs} 73.5
TM1765 N utilization substance prt B (nusB) {Hi} 59.7
TM0453 N utilization substance prt G (nusG) {Tm} 100
TM0902 RNA polymerase sigma-28 factor, put {Lpn} 61.1
TM1451 RNA polymerase sigma-A factor (rpoD) {Cac} 75.0
TM1598 RNA polymerase sigma-E factor (rpoE) {Ec} 61.7
TM0534 RNA polymerase sigma-H factor, put {Bl} 58.2
TM1706 transcription elong factor, greA/greB fam {Bs} 63.8
TM1470 transcription term factor Rho (rho) {Tm} 100
RNA processing
TM1568 16S rRNA processing prt, put {Ec} 60.6
TM1094 RNA methylTase, put {Bs} 57.5
TM0856 tRNA pseudouridine 55 Sase (truB) {Aa} 61.5

Translation
tRNA aminoacylation
TM1396 alanyl-tRNA Sase (alaS) {Aa} 69.0
TM1093 arginyl-tRNA Sase (argS) {Bs} 67.6
TM1441 aspartyl-tRNA Sase (aspS) {Aa} 75.0
TM0719 cysteinyl-tRNA Sase (cysS) {Aa} 69.3
TM1272 glutamyl tRNA-Gln amidoTase, sub A (gatA) {Bs} 74.3
TM1273 glutamyl tRNA-Gln amidoTase, sub B (gatB) {Aa} 75.2
TM0252 glutamyl tRNA-Gln amidoTase, sub C (gatC) {Bs} 63.2
TM1351 glutamyl-tRNA Sase (gltX-1) {Aa} 67.9
TM1875 glutamyl-tRNA Sase (gltX-2) {Aa} 65.4
TM0216 glycyl-tRNA Sase, alpha sub (glyQ) {Aa} 82.9
TM0217 glycyl-tRNA Sase, beta sub (glyS) {Bs} 57.0
TM1090 histidyl-tRNA Sase (hisS) {Bs} 66.8
TM1043 histidyl-tRNA Sase-rel prt {Bs} 50.8
TM1361 isoleucyl-tRNA Sase (ileS) {Tm} 100
TM0168 leucyl-tRNA Sase (leuS) {SPCC} 69.1
TM1705 lysyl-tRNA Sase (lysS) {Sa} 72.7
TM0528 methionyl-tRNA formylTase (fmt) {Ec} 61.7
TM1085 methionyl-tRNA Sase (metS) {Tm} 100
TM0821 phenylalanyl-tRNA Sase, alpha sub (pheS) {Bs} 73.3
TM0822 phenylalanyl-tRNA Sase, beta sub (pheT) {Aa} 61.4
TM0514 prolyl-tRNA Sase (proS) {Aa} 69.6
TM1379 seryl-tRNA Sase (serS) {Aa} 77.1
TM0740 threonyl-tRNA Sase (thrS) {Bs} 73.5
TM0492 tryptophanyl-tRNA Sase (trpS) {Bb} 62.8
TM0478 tyrosyl-tRNA Sase (tyrS) {Aa} 76.9
TM1817 valyl-tRNA Sase (valS) {Bst} 72.4
Degradation of proteins, peptides, and glycopeptides
TM0042 aminopeptidase P, put {Ll} 63.6
TM0365 aminopeptidase, put {Bb} 69.1
TM0198 ATP-dep Clp protease, ATPase sub (clpC-1) {Scy} 74.4
TM0873 ATP-dep Clp protease, ATPase sub (clpC-2), AFS {Scy} 74.0
TM1391 ATP-dep Clp protease, ATPase sub (clpC-3) {Scy} 74.0
TM0146 ATP-dep Clp protease, ATPase sub clpX {Ec} 81.7
TM0695 ATP-dep Clp protease, proteolytic sub (clpP) {Aa} 86.1
TM1633 ATP-dep protease LA (lon) {Bbv} 71.9
TM1869 ATP-dep protease LA, put {Hi} 55.6
TM0747 carboxyl-terminal protease {Aa} 62.6
TM1589 clostripain-rel prt {Chi} 50.3
TM0516 clostripain-rel prt {Chi} 56.0
TM0643 clostripain-rel prt {Chi} 49.0
TM1823 ftsH protease activity modulator HflC (hflC) {Vp} 62.0
TM1822 ftsH protease activity modulator HflK (hflK) {Vp} 58.4
TM0571 heat shock serine protease, periplasmic (htrA) {Ba} 61.1
TM0963 oligoendopeptidase, put {Bb} 49.9
TM1346 processing protease, put {Bs} 51.1
TM1713 proline dipeptidase, put {Bb} 59.2
TM0145 secreted metalloendopeptidase Gcp, put {Pha} 65.8
Protein modification
TM0704 L-isoaspartate(D-aspartate) O-methylTase, put {Tm} 100
TM0463 lipoprt signal peptidase {Aa} 61.2
TM1478 methionine aminopeptidase (map) {Bs} 69.4
TM1661 polypeptide deformylase (def) {Tm} 100
TM0158 prolipoprt diacylglyceryl Tase (lgt) {SPCC} 68.0
TM0109 pyruvate formate lyase activating enzyme, put {Tl} 59.0
TM1552 pyruvate formate-lyase activating enzyme, put {Af} 68.0
Ribosomal proteins: synthesis and modification
TM0264 16S pseudouridylate Sase (rsuA) {Ec} 59.8
TM0940 ribosomal large sub pseudouridine Sase C (rluC) {Ec} 57.7
TM0455 ribosomal prt L1 (rplA) {Tm} 100
TM1497 ribosomal prt L2 (rplB) {Tm} 99.3
TM1500 ribosomal prt L3 (rplC) {Tm} 100
TM1499 ribosomal prt L4 (rplD) {Tm} 99.1
TM1488 ribosomal prt L5 (rplE) {Tm} 99.5
TM1485 ribosomal prt L6 (rplF) {Bst} 72.2
TM0457 ribosomal prt L7/L12 (rplL) {Tm} 100
TM0863 ribosomal prt L9 (rplI) {Bst} 64.9
TM0456 ribosomal prt L10 (rplJ) {Tm} 99.4
TM0454 ribosomal prt L11 (rplK) {Tm} 100
TM1079 ribosomal prt L11 methylTase, put {Aa} 61.2
TM1454 ribosomal prt L13 (rplM) {Aa} 83.0
TM1490 ribosomal prt L14 (rplN) {Tm} 100
TM1481 ribosomal prt L15 (rplO) {Bst} 75.5
TM1493 ribosomal prt L16 (rplP) {Tm} 100
TM1471 ribosomal prt L17 (rplQ) {Hp} 65.0
TM1484 ribosomal prt L18 (rplR) {Aa} 84.6
TM1571 ribosomal prt L19 (rplS) {Bst} 86.8
TM1592 ribosomal prt L20 (rplT) {Bs} 69.6
TM1458 ribosomal prt L21 (rplU) {Bs} 67.0
TM1495 ribosomal prt L22 (rplV) {Tm} 99.4
TM1498 ribosomal prt L23 (rplW) {Tm} 100
TM1489 ribosomal prt L24 (rplX) {Tm} 100
TM1456 ribosomal prt L27 (rpmA) {Ec} 75.0
TM0255 ribosomal prt L28 (rpmB) {Aa} 78.9
TM1492 ribosomal prt L29 (rpmC) {Tm} 100
TM1482 ribosomal prt L30 (rpmD) {Bst} 75.0
TM1684 ribosomal prt L31 (rpmE) {Tm} 100
TM0150 ribosomal prt L32 (rpmF) {Aa} 72.4
TM0451 ribosomal prt L33 (rpmG) {Tm} 100
TM1591 ribosomal prt L35 (rpmI) {Bs} 65.6
TM1476 ribosomal prt L36 (rpmJ) {SPCC} 81.6
TM1445 ribosomal prt S1 (rpsA), AFS {Hi} 55.3
TM0762 ribosomal prt S2 (rpsB) {Bs} 81.4
TM1494 ribosomal prt S3 (rpsC) {Tm} 100
TM1473 ribosomal prt S4 (rpsD) {Hp} 69.9
TM1483 ribosomal prt S5 (rpsE) {Bst} 76.9
TM0603 ribosomal prt S6 (rpsF) {Aa} 59.0
TM1504 ribosomal prt S7 (rpsG) {Tm} 100
TM1486 ribosomal prt S8 (rpsH) {Bs} 78.8
TM1453 ribosomal prt S9 (rpsI) {Bst} 75.0
TM1501 ribosomal prt S10 (rpsJ) {Tm} 100
TM1474 ribosomal prt S11 (rpsK) {Bac} 78.7
TM1505 ribosomal prt S12 (rpsL) {Aa} 95.2
TM1475 ribosomal prt S13 (rpsM) {Aa} 80.0
TM1487 ribosomal prt S14 (rpsN) {Ta} 85.2
TM1344 ribosomal prt S15 (rpsO) {Bs} 78.7
TM1566 ribosomal prt S16 (rpsP) {Bs} 73.0
TM1491 ribosomal prt S17 (rpsQ) {Tm} 99.1
TM0605 ribosomal prt S18 (rpsR) {Bst} 75.0
TM1496 ribosomal prt S19 (rpsS) {Tm} 97.9
TM1657 ribosomal prt S20 (rpsT) {Bs} 61.2
tRNA modification
TM1574 pseudouridylate Sase I (hisT) {Bs} 59.2
TM1463 ribonuclease P prt component (rnpA) {Pp} 65.6
TM0574 S-adenosylmethionine tRNA ribosylTase (queA) {Bs} 70.3
TM0520 tRNA (5-methylaminomethyl-2-thiouridylate)-methylTase (trmU) {Bs} 58.5
TM0525 tRNA delta-2-isopentenylpyroP Tase (miaA) {Aa} 65.5
TM1561 tRNA guanine transglycosylase (tgt) {Bs} 69.9
TM1569 tRNA guanine-N1 methylTase (trmD) {Bs} 68.7
Translation factors
TM1363 peptide chain release factor RF-1 (prfA) {Aa} 75.6
TM1579 peptide chain release factor RF-2 (prfB) {Aa} 78.3
TM1399 ribosome recycling factor (frr) {Bs} 69.6
TM1503 translation elong factor G (fus-1) {Tm} 99.1
TM1651 translation elong factor G (fus-2) {Aa} 66.4
TM1763 translation elong factor P (efp) {Mt} 63.2
TM1605 translation elong factor Ts (tsf) {Ta} 75.6
TM1502 translation elong factor Tu (tuf) {Tm} 99.8
TM1477 translation init factor IF-1 (infA), AFS {Bs} 84.9
TM0775 translation init factor IF-2 (infB) {Aa} 69.8
TM1590 translation init factor IF-3 (infC) {Aa} 74.4
TM0911 translation init factor, aIF-2B alpha sub-rel {Af} 70.7
TM1440 translation init factor, eIF-2B alpha sub-rel {Rn} 53.0
Other
TM1626 peptidyl-tRNA hydrolase {SPCC} 62.4
TM0215 prt synthesis inhibitor, put {Bs} 71.8
TM0855 ribosome binding factor A {Sau} 57.5

Regulatory functions
TM0729 (p)ppGpp Sase (relA) {Bs} 64.5
TM1081 anti-sigma factor antagonist, put {Bl} 64.2
TM1442 anti-sigma factor antagonist, put {Bl} 53.3
TM0371 arginine repressor (argR) {Bs} 62.0
TM0251 carbon storage reg (csrA) {Ec} 78.3
TM0122 ferric uptake reg prt (fur-1) {Cj} 57.1
TM1515 ferric uptake reg prt (fur-2) {SPCC} 59.7
TM1776 ferric uptake reg prt (fur-3) {Ng} 56.5
TM1431 glycerol uptake operon antiterminator (glpP) {Bs} 61.1
TM1436 glycerol uptake operon antiterminator-rel prt {Bs} 55.2
TM0195 guanosine pentaP phosphohydrolase, put {Hp} 49.6
TM0851 heat shock operon repressor HrcA {Bs} 53.5
TM0510 Fe-dep transcriptional repressor, put {Mt} 62.6
TM0602 Fe-dep transcriptional repressor, put {Mt} 45.0
TM1330 lacI fam transcriptional reg, put {Mj} 70.4
TM1082 lexA repressor (lexA) {Tm} 100
TM1866 membrane bound prt LytR, put {Bs} 52.3
TM0403 nitrogen reg prt P-II (glnK) {Av} 69.6
TM1259 P regulon transcriptional reg prt PhoB (phoB) {Sd} 63.6
TM1260 P transport system reg PhoU (phoU) {Pa} 55.1
TM1734 P transport system reg PhoU, put {Aa} 57.7
TM1184 pleD-rel prt {SPCC} 50.8
TM0668 pleiotropic reg prt (degT) {Bst} 74.5
TM0467 reg prt, put {SPCC} 54.4
TM0490 reg prt, SIR2 fam {Af} 64.2
TM1655 response reg DrrA (drrA) {Tm} 100
TM0399 response reg {Ae} 65.9
TM1360 response reg {Bs} 65.8
TM0126 response reg {Bs} 60.1
TM0186 response reg {SPCC} 60.5
TM0842 response reg {SPCC} 56.7
TM0468 response reg {Tm} 69.2
TM0143 response regulator/GGDEF domain {SPCC} 53.0

TM1654 sensor histidine kinase HpkA (hpkA) {Tm} 100
TM1359 sensor histidine kinase {Ec} 56.6
TM0127 sensor histidine kinase {Pa} 54.1
TM0187 sensor histidine kinase {Syn} 55.0
TM0400 sensor histidine kinase {SPCC} 54.2
TM0853 sensor histidine kinase {Tm} 61.1
TM1258 sensor histidine kinase, PhoR-rel {Bs} 57.0
TM0742 serine/threonine prt PPase {Ma} 57.1
TM0733 sigma-B reg, put {Bl} 49.6
TM0565 sugar fermentation stimulation prt, put {Mta} 64.1
TM1602 transcriptional reg, biotin repressor fam {Bs} 59.1
TM1171 transcriptional reg, crp fam {Ps} 50.6
TM1069 transcriptional reg, DeoR fam {Bs} 55.0
TM0275 transcriptional reg, GntR fam {Bs} 54.0
TM0439 transcriptional reg, GntR fam {Bs} 58.3
TM0766 transcriptional reg, GntR fam {Bs} 65.0
TM0065 transcriptional reg, IclR fam {Ech} 59.1
TM1200 transcriptional reg, LacI fam {Bs} 61.0
TM1218 transcriptional reg, LacI fam {Bs} 53.3
TM1856 transcriptional reg, LacI fam {Bs} 52.0
TM0299 transcriptional reg, LacI fam {Ec} 58.7
TM0949 transcriptional reg, LacI fam {Hi} 55.2
TM0710 transcriptional reg, MarR fam {Bs} 57.0
TM1176 transcriptional reg, metal-sensing {Mta} 63.9
TM1580 transcriptional reg, put {Aa} 59.4
TM1005 transcriptional reg, put {Ec} 57.2
TM0816 transcriptional reg, put, Mar fam {Bs} 53.6
TM1228 transcriptional reg, RpiR fam {Cpe} 55.6
TM0326 transcriptional reg, RpiR fam {Ec} 56.9
TM0823 transcriptional reg, TetR fam {Aa} 52.9
TM1030 transcriptional reg, TetR fam {Bm} 48.0
TM1224 transcriptional reg, XylR-rel {Ath} 54.9
TM0032 transcriptional reg, XylR-rel {Ath} 51.2
TM0110 transcriptional reg, XylR-rel {Ath} 59.5
TM0411 transcriptional reg, XylR-rel {Bm} 54.2
TM0808 transcriptional reg, XylR-rel {Bm} 52.3
TM0393 transcriptional reg, XylR-rel {Bs} 52.3

Cell envelope

Surface structures

TM0673 basal-body rod modification prt FlgD (flgD) {Ec} 56.7
TM1364 flgr basal-body rod prt FlgB (flgB) {Bb} 67.5
TM1365 flgr basal-body rod prt FlgC (flgC) {Bb} 72.5
TM1543 flgr basal-body rod prt FlgF (flgF) {Cc} 56.1
TM1542 flgr basal-body rod prt FlgG (flgG) {Aa} 75.2
TM0908 flgr biosynth prt FlhA (flhA) {Bs} 71.1
TM0909 flgr biosynth prt FlhB (flhB) {Bs} 70.9
TM0907 flgr biosynth prt FlhF (flhF) {Bs} 65.5
TM0698 flgr biosynth prt FliP (fliP) {Aa} 75.2
TM0697 flgr biosynth prt FliQ (fliQ) {Bs} 66.3
TM0910 flgr biosynth prt FliR (fliR) {Bs} 55.9
TM0699 flgr biosynth prt FliZ, put {Bb} 57.0
TM0219 flgr export/assembly prt (fliH) {Td} 58.2
TM0674 flgr hook prt FlgE (flgE), AFS {Bb} 61.1
TM0083 flgr hook-associated prt 1 (flgK) {Bb} 49.9
TM1123 flgr hook-associated prt 2, put {Bs} 53.5
TM0082 flgr hook-associated prt 3 (flgL) {Bs} 56.4
TM1366 flgr hook-basal body complex prt FliE (fliE) {Bs} 66.7
TM1540 flgr L-ring prt (flgH) {Aa} 59.2
TM0221 flgr M-ring prt (fliF) {Ec} 51.7
TM0220 flgr motor switch prt FliG (fliG) {Bs} 79.9
TM0679 flgr motor switch prt FliM (fliM) {Bs} 59.7
TM0680 flgr motor switch prt FliY (fliY), AFS {Bs} 65.7
TM1539 flgr P-ring prt (flgI) {Aa} 59.6
TM1541 flgr prt FlgA, put {Ar} 48.0
TM1179 flgr prt FliS (fliS) {Sch} 59.5
TM0678 flgr prt, put {Aa} 41.8
TM0675 flgr prt, put {Bb} 70.1
TM0758 flagellin {Cac} 74.6
TM0132 flagellin, put {Bac} 56.4
TM0218 flagellum-specific ATP Sase (fliI) {Bs} 76.6
TM0676 motility prt A (motA) {Bm} 67.5
TM0677 motility prt B (motB) {Bf} 56.3
TM1362 motility prt PilT (pilT) {Aa} 66.7
TM0841 S-layer-like array prt (slpM) {Ta} 56.1
TM1271 type IV pilin-rel prt {Eco} 54.2

Biosynthesis of murein sacculus and peptidoglycan

TM0259 D-alanine—D-alanine ligase {Tm} 100
TM0148 glucosamine—fructose-6-P aminoTase, isomerizing (glmS) {Ta} 65.6
TM0590 penicillin-Bprt 2 {Aa} 55.3
TM0870 penicillin-Bprt 2 {Bs} 52.2
TM0886 penicillin-Bprt, class 1A {Pa} 55.6
TM0235 phospho-N-acetylmuramoyl-pentapeptide-Tase (mraY) {Ec} 65.6
TM1714 UDP-N-acetylenolpyruvoylglucosamine RDase (murB) {Bs} 64.5
TM0108 UDP-N-acetylglucosamine 1-carboxyvinylTase (murA) {Ec} 66.0
TM1629 UDP-N-acetylglucosamine pyrophosphorylase (glmU) {Bs} 68.0
TM0232 UDP-N-acetylglucosamine—N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine Tase (murG) {Bs} 57.0
TM0231 UDP-N-acetylmuramate—alanine ligase (murC) {Aa} 53.5
TM0234 UDP-N-acetylmuramoylalanine—D-glutamate ligase (murD) {Aa} 54.0
TM0237 UDP-N-acetylmuramoylalanyl-D-glutamate—2,6-diaminopimelate ligase (murE) {Aa} 60.4
TM0236 UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate—D-alanyl-D-alanyl ligase (murF) {Aa} 59.7

Biosynthesis of surface polysaccharides and lipopolysaccharides

TM0862 glucose-1-P thymidylylTase (rfbA) {SPCC} 71.3
TM0585 LPS biosynth prt BplA (wlbA) {Bp} 58.4
TM0627 LPS biosynth prt {Aa} 60.9
TM0610 LPS biosynth prt {Bp} 66.3
TM0583 LPS biosynth prt {Ph} 72.0
TM1548 LPS biosynth prt {Sa} 61.2
TM0631 LPS biosynth prt {Sa} 45.9
TM0620 LPS biosynth prt {St} 50.4
TM0622 LPS biosynth prt, put {Aa} 57.7
TM0760 LPS biosynth prt, put {Pa} 44.9
TM0818 LPS biosynth prt, put {Sp} 60.4
TM0621 LPS biosynth prt, put {St} 52.8
TM0572 LPS biosynth prt, put {SPCC} 63.9
TM0741 LPS core biosynth prt KdtB (kdtB) {Ec} 71.6
TM1033 mannose-1-P guanylylTase (manC) {Ec} 60.4
TM0638 polysaccharide export prt, put {Vc} 49.5
TM1034 UDP-N-acetylglucosamine 2-epimerase (wecB) {Ec} 74.6

Other

TM0102 basic membrane prt {Bb} 56.7
TM0912 basic membrane prt, put {Bb} 50.9
TM1053 cationic outer membrane prt (ompH) {Ye} 53.8
TM1461 inner membrane prt, put {Ec} 58.8
TM0477 outer membrane prt alpha {Tm} 99.8
TM1729 outer membrane prt {Tm} 54.1

Cellular processes

Cell division

TM1277 cell division prt FtsA, put {Aa} 49.7
TM0835 cell division prt FtsA, put {Aa} 51.3
TM0580 cell division prt FtsH (ftsH) {Bs} 74.3
TM0570 cell division prt FtsY (ftsY) {Bs} 74.0
TM0836 cell division prt FtsZ (ftsZ) {Tm} 99.7
TM0233 cell division prt, rodA/ftsW/spoVE fam {Hp} 58.4
TM1182 chromosome segregation SMC prt, put {Bs} 59.3
TM1606 cytoplasmic axial filament prt, put {Ec} 58.7
TM0588 rod shape-determining prt MreB (mreB-1) {Bs} 77.1
TM1544 rod shape-determining prt MreB (mreB-2) {Bce} 74.6
TM0839 rod shape-determining prt RodA (rodA) {Aa} 63.0
TM1047 septum site-determining prt MinC, put {Aa} 65.4
TM1870 septum site-determining prt MinD (minD) {Bs} 75.1

Chaperones

TM0849 dnaJ prt (dnaJ) {Aa} 63.3
TM0373 dnaK prt (dnaK) {Cac} 76.0
TM0506 groEL prt (groEL) {Tb} 84.9
TM0505 groES prt (groES) {Pg} 74.2
TM0850 grpE prt, put {Mta} 71.0
TM0522 heat shock prt HslU (hslU) {Aa} 81.6
TM0521 heat shock prt HslV (hslV) {Aa} 81.8
TM0374 heat shock prt, class I (hsp) {Aa} 70.4

Detoxification

TM0807 alkyl hydroperoxide RDase, put {Mj} 88.0
TM0998 heavy metal resistance transcriptional reg {Sa} 67.5
TM0379 NADH oxidase (nox) {Af} 70.4
TM0395 NADH oxidase, put {Af} 66.4
TM1056 periplasmic divalent cation tolerance prt (cutA) {Af} 66.3

Protein and peptide secretion

TM1117 general secretion pathway prt D, put {Pa} 47.2
TM0837 general secretion pathway prt E {Aa} 67.4
TM0094 general secretion pathway prt F, put {Aa} 58.0
TM1578 preprt translocase SecA sub (secA) {Aa} 74.4
TM0452 preprt translocase SecE sub (secE) {Tm} 100
TM1480 preprt translocase SecY sub (secY) {Aa} 70.7
TM0860 prt-export membrane prt SecD, put {Aa} 61.2
TM0861 prt-export membrane prt SecF, put {Aa} 63.3
TM1572 signal peptidase I, put {SPCC} 61.9
TM1565 signal recognition particle prt (ffh) {Hi} 76.7
TM0694 trigger factor, put {Bs} 55.3
TM1696 type IV prepilin peptidase {Aa} 61.9

Transformation

TM0088 comE prt, put {Hi} 56.7
TM1052 comEA prt-rel prt {Bs} 75.8
TM1584 comFC prt, put {Bs} 49.7
TM0513 comM prt {Hi} 64.3
TM0250 DNA processing chain A (dprA) {Hi} 61.2

Toxin production and resistance

TM0893 bacitracin resistance prt (bacA) {Ec} 61.2
TM0785 bacteriocin {Bli} 55.0
TM1576 hemolysin (tly) {Aa} 71.1
TM0805 lipophilic prt, put {SPCC} 57.6
TM1444 lytB prt (lytB) {Cj} 60.8
TM1549 methicillin resistance prt (llm) {Sa} 60.1
TM1774 phosphonopyruvate decarboxylase, put {Aa} 70.6

Other

TM0464 chemotactic methylTase (cheR) {Bs} 65.9
TM0903 chemotaxis methylation prt (cheD) {Ph} 69.2
TM0904 chemotaxis prt CheC (cheC) {Bs} 61.2
TM0700 chemotaxis response reg CheY (cheY) {Tm} 100
TM0702 chemotaxis sensor histidine kinase CheA (cheA) {Tm} 99.9
TM1143 methyl-accepting chemotaxis prt {Bs} 53.9
TM1146 methyl-accepting chemotaxis prt {Bs} 57.1
TM0023 methyl-accepting chemotaxis prt {Bs} 51.1
TM0429 methyl-accepting chemotaxis prt {Bs} 51.0
TM0918 methyl-accepting chemotaxis prt {Bs} 53.1
TM1428 methyl-accepting chemotaxis prt {Rc} 52.6
TM0014 methyl-accepting chemotaxis prt, put {Tm} 99.6
TM0408 prt-glutamate methylesterase (cheB) {Af} 69.2
TM0701 purine-binding chemotaxis prt (cheW-1) {Tm} 100
TM0718 purine-binding chemotaxis prt (cheW-2) {Ec} 61.4

Other categories

Adaptations and atypical conditions

TM1683 cold shock prt (cspB) {Tm} 100
TM1874 cold shock prt (cspL) {Tm} 84.8
TM1627 general stress prt Ctc (ctc) {Bs} 48.0
TM1662 stationary phase survival prt (surE) {Tm} 100

Phage-related functions and prophages

TM0526 host factor I (hfq) {Aa} 78.7

Transposon-related functions

TM1832 transposase {Vc} 62.4
TM0777 transposase {Vc} 63.2
TM0048 transposase {Vc} 69.9
TM0131 transposase, AFS {Cpe} 69.9
TM1044 transposase, IS605-TnpB fam (tnpB) {Hp} 62.8
TM0135 transposase, put {Ec} 52.8
TM1026 transposase, put {Mj} 55.7
TM1677 transposase, put {Mj} 57.0
TM1831 transposase, put {Mj} 55.7
TM0047 transposase, put {Mj} 56.8
TM0776 transposase, put {Mj} 56.4
TM1003 transposase-rel prt {Mj} 52.8

Other

TM0884 methylTase, put {Mo} 48.9

Unknown

TM0482 (R)-2-hydroxyglutaryl-CoA dehydratase activator-rel prt {Mj} 53.5
TM1519 2,3,4,5-tetrahydropyridine-2-carboxylate N-succinylTase-rel prt {Bs} 79.4
TM0435 acetyl xylan esterase-rel prt {Tn} 86.1
TM1178 acetylTase-rel prt {Aa} 50.4
TM0157 actinorhodin polyketide dimerase-rel prt {Sco} 50.4
TM1726 alanyl-tRNA Sase-rel prt {Hi} 50.2
TM1231 alpha-mannosidase-rel prt {Rn} 50.2
TM0384 anaerobic ribonucleoside-triP RDase-rel prt {T4} 53.9
TM0285 araM prt, put {Bs} 53.4
TM0824 astB/chuR-rel prt {Af} 51.5
TM0825 astB/chuR-rel prt {Af} 55.5
TM1317 astB/chuR-rel prt {Bt} 47.0
TM1301 astB/chuR-rel prt {Ec} 49.5
TM1324 astB/chuR-rel prt {Mj} 58.4
TM1325 astB/chuR-rel prt {Mj} 50.2
TM0780 bacterioferritin comigratory prt, ahpC/TSA fam {Hi} 56.9
TM0386 bacterioferritin comigratory prt/NADH DHase {SPCC} 59.4
TM1846 beta transducin-rel prt {Pan} 49.5
TM1618 cheX prt {Tp} 52.0
TM0416 D-tagatose 3-epimerase-rel prt {Pc} 52.7
TM0681 dehydrase-rel prt {Af} 49.5
TM0561 divalent cation transport-rel prt {Af} 71.2
TM0771 DNA polymerase III, gamma sub-rel prt {Aa} 56.3
TM1803 dnaJ-rel prt {Bst} 65.4
TM1753 excinuclease ABC, sub B-rel prt {Aa} 85.7
TM0632 extracellular polysaccharide biosynth-rel prt {Ec} 62.2
TM1672 flgr biosynth-rel prt {Bs} 67.9
TM0633 flgr-rel prt {Vc} 54.8
TM0756 galactosylTase-rel prt {Sp} 61.8
TM0891 gcpE prt {Bs} 71.1
TM1058 glutamate Sase-rel prt {Mta} 48.7
TM1738 gufA prt {Mx} 62.8
TM0159 ham1 prt {Aa} 63.7
TM1411 helicase-rel prt {Ph} 82.8
TM0845 hemolysin-rel prt {Bb} 61.6
TM1354 inosine-5-monoP DHase-rel prt {Mj} 60.3
TM1415 inositol monoPPase fam prt, put {Tm} 100
TM0133 isochorismatase-rel prt {Bs} 52.0
TM1460 jag prt, put {Bs} 53.9
TM0961 lemA prt {Lmo} 61.6
TM1623 lepA prt (lepA) {Aa} 76.6
TM1405 LPS biosynth prt-rel prt {Mj} 53.7
TM0671 M-rel prt {Streptococcus pyogenes} 47.4
TM1556 maf prt (maf) {Aa} 66.5
TM0913 mazG prt {Hi} 70.2
TM0360 mazG-rel prt {Hi} 57.4
TM1833 methy-accepting chemotaxis-rel prt {Bs} 51.1
TM0973 methyl-accepting chemoreceptor-rel prt {Dv} 71.1
TM1537 Mg-protoporphyrin IX monomethyl ester oxidative cyclase-rel prt {Mta} 55.9
TM1132 moxR prt, put {Bs} 77.1
TM1419 myo-inositol-1-P Sase-rel prt {Af} 50.9
TM0624 N-acetylglucosaminyl-phosphatidylinositol biosynth-rel prt {Sc} 43.5
TM1372 nitrogen fixation prt NifU-rel prt {Ml} 55.1
TM1284 oxidase-rel prt {Sco} 53.5
TM1340 oxidase-rel prt {Sco} 57.7
TM1507 phoH-rel prt {Aa} 72.1
TM0495 phoH-rel prt {Bs} 70.3
TM0184 phosphoglucomutase/phosphomannomutase fam prt {Bs} 67.7
TM0965 PPRaminoimidazole carboxylase-rel prt {Mj} 63.5
TM0426 PHT4-rel prt {Pp} 54.7
TM0727 pmbA-rel prt {Aa} 52.1
TM0475 pyrazinamidase/nicotinamidase-rel prt {Af} 66.1
TM0878 pyruvate:ferredoxin 2-oxidoRDase-rel prt {Ph} 68.2
TM0696 ray-rel prt {Mmu} 59.0
TM1858 recX prt, put (recX) {Ms} 50.0
TM1296 ribonuclease H-rel prt {Sc} 54.0
TM1847 ROK fam prt {Bs} 52.9
TM0254 small prt B, AFS {Ef} 75.3
TM1059 spoVS-rel prt {Bs} 91.9
TM0192 spoVS-rel prt {Bs} 93.0
TM0897 spoVS-rel prt {Bs} 73.5
TM1512 sun prt {Bs} 57.3
TM0267 thiophene oxidation prt ThdF-rel GTPase {Aa} 69.9
TM0134 thioredoxin RDase-rel prt {Eac} 54.1
TM0449 thy1 prt {Bb} 67.3
TM0295 transaldolase-rel prt {Aa} 85.1
TM0715 tRNA nucleotidyl Tase-rel prt {Aa} 56.7
TM1389 ubiquinone/menaquinone biosynth methylTase-rel prt {Mj} 53.7
TM0318 ubiquinone/menaquinone biosynth-rel prt {Af} 65.0
TM0751 uridine kinase-rel prt {Mp} 52.1
TM0722 vacB prt (vacB) {Aa} 61.2
TM0086 virulence factor MviN-rel prt {Sch} 56.0
TM0113 xylU-rel prt {Ct} 67.7