17. Fülöp, V., Moir, J. W. B., Ferguson, S. J. & Hajdu, J. Crystallisation and preliminary crystallographic study of cytochrome cd_1 nitrite reductase from *Thiosphaera pantotropha*. *J. Mol. Biol.* **232,** 1211–1212 (1993).
18. Berger, H. & Wharton, D. C. Small angle X-ray scattering studies of oxidised and reduced cytochrome oxidase from *Pseudomonas aeruginosa*. *Biochim. Biophys. Acta* **622,** 355–359 (1980).
19. Moore, G. R. & Pettigrew, G. W. *Cytochromes c: Evolutionary, Structural and Physicochemical Aspects* (Springer, Berlin, 1990).
20. Pettigrew, G. W. & Moore, G. R. *Cytochromes c: Biological Aspects* (Springer, Berlin, 1987).
21. Harutunyan, E. H. *et al*. The binding of carbon monoxide and nitric oxide to leghaemoglobin in comparison with other haemoglobins. *J. Mol. Biol.* **264,** 152–161 (1996).
22. Edwards, S. L., Kraut, J. & Poulos, T. L. Crystal structure of nitric oxide inhibited cytochrome-c peroxidase. *Biochemistry* **27,** 8074–8081 (1988).
23. Adman, E. T., Godden, J. W. & Turley, S. The structure of copper nitrite reductase from *Achromobacter cycloclastes* at five pH values, with NO_2^- bound and with type II copper depleted. *J. Biol. Chem.* **270,** 27458–27474 (1995).
24. Williams, P. A. thesis, Oxford Univ. (1996).
25. Poulos, T. L. Ligands and electrons and haem proteins. *Nature Struct. Biol.* **3,** 401–403 (1996).
26. Wittung, P. & Malmstrom, B. G. Redox-linked conformational changes in cytochrome *c* oxidase. *FEBS Lett.* **388,** 47–49 (1996).
27. Pascher, T., Chesick, J. P., Winkler, J. R. & Gray, H. B. Protein folding triggered by electron transfer. *Science* **271,** 1558–1560 (1996).
28. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein. *J. Appl. Crystallogr.* **24,** 946–950 (1991).
29. Merritt, E. A. & Murphy, M. E. P. Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D* **50,** 869–873 (1994).
30. Brünger, A. T. The free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355,** 472–474 (1992).

Acknowledgements. We thank the ESRF and SRS Daresbury for data collection facilities; the EMBL outstation, Grenoble, for use of an image plate detector; M.L.D. Page for expert advice; R. Bryan and R. Esnouf for computing; K. Harlos for help with in-house data collection; F. Armstrong and J. Hirst for providing electrochemically reduced methyl viologen. This work was supported by MRC, BBSRC and EU-BIOTECH. The Oxford Centre for Molecular Sciences is funded jointly by BBSRC, EPSRC and MRC. N.F.W.S. was supported by a Wellcome Trust prize studentship. V.F. is a Royal Society university research fellow.

Correspondence and requests for materials should be addressed to P.A.W. (e-mail: pamela@scripps.edu), V.F. (e-mail: vilmos@biop.ox.ac.uk) or J.H. (e-mail: janos@xray.bmc.uu.se).

errata

The yeast genome directory

Nature **387** (suppl.) (1997)

In the list of authors given on page 5 of this supplement, the names of some authors were omitted or misspelled (asterisks). These were: R. Altmann; W. Arnold*; M. de Haan*; K. Hamberg; K. Hinni; L. Jones; W. Kramer; H. Küster*; K. C. T. Maurer*; D. Niblett; N. Paricio*; A. G. Parle-McDermott*; C. Rebischung; C. Richards; L. Rifkin*; J. Robben; C. Rodrigues-Pousada*; I. Schaaff-Gerstenschläger*; P. H. M. Smits*; Y. Su*; Q. J. M. van der Aart*; J. C. van Vliet-Reedijk*; A. Wach; M. Yamazaki*. □

Measurements of elastic anisotropy due to solidification texturing and the implications for the Earth's inner core

Michael I. Bergman

Nature **389,** 60–63 (1997)

Owing to a typographical error, this Letter appeared under the title "Measurements of electric anisotropy due to solidification texturing and the implications for the Earth's inner core". The word 'elastic' in the first line was erroneously replaced with 'electric'. □

cAMP-induced switching in turning direction of nerve growth cones

Hong-jun Song, Guo-li Ming & Mu-ming Poo

Nature **388,** 275–279 (1997)

The order of panels in Fig. 3 of this Letter is incorrect as published. Figure 3a–e should be labelled as f–j, and Fig. 3f–j should be labelled a–e. □

corrections

Synthesis and X-ray structure of dumb-bell-shaped C_{120}

Guan-Wu Wang, Koichi Komatsu, Yasujiro Murata & Motoo Shiro

Nature **387,** 583–586 (1997)

In this Letter, we overlooked a citation of G. Oszlanyi *et al*., *Phys. Rev. B* **54,** 11849 (1996), who reported the observation of covalently bound $(C_{60})_2^{2-}$ dianions from the X-ray powder diffraction patterns of the metastable phases of KC_{60} and RbC_{60}. □

The complete genome sequence of the gastric pathogen *Helicobacter pylori*

Jean-F. Tomb, Owen White, Anthony R. Kerlavage, Rebecca A. Clayton, Granger G. Sutton, Robert D. Fleischmann, Karen A. Ketchum, Hans Peter Klenk, Steven Gill, Brian A. Dougherty, Karen Nelson, John Quackenbush, Lixin Zhou, Ewen F. Kirkness, Scott Peterson, Brendan Loftus, Delwood Richardson, Robert Dodson, Hanif G. Khalak, Anna Glodek, Keith McKenney, Lisa M. Fitzegerald, Norman Lee, Mark D. Adams, Erin K. Hickey, Douglas E. Berg, Jeanine D. Gocayne, Teresa R. Utterback, Jeremy D. Peterson, Jenny M. Kelley, Matthew D. Cotton, Janice M. Weidman, Claire Fujii, Cheryl Bowman, Larry Watthey, Erik Wallin, William S. Hayes, Mark Borodovsky, Peter D. Karp, Hamilton O. Smith, Claire M. Fraser & J. Craig Venter

Nature **388,** 539–547 (1997)

In this Article, we incorrectly stated that the amino acids lysine and arginine are twice as abundant in *H. pylori* proteins as they are in those of *Haemophilus influenzae* and *Escherichia coli*. This statement was derived from amino-acid analyses that compared absolute differences in abundance, but these do not reflect the frequencies with which amino acids are found in the organisms in question. The actual abundance of arginine in *H. pylori*, *H. influenzae* and *E. coli* is 3.5, 4.5 and 5.5%, respectively; the abundance of lysine in these organisms is 8.9, 6.3 and 4.4%, respectively. This oversight is particularly unfortunate because Russell H. Doolittle, who wrote an accompanying News and Views on our Article and brought this to our attention, was led to comment on the significance of our inaccurate observation. We regret this and any other misunderstanding that our error may have caused. □

The complete genome sequence of the gastric pathogen *Helicobacter pylori*

Jean-F. Tomb*, Owen White*, Anthony R. Kerlavage*, Rebecca A. Clayton*, Granger G. Sutton*, Robert D. Fleischmann*, Karen A. Ketchum*, Hans Peter Klenk*, Steven Gill*, Brian A. Dougherty*, Karen Nelson*, John Quackenbush*, Lixin Zhou*, Ewen F. Kirkness*, Scott Peterson*, Brendan Loftus*, Delwood Richardson*, Robert Dodson*, Hanif G. Khalak*, Anna Glodek*, Keith McKenney*, Lisa M. Fitzegerald*, Norman Lee*, Mark D. Adams*, Erin K. Hickey*, Douglas E. Berg†, Jeanine D. Gocayne*, Teresa R. Utterback*, Jeremy D. Peterson*, Jenny M. Kelley*, Matthew D. Cotton*, Janice M. Weidman*, Claire Fujii*, Cheryl Bowman*, Larry Watthey*, Erik Wallin‡, William S. Hayes§, Mark Borodovsky§, Peter D. Karp‖, Hamilton O. Smith¶, Claire M. Fraser* & J. Craig Venter*

* *The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA*
† *Department of Molecular Biology, School of Medicine, Washington University St Louis, 660 S. Euclid Avenue, St Louis, Missouri 63110, USA*
‡ *Department of Biochemistry, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm, Sweden*
§ *School of Biology, Georgia Tech, Atlanta, Georgia 30332, USA*
‖ *SRI International, Artificial Intelligence Center, 333 Ravenswood Avenue, Menlo Park, California 94025, USA*
¶ *Department of Molecular Biology and Genetics, School of Medicine, Johns Hopkins University, 725 N. Wolfe Street, Baltimore, Maryland 21205, USA*

***Helicobacter pylori*, strain 26695, has a circular genome of 1,667,867 base pairs and 1,590 predicted coding sequences. Sequence analysis indicates that *H. pylori* has well-developed systems for motility, for scavenging iron, and for DNA restriction and modification. Many putative adhesins, lipoproteins and other outer membrane proteins were identified, underscoring the potential complexity of host–pathogen interaction. Based on the large number of sequence-related genes encoding outer membrane proteins and the presence of homopolymeric tracts and dinucleotide repeats in coding sequences, *H. pylori*, like several other mucosal pathogens, probably uses recombination and slipped-strand mispairing within repeats as mechanisms for antigenic variation and adaptive evolution. Consistent with its restricted niche, *H. pylori* has a few regulatory networks, and a limited metabolic repertoire and biosynthetic capacity. Its survival in acid conditions depends, in part, on its ability to establish a positive inside-membrane potential in low pH.**

For most of this century the cause of peptic ulcer disease was thought to be stress-related and the disease to be prevalent in hyperacid producers. The discovery[1] that *Helicobacter pylori* was associated with gastric inflammation and peptic ulcer disease was initially met with scepticism. However, this discovery and subsequent studies on *H. pylori* have revolutionized our view of the gastric environment, the diseases associated with it, and the appropriate treatment regimens[2].

Helicobacter pylori is a micro-aerophilic, Gram-negative, slow-growing, spiral-shaped and flagellated organism. Its most characteristic enzyme is a potent multisubunit urease[3] that is crucial for its survival at acidic pH and for its successful colonization of the gastric environment, a site that few other microbes can colonize[2]. *H. pylori* is probably the most common chronic bacterial infection of humans, present in almost half of the world population[2]. The presence of the bacterium in the gastric mucosa is associated with chronic active gastritis and is implicated in more severe gastric diseases, including chronic atrophic gastritis (a precursor of gastric carcinomas), peptic ulceration and mucosa-associated lymphoid tissue lymphomas[2]. Disease outcome depends on many factors, including bacterial genotype, and host physiology, genotype and dietary habits[4,5]. *H. pylori* infection has also been associated with persistent diarrhoea and increased susceptibility to other infectious diseases[6].

Because of its importance as a human pathogen, our interest in its biology and evolution, and the value of complete genome sequence information for drug discovery and vaccine development, we have

Table 1 Genome features

General	
Coding regions (91.0%)	
Stable RNA (0.7%)	
Non-coding repeats (2.3%)	
Intergenic sequence (6.0%)	
RNA	
Ribosomal RNA	Coordinates
23S-5S	445,306–448,642 bp
23S-5S	1,473,557–1,473,919 bp
16S	1,209,082–1,207,584 bp
16S	1,511,138–1,512,635 bp
5S	448,041–448,618 bp
Transfer RNA	
36 species (7 clusters, 12 single genes)	
Structural RNA	
1 species (ssrD)	629,845–630,124 bp
DNA	
Insertion sequences	
IS605 13 copies (5 full-length, 8 partial)	
IS606 4 copies (2 full-length, 2 partial)	
Distinct G + C regions	Associated genes
region 1 (33% G + C) 452–479 kb	IS605, 5SRNA and repeat 7; *virB4*
region 2 (35% G + C) 539–579 kb	cag PAI (Fig. 4)
region 3 (33% G + C) 1,049–1,071 kb	IS605, 5SRNA and repeat 7
region 4 (43% G + C) 1,264–1,276 kb	β and β′ RNA polymerase, EF-G (fusA)
region 5 (33% G + C) 1,590–1,602 kb	two restriction/modification systems
Coding sequences	
1,590 coding sequences (average 945 bp)	
1,091 identified database match	
499 no database match	

sequenced the genome of a representative *H. pylori* strain by the whole-genome random sequencing method as described for *Haemophilus influenzae*[7], *Mycoplasma genitalium*[8] and *Methanococcus jannaschii*[9].

General features of the genome

Genome analysis. The genome of *H. pylori* strain 26695 consists of a circular chromosome with a size of 1,667,867 base pairs (bp) and average G + C content of 39% (Figs 1 and 2). Five regions within the genome have a significantly different G + C composition (Table 1 and Fig. 1). Two of them contain one or more copies of the insertion sequence IS605 (see below) and are flanked by a 5S ribosomal RNA sequence at one end and a 521 bp repeat (repeat 7) near the other. These two regions are also notable because they contain genes involved in DNA processing and one contains 2 orthologues of the *virB4/ptl* gene, the product of which is required for the transfer of oncogenic T-DNA of *Agrobacterium* and the secretion of the *pertussis* toxin by *Bordetella pertussis*[10]. Another region is the *cag* pathogenicity island (PAI), which is flanked by 31-bp direct repeats, and appears to be the product of lateral transfer[11].

RNA and repeat elements. Thirty-six tRNA species were identified using tRNAscan-SE[12]. These are organized into 7 clusters plus 12 single genes. Two separate sets of 23S–5S and 16S ribosomal RNA (rRNA) genes were identified, along with one orphan 5S gene and one structural RNA gene (Table 1). Associated with each of the two 23S–5S gene clusters is a 6-kilobase (kb) repeat containing a possible operon of 5 ORFs that have no database matches.

Eight repeat families (>97% identity) varying in length from 0.47 to 3.8 kb were found in the chromosome (Figs 1 and 2). Members of repeat 7 are found in intergenic regions, while the others are associated with coding sequences and may represent gene duplications. Repeats 1, 2, 3 and 6 are associated with genes that encode outer-membrane proteins (OMP) (Fig. 3).

Two distinct insertion sequence (IS) elements are present. There are five full-length copies of the previously described IS605[11,13] and two of a newly discovered element designated IS606. In addition, there are eight partial copies of IS605 and two partial copies of IS606. Both elements encode two divergently transcribed transposases (TnpA and TnpB). IS606 has less than 50% nucleotide identity with IS605 and the IS606 transposases have 29% amino-acid identity with their IS605 counterpart. Both copies of the IS606 TnpB may be non-functional owing to frameshifts.

Origin of replication. As a typical eubacterial origin of replication was not identified[14], we arbitrarily designated basepair one at the start of a 7-mer repeat, $(AGTGATT)_{26}$, that produces translational stops in all reading frames, as this repeated DNA is unlikely to contain any coding sequence.

Open reading frames. One thousand five hundred and ninety predicted coding sequences were identified. They were searched against a non-redundant protein database resulting in 1,091 putative identifications that were assigned biological roles using a classification system adapted from Riley[15] (Table 2). The 1,590 predicted genes had an average size of 945 bp, similar to that observed in other prokaryotes[7–9], and no genome-wide strand bias was observed (Fig. 2). More than 70% of the predicted proteins in *H. pylori* have a calculated isoelectric point (p*I*) greater than 7.0, compared to ~40% in *H. influenzae* and *E. coli*. The basic amino acids, arginine and lysine, occur twice as frequently in *H. pylori* proteins as in those of *H. influenzae* and *E. coli*, perhaps reflecting an adaptation of *H. pylori* to gastric acidity.

Paralagous families. Ninety-five paralogous gene families comprising 266 gene products (16% of the total) were identified (www.tigr.org/tdb/mdb/hpdb/hpdb.html). Of these, 67 (173 proteins) have an assigned role. Sixty-four have only 2 members, while the porin/adhesin-like outer membrane protein family (Fig. 2) is the largest with 32 members. The largest number of paralogues with assigned roles fall into the functional categories of cell envelope, transport and binding proteins, and proteins involved in replication. The large number of cell envelope proteins might reflect either a reduced biosynthetic capacity or a need to adapt to the challenging gastric environment.

Cell division and protein secretion

The gene content of *H. pylori* suggests that the basic mechanisms of replication, cell division and secretion are similar to those of *E. coli* and *H. influenzae*. However, important differences are noted. For example, apparently missing from the *H. pylori* genome are orthologues of DnaC, MinC, and the secretory chaperonin, SecB. In oriC-type primosome formation, the DnaB and DnaC proteins form a B–C complex that delivers the DnaB helicase to the developing primosome complex[16]. The apparent absence of DnaC in *H. pylori* suggests that either a novel mechanism for recruiting DnaB exists or a DnaC orthologue with no detectable sequence similarity is present. Similar arguments can be made for other seemingly missing important functions.

H. pylori has a classical set of bacterial chaperones (DnaK, DnaJ, CbpA, GrpE, GroEL, GroES, and HtpG). The transcriptional regulation of *H. pylori* chaperone genes is likely to be different from that in *E. coli*, as it seems not to have the sigma factors that upregulate chaperone synthesis in *E. coli* (heat-shock sigma 32 and stationary-phase sigma S).

In addition to the SecA-dependent secretory pathway, *H. pylori* has two specialized export systems. One is associated with the *cag* pathogenicity island[11] and the other is the flagellar export pathway which is assembled from orthologues of FliH, FliI, FliP, FlhA, FlhB, FliQ, FliR and FliP[17]. Apparently absent from *H. pylori* is a type IV signal peptidase and orthologues of the dsbABC system, which in other species are required for the maturation of pili and pilin-like structures[18] and assembly of surface structures involved in virulence and DNA transformation[19].

Recombination, repair and restriction systems

Systems for homologous recombination and post-replication, mismatch, excision and transcription-coupled repair appear to be present in *H. pylori*. Also present are genes with similarity to DNA glycosylases which have associated AP endonuclease activity. The RecBCD pathway, which mediates homologous recombination and double-strand break repair, and RecT and RecE orthologues, proteins involved in strand exchange during recombination[20], seem to be absent. The ability of *H. pylori* to perform mismatch repair is suggested by the presence of methyl transferases, mutS and uvrD. However, orthologues of MutH and MutL were not identified. Components of an SOS system also appear to be absent.

Bacteria commonly use restriction and modification systems to degrade foreign DNA. In *H. pylori*, this defence system is well developed with eleven restriction-modification systems identified on the basis of gene order and similarity to endonucleases, methyltransferases, and specificity subunits. Three type I, one type II, and three type IIS systems were identified, as well as four type III systems, including the recently identified epithelial responsive

Figure 1 Linear representation of the *H. pylori* 26695 chromosome illustrating the location of each predicted protein-coding region, RNA gene, and repeat elements in the genome. Symbols are as follows: ++, Co^{2+}, Zn^{2+}, Cd^{2+}; ?, unknown; A/G/S, D-alanine/glycine/D-serine; B12, B12/ferric siderophores; E, glutamate; Mo, molybdenum; P, proline; P/G, proline/glycine betaine; Q, glutamine; S, serine; a-k, α-ketoglutarate; a/o, arginine/ornithine; aa, amino acids (specificity unknown); aa2, dipeptides; aaX, oligopeptides; fum, fumarate, succinate; glu, glucose/galactose; h, hemin; lac, L-lactate; mal, malate 2-oxoglutarate; nic, nicotinamide mononucleotides; pyr, pyrimidine nucleosides. Numbers associated with tRNA symbols represent the number of tRNAs at a locus. Numbers associated with GES represent the number of membrane-spanning domains according to the Goldman, Engelman and Steitz scale as calculated by TopPred[47].

endonuclease, *iceA1*, and its associated DNA adenine methyltransferase (M. HypI) genes[21,22]. In addition to the complete systems, seven adenine-specific, and four cytosine-specific methyltransferases, and one of unknown specificity were found. Each of these has an adjacent gene with no database match, suggesting that they may function as part of restriction-modification systems.

Transcription and translation

Although analysis of gene content suggests that *H. pylori* has a basic transcriptional and translational machinery similar to that of *E. coli*, interesting differences are observed. For example, no genes for a catalytic activity in tRNA maturation (*rnd*, *rph*, or *rnpB*) were identified and of the three known ribonucleases involved in mRNA degradation, only polyribonucleotide phosphorylase was found. Twenty-one genes coding for 18 of the 20 tRNA synthetases normally required for protein biosynthesis were found.

As in most other completely sequenced bacterial genomes, the gene for glutaminyl-tRNA synthetase, *glnS*, is missing, and the existence of a transamidation process is assumed. It is also possible that the product of the second glutamyl-tRNA synthetase gene, *gltX*, present in *H. pylori*, may have acquired the glutaminyl-tRNA synthetase function. *H. pylori* provides the first example of a bacterial genome apparently lacking an asparaginyl-tRNA synthetase gene, *asnS*. A transamidation process to form *Asn-tRNAAsn* from *Asp-tRNAAsn* has been reported for the archaeon *Haloferax volcanii*[22] and may also operate in *H. pylori*. Most intriguing, however, is the finding that in *H. pylori* the genes encoding the β and β′ subunits of RNA polymerase are fused. In all studied prokaryotes the two genes are contiguous, but separate, and are part of the same transcriptional unit. Whether this gene fusion in *H. pylori* results in a fused protein, or whether the transcriptional or translational product of the fusion is subject to splicing, is currently not known. It is worth noting that an artificial fusion of the *E. coli* *rpoB* and *rpoC* genes is viable and results in a transcriptional complex, which has the same stoichiometry as the native complex (K. Severinov, personal communication).

Adhesion and adaptive antigenic variation

Most pathogens show tropism to specific tissues or cell types and often use several adherence mechanisms for successful attachment. *H. pylori* may use at least five different adhesins to attach to gastric epithelial cells[5]. One of them, HpaA (HP0797), was previously identified as a lipoprotein in the flagellar sheath and outer membrane[5,23]. In addition to the HpaA orthologue, we have identified 19 other lipoproteins. Few have an identifiable function, but some are likely to contribute to the adherence capacity of the organism.

Two adhesins[24–26], one of which mediates attachment to the Lewis^b histo-blood group antigens, belong to the large family of outer membrane proteins (OMP) (Fig. 3) (T. Boren and R. Haas, personal communication). It is conceivable that other members of these closely related proteins also act as adhesins. Given the large number of sequence-related genes encoding putative surface-exposed proteins, the potential exists for recombinational events leading to mosaic organization. This could be the basis for antigenic variation in *H. pylori* and an effective mechanism for host defence evasion, as seen in *M. genitalium*[27].

At least one other mechanism for antigenic variation could operate in *H. pylori*. The DNA sequence at the beginning of eight genes, including five members of the OMP family, contain stretches of CT or AG dinucleotide repeats (Table 3a). In addition, poly(C) or poly(G) tracts occur within the coding sequence of nine other genes (Table 3b). Slipped-strand mispairing within such repeats are documented features of one mechanism of genotypic variation[28,29]. These mechanisms may have evolved in bacterial pathogens to increase the frequency of phenotypic variation in genes involved in

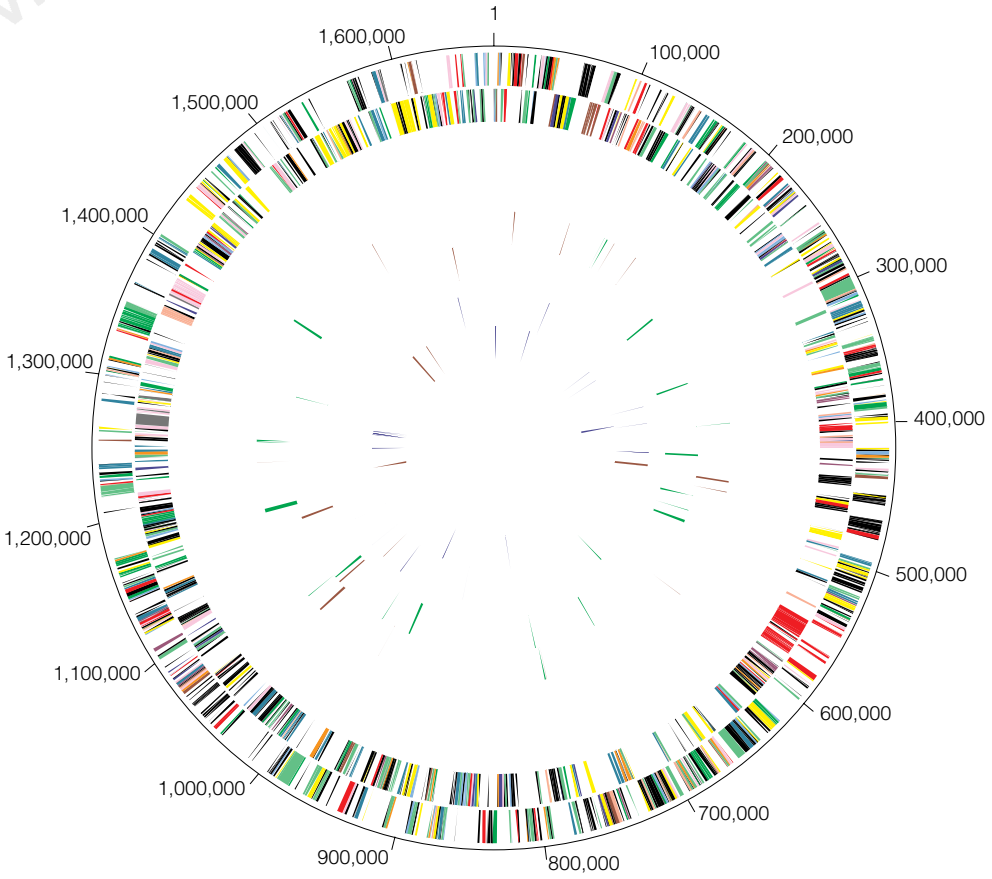

Figure 2 Circular representation of the *H. pylori* 26695 chromosome. Outer concentric circle: predicted coding regions on the plus strand classified as to role according to the colour code in Fig. 1 (except for unknowns and hypotheticals, which are in black). Second concentric circle: predicted coding regions on the minus strand. Third and fourth concentric circles: IS elements (red) and other repeats (green) on the plus and minus strand, respectively. Fifth and sixth concentric circles: tRNAs (blue), rRNAs (red), and sRNAs (green) on the plus and minus strand, respectively.

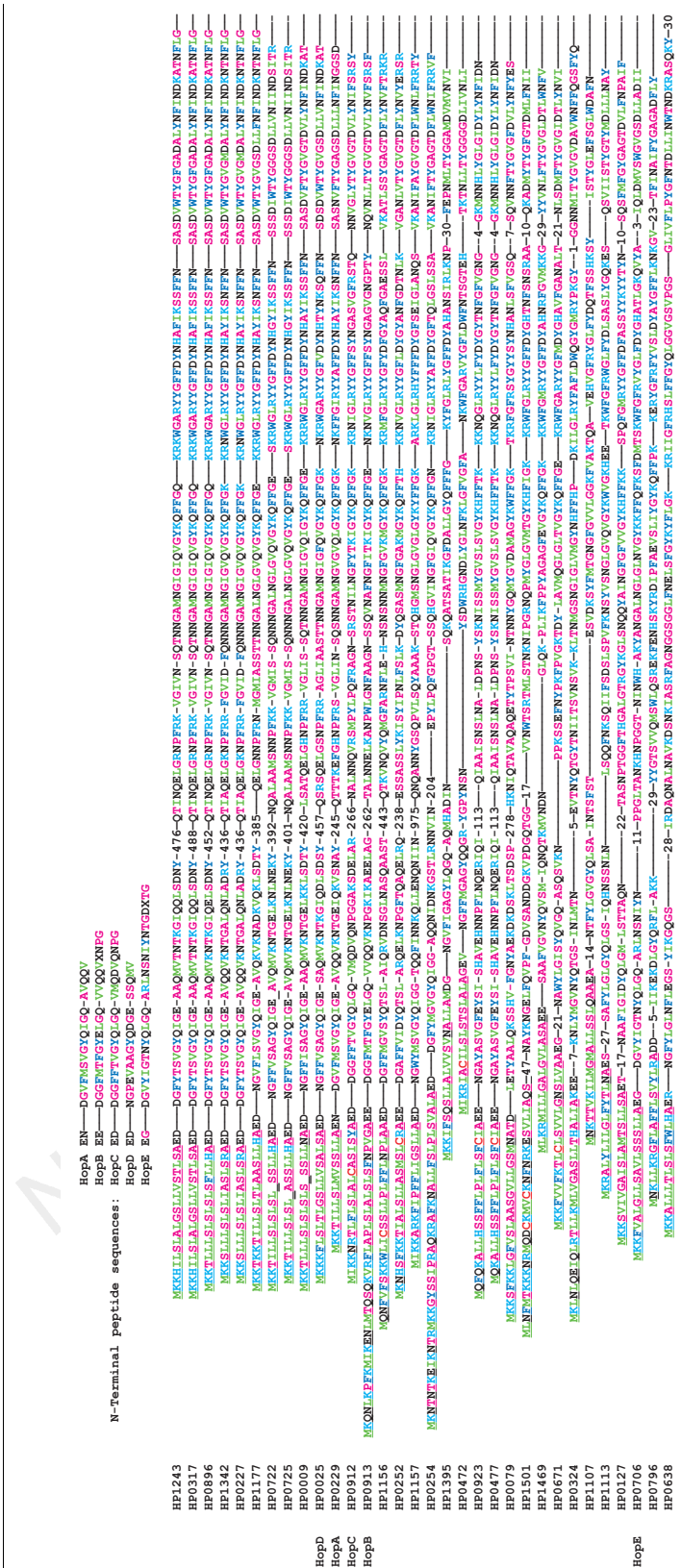

Figure 3 Multiple sequence alignment of members of the outer membrane protein family of *H. pylori*. These proteins were identified as OMPs based on the characteristic alternating hydrophobic residues at their carboxy termini. All members of this family have one domain of similarity at the amino-terminal end and seven domains of similarity at their carboxy-terminal end. Note that the first 11 of these OMPs share extensive similarity over their entire length. Four of the OMPs were identified as porins (Hops) based on identity to published amino-terminal sequences, represented at the top of the alignment[50]. The most likely candidate for HopD is HP0913, which has 15 matches to the first 20-residue N-terminal peptide sequence[50]. These differences may be due to strain variability. The program Signal-P[48] was used to identify cleavage sites and signal peptides (underlined). Four of the OMPs have TTG start codons (HP1156, HP0252, HP1113, HP0796). Numbers embedded in the sequences represent amino acids omitted from the alignment. The star symbols indicate that HP722, HP725 and HP9 proteins contain a frameshift in their signal-peptide-coding region. These frame-shifts are associated with the presence of dinucleotide repeats (Table 3).

critical interactions with their hosts[28]. Such 'contingency' genes encode surface structures like pilins, lipoproteins or enzymes that produce lipopolysaccharide molecules[28]. Our analysis suggests that the seventeen genes reported in Table 3a,b belong to this category and thus may provide an example of adaptive evolution in *H. pylori*.

Phenotypic variation at the transcriptional level may also operate in *H. pylori*. Examples of repetitive DNA mediating transcriptional control have been documented by the presence of oligonucleotide repeats in promoter regions[29]. Homopolymeric tracts of A or T in potential promoter regions of eighteen genes were found, including eight members of the OMP family (Table 3c).

Virulence

The virulence of individual *H. pylori* isolates has been measured by their ability to produce a cytotoxin-associated protein (CagA) and an active vacuolating cytotoxin (VacA)[5]. The *cagA* gene, though not a virulence determinant, is positioned at one end of a pathogenecity island containing genes that elicit the production of interleukin (IL)-8 by gastric epithelial cells[11,30]. Consistent with its more virulent character, *H. pylori* strain 26695 contains a single contiguous PAI region[11] (Fig. 4).

VacA induces the formation of acidic vacuoles in host epithelial cells, and its presence is associated epidemiologically with tissue damage and disease[31]. VacA may not be the only ulcer-causing factor as 40% of *H. pylori* strains do not produce detectable amounts of the cytotoxin *in vitro*[5]. Sequence differences at the amino terminus and central sections are noted among VacA proteins derived from Tox^+ and Tox^- strains[31]. This Tox^+ *H. pylori* strain contains the more toxigenic S1a/m1 type cytotoxin and three additional large proteins with moderate similarities to the carboxy-terminal end of the active

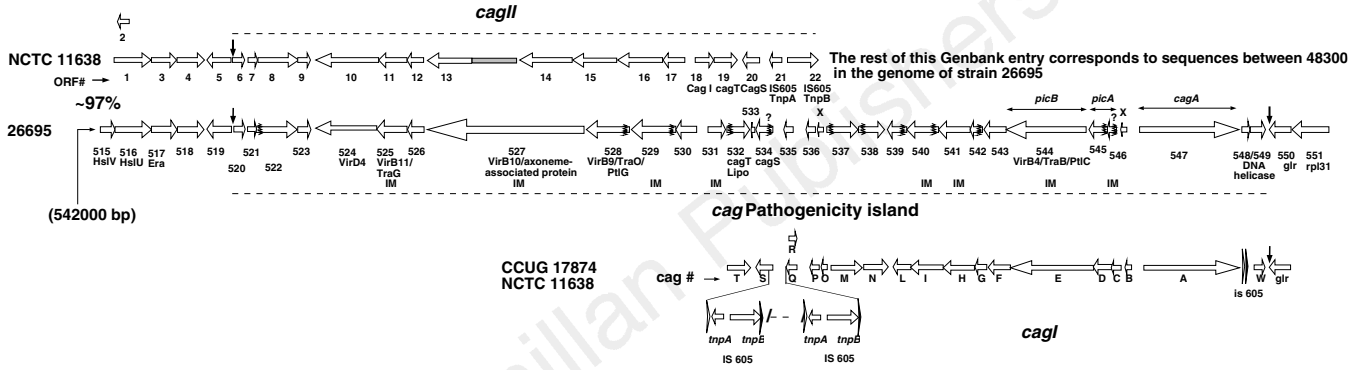

Figure 4 Comparison between the Cag pathogenicity islands of the sequenced strain, 26695 and the NCTC11638 strain. The twenty nine ORFs of the contiguous PAI in strain 26695 are represented together with the corresponding ORFs from the PAI present in NCTC11638 (AC000108 and U60176). The PAI in NCTC11638 is divided by the IS 605 elements into two regions, *cagI* and *cagII*. The PAI in NCTC11638 is flanked by a 31-bp (TTACAATTTGAGCCCATTCTTTAGCTTGTTTT) direct repeat (vertical arrows) as described[11]. Some of the genes encode proteins with similarity to proteins involved either in DNA transfer (Vir and Tra proteins) or in export of a toxin (Ptl protein)[10]. However, these genes do not have the conserved contiguous arrangement found in the VirB, Tra and Ptl operons, suggesting that this PAI is not derived from these systems. Most genes of the PAI have no database match, contrary to a previous suggestion[11]. Thirteen of the proteins have a signal peptide (squiggle line), three of them with a weaker probability (squiggled line+?). The average length of the signal peptides is 25 amino acids, suggesting that this PAI is of Gram-negative origin. Eight proteins are predicted to have at least two membrane-spanning domains and to be integral membrane proteins (IM)[47]. Although the two PAI are ~97% identical at the nucleotide level, there are several notable and perhaps biologically relevant differences between the two sequences. Four of the genes differ in size. In the PAI of strain 26695, HP 520 and 521 are shorter, whereas HP523 is longer, and HP 527 actually spans both ORF13 and 14. In addition, the N-terminal part of HP527 is 129 amino acids longer than the corresponding region in ORF14. HP548/549 contains a frameshift and is therefore probably inactive in strain 26695. The stippled box preceding ORF13 represents an N-terminal extension not annotated in the Genbank entry for the PAI of NCTC11638. The 'x' indicates ORFs that are neither GeneMark-positive nor GeneSmith-positive, so were not included in our gene list. However, these ORFs may be biologically significant. We do not represent cagR as an ORF, because it is completely contained within ORFQ, and is GeneMark-negative.

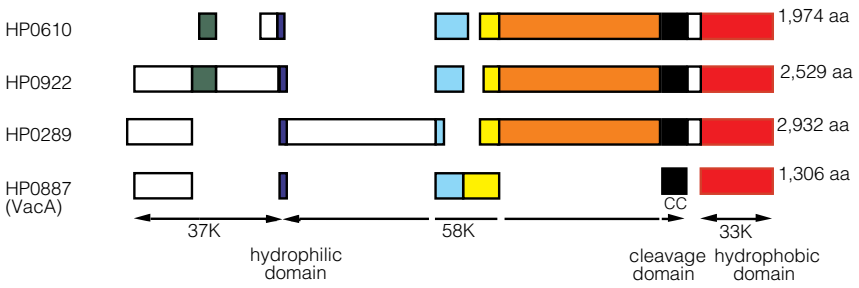

Figure 5 Conserved domains of VacA and related proteins. HP887 is the vacuolating cytotoxin (vacA) gene from *H. pylori* 26695 strain. HP610, HP922 and HP289 are related proteins. Blocks of aligned sequence and the length of each protein are shown. Arrows designate the extents of each VacA domain. The hydrophilic domain (blue boxes) contains the site in VacA at which the N-terminal domain is cleaved into 37K and 58K fragments. The putative cleavage site (ANNNQQNS) differs from that of three cytotoxic strains (CCUG 1784, 60190, G39; AKNDKXES) and is not conserved in the other three VacA-related proteins. The cleavage domain (black boxes) of VacA contains a pair of Cys residues 60 residues upstream from the site at which the C terminus is cleaved. These residues are not conserved in the other three proteins. The 33K C-terminal hydrophobic domain (red boxes) in VacA is thought to form a pore through which the toxin is secreted. The other three proteins show 26–31% sequence similarity to VacA in this region. The other coloured boxes represent regions of similarity.

cytotoxin (~26–31%) (Fig. 5). However, they lack the paired-cysteine residues and the cleavage site required for release of the VacA toxin from the bacterial membrane[31] (Fig. 5). We propose that these proteins may be retained on the outside surface of the cell membrane and contribute to the interaction between *H. pylori* and host cells.

The surface-exposed lipopolysaccharide (LPS) molecule plays an important role in *H. pylori* pathogenesis[32]. The LPS of *H. pylori* is several orders of magnitude less immunogenic than that of enteric bacteria[33] and the O antigen of many *H. pylori* isolates is known to mimic the human Lewis^x and Lewis^y blood group antigen[32]. Genes for synthesis of the lipid A molecule, the core region, and the O antigen were identified. Two genes with low similarity to fucosyltransferases (HP379, HP651) were found and may play a role in the LPS-Lewis antigen molecular mimicry. Our analysis also suggests that three genes, two glycosyltransferases (HP208 and HP619) and one fucosyltransferase (HP379), may be subject to phase variation (Table 3a, b).

As with other pathogens, *H. pylori* probably requires an iron-scavenging system for survival in the host[5]. Genome analysis suggests that *H. pylori* has several systems for iron uptake. One is analogous to the siderophore-mediated iron-uptake *fec* system of *E. coli*[34], except that it lacks the two regulatory proteins (FecR and FecI) and is not organized in a single operon. Unlike other studied systems, *H. pylori* has three copies of each of *fecA*, *exbB* and *exbD*. A second system, consisting of a *feoB*-like gene without *feoA*, suggests that *H. pylori* can assimilate ferrous iron in a fashion similar to the anaerobic feo system of *E. coli*. Other systems for iron uptake present in *H. pylori* consist of the three *frpB* genes which encode proteins similar to either haem- or lactoferrin-binding proteins. Finally, *H. pylori* contains NapA, a bacterioferritin[34], and Pfr, a non-haem cytoplasmic iron-containing ferritin used for storage of iron[35]. The global ferric uptake regulator (Fur) characterized in other bacteria is also present in *H. pylori*. Consensus sequences for Fur-binding boxes were found upstream of two *fecA* genes, the three *frnB* genes and *fur*.

H. pylori motility is essential for colonization[36]. It enables the bacterium to spread into the viscous mucous layer covering the gastric epithelium. At least forty proteins in the *H. pylori* genome appear to be involved in the regulation, secretion and assembly of the flagellar architecture. As has bene reported for the *flaA* and *flaB* genes, we identified sigma 28 and sigma 54-like promoter elements upstream of many flagellar genes, underscoring the complexity of the transcriptional regulation of the flagellar regulon[5].

Acidity, pH and acid tolerance

H. pylori is unusual among pathogenic bacteria in its ability to colonize host cells in an environment of high acidity. As it enters the gastric environment by oral ingestion, the organism is transiently subjected to the extreme pH of the lumen side of the gastric mucous layer (pH ~2). The survival of *H. pylori* in acidic environments is probably due to its ability to establish a positive inside-membrane potential[37] and subsequently to modify its microenvironment through the action of urease and the release of factors that inhibit acid production by parietal cells[5]. A switch in membrane polarity provides an electrical barrier that prevents the entry of protons (H^+). A positive cell interior can be created by the active extrusion of anions or by a proton diffusion potential. The latter model appears more likely as no clear mechanism for electrogenic anion efflux is apparent in the genome. A proton diffusion potential would require the anion permeability of the cytoplasmic membrane to be low and, thus far, only three anion transporters have been identified. However, it remains to be determined whether anion conductances are associated with other proteins: the MDR-like transporters (HP600, HP1082 and HP1206) or hypotheticals. Although it has been suggested that proton-translocating P-type ATPases could mediate survival in acid conditions by the extrusion of protons from the cytoplasm[38], this idea is not supported by the identified transporter

Table 3 Homopolymeric tracts and dinucleotide repeats in *H. pylori*

HP no.	ID	No. of repeats	Gene status	Poly(A) or Poly(T) tracts in 5′ intergenic region
9	OMP	11 CT	Off	Poly(A)
208	glycos. transf.	11 AG	Truncated	Poly(A)
638	OMP	6 CT	On	No
722	OMP	8 CT	Off	Poly(T)
725	OMP	6 CT	Off	Poly(T)
744	Hypo	9 AG	Truncated	No
896	OMP	11 CT	On	Poly(A)
1417	Cons. Hypo	9 AG	Truncated	No

Nucleotide sequence at the beginning of HP0722 showing the CT dinucleotide repeat and the poly T tract. The putative ribosome binding site is shown in green. Translation starting at the designated methionine leads to a truncated product. The addition or deletion of two CT repeats, by 'slipped-strand mispairing', will restore the frame.

CCAAAAATCTTTTTTTTTTTTTTTTGAAATCCAATAAATTTATGGTAAAGT-37bp-TTTACAATAAAAAAATTACTTTAAGGAACATTT

T**ATG**AAAAAGACAATTCTACTCTCTCTCTCTCTCTCGCTTCATCGCTCTTGCACGCTGAAGACAACGGCTTTTTTGTGAGCGCCGGCT

Y E K D N S T L S L S L S L A S S L L H A E D N G F F V S A G Y

M K K T I L L S L S L S L H R S C T L K T T A F L *

(b) Homopolymeric poly(C) and poly(G) tracts within coding sequence

HP no.	ID	Tract length	Gene status
58	Hypo	C15	Off
217	Hypo	G12	On
379	fucosyl transf.	C13	On
464	TypeI R	C15	On
619	glycos. transf.	C13	Truncated
651	Hypo	C13	On
1353	Hypo	C15	Truncated
1471	TypeIIS-R	G14	On
1522	Methyl ase	G12	Truncated

Genes possibly regulated by homopolymeric poly(A) or poly(T) tracts in 5′ intergenic regions

HP no.	ID	Tract	HP no.	ID	Tract	HP no.	ID	Tract
9	OMP	A14	25	OMP	T15	208	*rfaJ*	A11
227	OMP	T14	228	IMP	A14	349	*pyrG*	T15
350	IMP	A15	547	*cagA*	A14	629	Hypo	T15
722	OMP	T16	725	OMP	T14	733	Hypo	T13
876	*frpB*	T16	896	OMP	A14	912	OMP	T13
1342	OMP	A14	1400	*fecA*	A16			

genes. The P-type ATPase sequences in *H. pylori* (*copAP*, HP791, and HP1503) are more closely related to divalent cation transporters than to ATPases with specificity for protons or monovalent cations. One of them, HP0791, is involved in Ni^{2+} supply, an essential component of urease activity[39]. The others may be involved in the elimination of toxic metals from the cytoplasm and not in pH regulation.

Additional mechanisms of pH homeostasis may well contribute to *H. pylori* survival. A change in protein content observed in response to a shift of extracellular pH from 7.5 to 3.0 suggests the presence of an acid-inducible response[40]. Although *H. pylori* lacks most orthologues of the genes that are acid-induced in *E. coli* and *Salmonella typhimurium*, including the amino-acid decarboxylases and formate hydrogen lyase, certain virulence factors, outer membrane

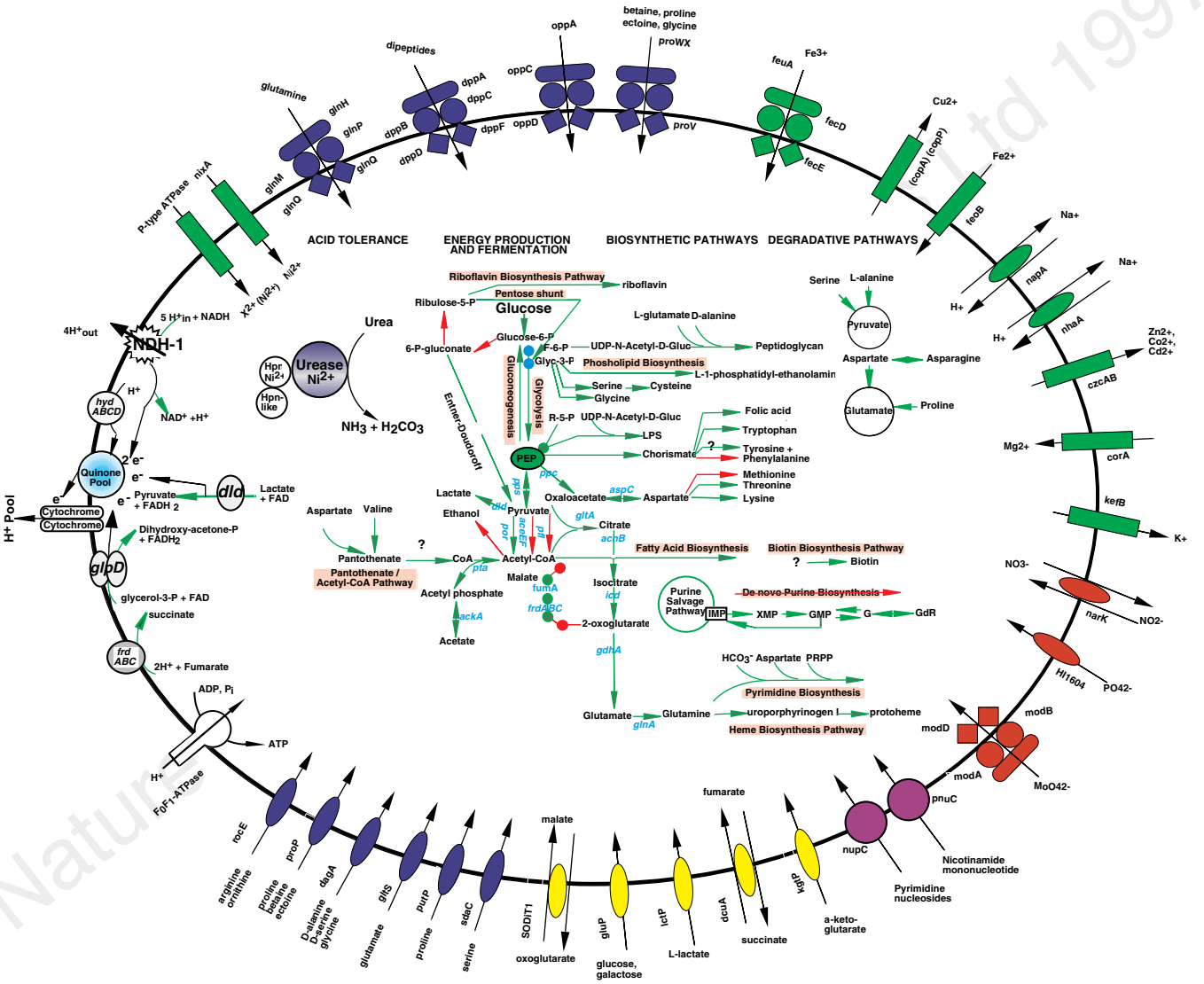

Figure 6 Solute transport and metabolic pathways of *Helicobacter pylori*. Transporters identified by sequence comparisons are characteristc of Gram-negative bacteria. Colours correspond to transport role categories defined by Riley[15]: blue, amino acids, peptides and amines; red, anions; yellow, carbohydrates, organic alcohols and acids; green, cations; and purple, nucleosides, purines and pyrimidines. Numerous permeases (ovals) with specificity for amino acids (*recE*, *proP*, *dagA*, *gltS*, *putP* and *sdaC*) or carbohydrates (*SODiTI*, *gluP*, *lactP*, *cduA*, *kgtP*) import organic nutrients. Structurally related permease proteins maintain ionic homeostasis by transporting HPO_4^{2-} (*Hl1604*), NO_3^{2-} (*narK*), and Na^+ (*nhA*, *napA*). Primary active-transport systems, independent of the proton cycle, are also apparent. Included in this group are ATP-binding protein-cassette (ABC) transporters (composite figures of 2 diamonds, 2 circles, 1 oval) for the uptake of oligopeptides (*oppACD*), dipeptides (*dppABCDF*), proline (*proVWX*), glutamine (*glnHMPQ*), molybdenum (*modABD*), and iron III (*fecED*), P-type ATPases that extrude toxic metals from the cell (*copAP* and *cadA*), and the glutathione-regulated potassium-efflux protein (*kefB*). Transporters for the accumulation of ionic cofactors are encoded by *nixA* (Ni^{2+} for urease activation), *corA* (Mg^{2+} for phosphohydrolases, phosphotransferases, ATPases) and *feoB* (Fe^{2+} import under anaerobic conditions for cytochromes, catalase). An integrated view of the main components of the central metabolism of *H. pylori* strain 26695 is presented. The use of glucose as the sole carbohydrate source is emphasized. Urease, a multisubunit Ni^{2+}-binding enzyme, is crucial for colonization and for survival of *H. pylori* at acid pH, and is indicated as a complex (purple circle) with Hpn, a Ni^{2+}-binding cofactor, and a newly identified Hpn-like protein (HP1432). A question mark is attached to pathways that could not be completely elucidated. Pathways or steps for which no enzymes were identified are represented by a red arrow. Pathways for macromolecular biosynthesis (RNA, DNA and fatty acids) have been omitted. *ackA*, acetate kinase; *acnB*, aconitase B; *aspC*, aspartate aminotransferase; *dld*, D-lactate dehydrogenase; *gdhA*, glutamate dehydrogenase; *glnA*, glutamine synthetase; *gltA* citrate synthase; *HydABC*, hydrogenase complex; *icd*, isocitrate dehydrogenase; *pfl*, pyruvate formate lyase; *por*, pyruvate ferredoxin oxidoreductase; *ppc*, phosphoenolpyruvate carboxylase; *pps*, phosphoenolpyruvate synthase; *pta*, phosphate acetyltransferase; *gldD*, glycerol-3-phosphate dehydrogenase; NDH-1, NADH–ubiquinone oxidoreductase complex.

proteins, sensor-regulator pairs and other proteins may be acid-induced.

Regulation of gene expression

Bacteria regulate the transcription of their genes in response to many environmental stimuli, such as nutrient availability, cell density, pH, contact with target tissue, DNA-damaging agents, temperature and osmolarity. In the case of pathogens, the regulated expression of certain key genes is essential for successful evasion of host responses and colonization, adaptation to different body sites, and survival as the pathogen passes to new hosts. In *H. pylori*, global regulatory proteins are less abundant than in *E. coli*. For example, orthologues of many DNA-binding proteins that regulate the expression of certain operons such as OxyR (oxidative stress), Crp (carbon utilization), RpoH (heat shock), and Fnr (fumarate and nitrate regulation) are absent. Only four *H. pylori* proteins have a perfect match to helix–turn–helix (HTH) motifs, a signature of transcription factors; a putative heat-shock protein (HspR), two proteins with no database match (HP1124 and HP1349) and SecA, a component of the general secretory machinery. In contrast, 34 proteins containing an HTH motif were found in *H. influenzae* and 148 in *E. coli*. We identified several other putative regulatory functions, including SpoT and CstA for 'stringent response' to amino-acid starvation and to carbon starvation, respectively.

Environmental response requires sensing changes and transmission of this information to cellular regulatory networks. Two-component regulator systems, consisting of a membrane histidine kinase sensor protein and a cytoplasmic DNA-binding response regulator, provide a well studied mechanism for such signal transduction. Four sensor proteins and seven response regulators were found in *H. pylori*, similar to the number found in *H. influenzae*[7]. This is approximately one third the number found in *E. coli* which, in contrast to *H. pylori* and *H. influenzae*, may be exposed to more environments.

Metabolism

Metabolic pathway analysis of the *H. pylori* genome suggests the following features. *H. pylori* uses glucose as the only source of carbohydrate and the main source for substrate-level phosphorylation. It also derives energy from the degradation of serine, alanine, aspartate and proline. The glycolysis–gluconeogenesis metabolic axis constitutes the backbone of energy production and the start point of many biosynthetic pathways. The biosynthesis of peptidoglycan, phospholipids, aromatic amino acids, fatty acids and cofactors is derived from acetyl-CoA or from intermediates in the glycolytic pathway (Fig. 6). The metabolism of pyruvate reflects the microaerophilic character of this organism. Neither the aerobic pyruvate dehydrogenase (*aceEF*) nor the strictly anaerobic pyruvate formate lyase (*pfl*) associated with mixed-acid fermentation are present. The conversion of pyruvate to acetyl CoA is performed by the pyruvate ferrodoxin oxidoreductase (POR), a four-subunit enzyme thus far only described in hyperthermophilic organisms[41]. The tricarboxylic acid cycle (TCA) is incomplete and the glyoxylate shunt is absent. The analysis of degradative pathways, uptake systems and biosynthetic pathways for pyrimidine, purine and haem suggests that *H. pylori* uses several substrates as nitrogen source, including urea, ammonia, alanine, serine and glutamine. The assimilation of ammonia, an abundant product of urease activity, is achieved by the glutamine synthase enzyme and α-ketoglutarate is transformed into glutamate by glutamate dehydrogenase rather than by the glutamate synthase enzyme.

In *H. pylori*, proton translocation is mediated by the NDH-1 dehydrogenase and the different cytochromes, including the primitive-type cytochrome cbb3 (Table 2). Four respiratory electron-generating deydrogenases have been identified, glycerol-3-phosphate dehydrogenase (GlpD), D-lactate dehydrogenase, NADH–ubiquinone oxidoreductase complex (NDH-1), and a hydrogenase complex (HydABC). Our analysis also suggests that *H. pylori* is not able to use nitrate, nitrite, dimethylsulphoxide, trimethylamine *N*-oxide or thiosulphate as electron acceptors. Much of our metabolic analysis is supported by experimental evidence[41,42].

Evolutionary relationships of *H. pylori*

H. pylori is currently classified in the Proteobacteria, a large, diverse division of Gram-negative bacteria which includes two other completely sequenced species, *H. influenzae* and *E. coli*. Given this taxonomic placement, based primarily on 16S rRNA sequence comparisons, one might expect the proteins of *H. pylori* more closely to resemble their *H. influenzae* and *E. coli* homologues rather than those in other genomes such as *Synechocystis* sp., *M. genitalium*, *M. pneumoniae*, *M. jannaschii*, and *Saccharomyces cerevisae*. This is indeed the case for many proteins. There are, however, many examples of *H. pylori* proteins in amino-acid biosynthesis, energy metabolism, translation and cellular processes that have greater sequence similarity to those found in non-Proteobacteria. For example, Dhs1, the initial enzyme in the chorismate biosynthesis pathway is 75.5% similar to *Arabidopsis thaliana* chloroplast Dhs1 gene product, and has minimal sequence similarity to the equivalent *E. coli* AroH, AroF or AroG gene products. The remaining enzymes in this pathway have strong sequence similarity to their *E. coli* counterpart. Similarly, the *H. pylori* prephenate dehydrogenase (TyrA), which converts chorismate to tyrosine, and six out of 15 enzymes in the aspartate amino acid biosynthetic pathways, resemble those from *B. subtilis*. A similar pattern can be seen in a different functional category. Nearly all *H. pylori* tRNA synthetases have eubacterial homologues, mostly with best matches to Proteobacteria species. However, histidyl-tRNA synthetase shows several amino-acid sequence signatures in common with eukaryotic and archaeal (*M. jannaschii*) homologues.

Such observations of discordant sequence similarity are often interpreted as evidence of lateral gene transfer in the evolutionary history of an organism. It is also possible that *H. pylori* diverged early from the lineage that led to the gamma Proteobacteria, and retained more ancient forms of enzymes that have been subsequently replaced or have diverged extensively in *H. influenzae* and *E. coli*.

Conclusion

Our whole-genome analysis of *H. pylori* gives new insight into its pathogenesis, acid tolerance, antigenic variation and microaerophilic character. The availability of the complete genome sequence will allow further assessment of *H. pylori* genetic diversity. This is an important aspect of *H. pylori* epidemiology as allelic polymorphism within several loci has already been associated with disease outcome[5,21,31]. The extent of molecular mimicry between *H. pylori* and its human host, an underappreciated topic, can now be fully explored[43]. The identification of many new putative virulence determinants should allow critical tests of their roles and thus new insight into mechanisms of initial colonization, persistence of this bacterium during long-term carriage, and the mechanisms by which it promotes various gastroduodenal diseases.

Methods

H. pylori strain 26695 (ref. 44) was originally isolated from a patient in the United Kingdom with gastritis (K. Eaton, personal communication) and was chosen because it colonizes piglets and elicits immune and inflammatory responses. It is also toxigenic, and transformable, and thus amenable to mutational tests of gene function.

The *H. pylori* genome sequence was obtained by a whole-genome random sequencing method previously applied to genomes of *Haemophilus influenzae*[7], *Mycoplasma genitalium*[8], and *Methanococcus jannaschii*[9]. Ninety-two per cent of the genome was covered by at least one λ clone and only 0.56% of the genome had single-fold coverage.

Open reading frames (ORFs) and predicted coding regions were identified using three methods. The predicted protein-coding regions were initially defined by searching for ORFs longer than 80 codons. Coding potential analysis of the entire genome was performed with a version of GeneMark[45] trained with a set of *H. pylori* ORFs longer than 600 nucleotides. Coding sequences and potential starts of translation were also determined using GeneSmith (H.S., unpublished), a program that evaluates ORF length, separation of ORFs and overlap and quality of ribosome binding site. ORFs with low GeneMark coding potential, no database match, and not retained by GeneSmith were eliminated. GeneSmith identified 25 ORFs that are smaller than 100 codons, had no database match and were GeneMark negative. Frameshifts were detected by inspecting pairwise alignments, families of orthologues (similar proteins derived from different species) and paralogues (similar proteins from within the same organism), and regions containing homopolymer stretches and dinucleotide repeats. Ambiguities were resolved by an alternative sequencing chemistry (terminator reactions), and by sequencing PCR products obtained using the genomic DNA as template. Frameshifts that remain in the genome are considered authentic and not sequencing artefacts.

To determine their identity, ORFs were searched against a non-redundant amino-acid database as previously described[9]. ORFs were also analysed using 175 hidden Markov models constructed for a number of conserved protein families (pfam v1.0) using hmmer[43]. In addition, all ORFs were searched against the prosite motif database using MacPattern[46]. Families of paralogues were constructed by pairwise searches of proteins using FASTA. Matches that spanned at least 60% of the smaller of the protein pair were retained and visually inspected.

A unix version of the program TopPred[47] was used to identify membrane-spanning domains (MSD) in proteins. Six hundred and sixty three proteins containing at least one MSD were found; of these, 300 had 2 potential MSDs or more. The presence of signal peptides and the probable position of the cleavage site in secreted proteins were detected using Signal-P, a neural net program that had been trained on a curated set of secreted proteins from Gram-negative bacteria[48]. 367 proteins were predicted to have a signal peptide. Lipoproteins were identified by scanning for the presence of a lipobox in the first 30 amino acids of every protein; 20 lipoproteins were identified, eighteen of which were Signal-P positive. Outer-membrane proteins were found by searching for aromatic amino acids at the end of the proteins.

Homopolymer and dinucleotide repeats were found by using RepScan (H.O.S., unpublished) which finds direct repeats of any length. All features identified using these programs were validated by visual inspection to remove false positives. Metabolic pathways were curated by hand and by reference to EcoCyc[49].

Received 16 May; accepted 1 July 1997.

1. Warren, J. R. & Marshall, B. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. *Lancet* **1,** 1273–1275 (1983).
2. Cover, T. L. & Blaser, M. J. *Helicobacter pylori* infection, a paradigm for chronic mucosal inflammation: pathogenesis and implications for eradication and prevention. *Adv. Int. Med.* **41,** 85–117 (1996).
3. Mobley, H. L. T., Island, M. D. & Hausinger, R. P. Molecular Biology of Microbial Ureases. *Microbiol. Rev.* **59,** 451–480 (1995).
4. Go, M. F. & Graham, D. Y. How dos *Helicobacter pylori* cause duodenal ulcer disease: The bug, the host, or both? *J. Gastroentrol. Hepatol.* (suppl.) **9,** 8–12 (1994).
5. Labigne, A. & de Reuse, H. Determinants of *Helicobacter pylori* pathogenicity. *Infect. Agents Disease* **5,** 191–202 (1996).
6. Clemens, J. *et al.* Impact of infection by *Helicobacter pylori* on the risk and severity of endemic cholera. *J. Inf. Dis.* **171,** 1653–1656 (1995).
7. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae Rd. Science* **269,** 496–512 (1995).
8. Fraser, C. M. *et al.* The *Mycoplasma genitalium* genome sequence reveals a minimal gene complement. *Science* **270,** 397–403 (1995).
9. Bult, C. J. *et al.* Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii. Science* **273,** 1058–1073 (1996).
10. Winans, S. C., Burns, D. L. & Christie, P. J. Adaptation of a conjugal transfer system for the export of pathogenic macromolecules. *Trends Microbiol.* **4,** 64–68 (1996).
11. Censini, S. *et al.* Cag, a pathogenicity island of *Helicobacter pylori*, encodes typeI-specific and disease-associated virulence factors. *Proc. Natl Acad. Sci. USA* **93,** 14648–14653 (1996).
12. http://genome.wustl.edu/eddy/low/tRNAscan-SE-Manual/Manual.html
13. Akopyants, N. S., Kersulyte, D. & Berg, D. E. DNA rearrangement in the 40 kb cag (virulence) region in the *Helicobacter pylori* genome. *Gut* **39** (suppl. 2), A67 (1996).
14. Marczynski, G. T. & Shapiro, L. Bacterial chromosome origins of replication. *Curr. Opin. Gen. Dev.* **3,** 775–782 (1993).
15. Riley, M. Functions of gene products of *Escherichia coli*. *Microbiol. Rev.* **57,** 862–952 (1993).
16. Kornberg, A. & Baker, T. A. Replication mechanisms and operations in DNA replication. (ed. Kornberg, A. & Baker, T.) 471–510 (Freeman, New York, 1992).
17. Macnab, R. M. in Escherichia coli *and* Salmonella *Cellular and Molecular Biology* (eds Neidhardt, F. C. *et al.*) 123–145 (ASM, Washington DC, 1996).
18. Strom, M. S., Nunn, D. N. & Lory, S. Posttranslational processing of type IV prepilin and homologs by PilD of Pseudomonas aeruginosa. *Meth. Enzymol.* **235,** 527–540 (1994).
19. Bardwell, J. C. Building bridges: disulphide bond formation in the cell. *Mol. Microbiol.* **14,** 199–205 (1994).
20. Linn, S. in Escherichia coli and *Salmonella Cellular and Molecular Biology* (eds Neidhardt, F. C. *et al.*) 764–772 (ASM, Washington D.C., 1996).
21. Peek, R. M., Thompson, S. A., Atherton, J. C., Blaser, M. J. & Miller, G. G. Expression of iceA, a novel ulcer-associated *Helicobacter pylori* gene, is induced by contact with gastric epithelial cells and is associated with enhanced mucosal IL-8. *Gut* **39** (suppl. 2), A71 (1996).
22. Curnow, A. W., Ibba, M. & Soll, D. tRNA-dependent asparagine formation. *Nature* **382,** 589–590 (1996).
23. Jones, A. C., Foynes, S., Cockayne, A. & Penn, C. W. Gene cloning of a flagellar sheath protein of *Helicobacter pylori* shows its identity with the putative adhesin, HpaA. *Gut* **39** (suppl. 2), A62 (1996).
24. Boren, T., Falk, P., Roth, K. A., Larson, G. & Normark, S. Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens. *Science* **262,** 1892–1895 (1993).
25. Ilver, D. *et al.* The *Helicobacter pylori* blood group antigen binding adhesin. *Gut* **39** (suppl. 2), A55 (1996).
26. Odenbreit, S., Till, M. & Haas, R. Optimized blaM-transposon shuttle mutagenesis of *Helicobacter pylori* allows identification of novel genetic loci involved in bacterial virulence. *Mol. Microbiol.* **20,** 361–373 (1996).
27. Peterson, S. N. *et al.* Characterization of repetitive DNA in the *Mycoplasma genitalium* genome: possible role in the generation of antigenic variation. *Proc. Natl Acad. Sci. USA* **92,** 11829–11833 (1995).
28. Moxon, E. R., Rainey, P. B., Nowak, M. A. & Lenski, R. E. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* **4,** 24–33 (1994).
29. Jonsson, A. B., Nyberg, G. & Normark, S. Phase variation of gonococcal pili by frameshift mutation in pilC, a novel gene for pilus assembly. *EMBO J.* **10,** 477–488 (1991).
30. Tummuru, M. K. R., Sharma, S. A. & Blaser, M. J. *Helicobacter pylori* picB, a homologue of the *Bordetella pertussis* toxin secretion protein, is required for induction of IL-8 in gastric epithelial cells. *Mol. Microbiol.* **18,** 867–876 (1995).
31. Atherton, J. C. *et al.* Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. Association of specific vacA types with cytotoxin production and peptic ulceration. *J. Biol. Chem.* **270,** 17771–17777 (1995).
32. Moran, A. P. The role of lipopolysaccharide in *Helicobacter pylori* pathogenesis. *Aliment. Pharmacol. Ther.* **10** (suppl. 1), 39–50 (1996).
33. Baker, P. J. *et al.* Molecular structures that influence the immunomodulatory properties of the lipid A and inner core region oligosaccharides of bacterial lipopolysaccharides. *Infect. Immun.* **62,** 2257–2269 (1994).
34. Earhart, C. F. in Escherichia coli *and* Salmonella *Cellular and Molecular Biology* (eds Neidhardt, F. C. *et al.*) 1075–1090 (ASM, Washington DC, 1996).
35. Evans, D. J. Jr, Evans, D. G., Lampert, H. C. & Nakano, H. Identification of four new prokaryotic bacterioferritins, from *Helicobacter pylori*, *Anabaena variabilis*m, *Bacillus subtilis* and *Treponema pallidum*, by analysis of gene sequences. *Gene* **153,** 123–127 (1995); Frazier, B. A. *et al.* Paracrystalline inclusions of a novel ferritin containing nonheme iron, produced by the human gastric pathogen *Helicobacter pylori*: evidence for a third class of ferritins. *J. Bacteriol.* **175,** 966–972 (1993).
36. Suerbaum, S. The complex flagella of gastric *Helicobacter* species. *Trends Microbiol.* **3,** 168–170 (1995).
37. Matin, A., Zychlinsky, E., Keyhan, M. & Sachs, G. Capacity of *Helicobacter pylori* to generate ionic gradients at low pH is similar to that of bacteria which grow under strongly acidic conditions. *Infect. Immun.* **64,** 1434–1436 (1996).
38. Melchers, K. *et al.* Cloning and membrane topology of a P type ATPase from *Helicobacter pyroli*. *J. Biol. Chem.* **271,** 446–457 (1996).
39. Melchers, K. *et al.* Cloning and analysis of two P type ion pumps of *Helicobacter pylori*, a cation resistance ATPase and a membrane pump necessary for urease activity. *Gut* **39** (suppl. 2), A67 (1996).
40. McGowan, C. C., Cover, T. L. & Blaser, M. J. *Helicobacter pylori* and gastric acid: biological and therapeutic implications. *Gastroenterology* **110,** 926–938 (1996).
41. Hughes, N. J., Chalk, T. L., Clayton, C. L. & Kelly, D. J. Identification of carboxylation enzymes and characterization of a novel four-subunit pyruvate:flavodoxin oxidoreductase from *Helicobacter pylori*. *J. Bacteriol.* **177,** 3953–3959 (1995).
42. Mendz, G. L. & Hazell, S. L. Aminoacid utilizaiton by *Helicobacter pylori*. *Int. J. Biochem. Cell. Biol.* **27,** 1085–1093 (1995).
43. Sonnhammer, E. L. L., Eddy, S. R. & Durbin, R. Pfam: A comprehensive database of protein families based on seed alignments. *Proteins* (in the press).
44. Akopyants, N. S., Eaton, K. A. & Berg, D. E. Adaptive mutation and co-colonization during *Helicobacter pylori* infection of gnotobiotic piglets. *Infect. Immun.* **63,** 116–121 (1995).
45. Borodovsky, M., Rudd, K. E. & Koonin, E. V. Intrinsic and extrinsic approaches for detecting genes in a bacterial genome. *Nucleic Acids Res.* **22,** 4756–4767 (1994).
46. Fuchs, R. MacPattern: protein pattern searching on the Apple MacIntosh. *Comput. Appl. Biosci.* **7,** 105–106 (1991).
47. Claros, M. G. & von Heijne, G. TopPred II: an improved software for membrane protein structure predictions. *Comput. Appl. Biosci.* **10,** 685–686 (1994).
48. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10,** 1–6 (1997).
49. Karp, P. D., Riley, M., Paley, S. M., Pellegrini-Toole, A. & Krummenacker, M. EcoCyc: Encyclopedia of *Escherichia coli* genes and metabolism. *Nucleic Acids Res.* **25,** 43–51 (1997).
50. Doig, P., Exner, M. M., Hancock, R. E. & Trust, T. J. Isolation and characterization of a conserved porin protein from *Helicobacter pylori*. *J. Bacteriol.* **177,** 5447–5452 (1995).

Acknowledgements. D.E.B., M.B. and W.H. are supported by grants from the NIH; P.K. is supported by a grant from the National Center for Research Resources. We thank N. S. Akopyants for preparing high quality chromosomal DNA from *H. pylori* strain 26695; M. Heaney, J. Scott, A. Saeed and R. Shirley for software and database support; and V. Sapiro, B. Vincent, J. Meehan and D. Mass for computer system support.

Correspondence and requests for materials should be addressed to J.-F.T. (e-mail: ghp@tigr.org). The annotated genome sequence and gene family alignments are available on the World-Wide Web site at http://www.tigr.org/tdb/mdb/hpdb/hpdb.html. The sequence has been deposited with GenBank under accession number AE000511.

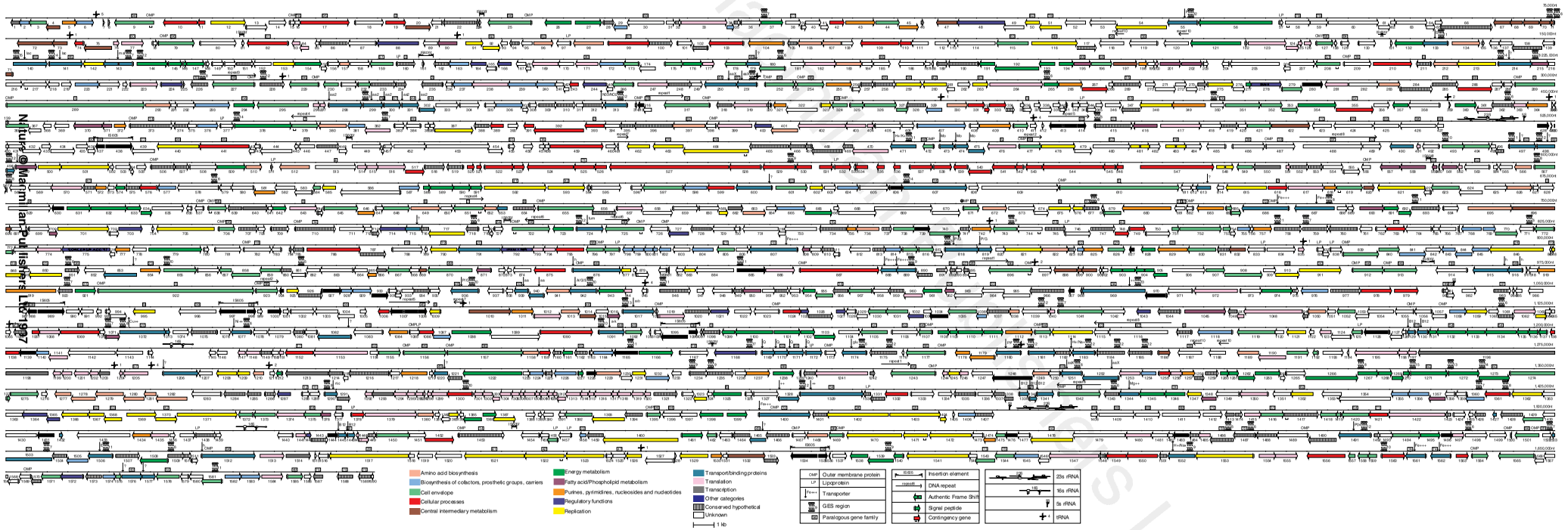

Table 2. List of *H. pylori* genes with putative identifications. Gene numbers correspond to those in Fig. 1. Each identified gene has been assigned a putative role category adapted from ref. 15. Percentages represent per cent identities.

AMINO-ACID BIOSYNTHESIS		
General		
HP0695	hydantoin utilization protein A (hyuA)	28.6%
Aromatic amino-acid family		
HP1038	3-dehydroquinase type II (aroQ)	99.4%
HP0283	3-dehydroquinate synthase (aroB)	38.1%
HP0134	3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (dhs1)	54.6%
HP0401	3-phosphoshikimate 1-carboxyvinyltransferase (aroA)	53.6%
HP1279	anthranilate isomerase (trpC)	47.0%
HP1282	anthranilate synthase component I (trpE)	47.9%
HP1280	anthranilate synthase component II (trpD)	42.5%
HP1281	anthranilate synthase component II (trpD)	40.2%
HP0663	chorismate synthase (aroC)	47.2%
HP1380	prephenate dehydrogenase (tyrA)	30.2%
HP1249	shikimate 5-dehydrogenase (aroE)	36.6%
HP0157	shikimic acid kinase I (aroK)	36.1%
HP1277	tryptophan synthase, alpha subunit (trpA)	46.5%
HP1278	tryptophan synthase, beta subunit (trpB)	66.1%
Aspartate family		
HP0649	aspartate ammonia-lyase (aspA)	55.5%
HP1189	aspartate-semialdehyde dehydrogenase (asd)	45.7%
HP1229	aspartokinase (lysC) 48.0%	
HP0106	cystathionine gamma-synthase (metB)	47.7%
HP0290	diaminopimelate decarboxylase (dap decarboxylase) (lysA)	42.7%
HP0566	diaminopimelate epimerase (dapF)	30.0%
HP0510	dihydrodipicolinate reductase (dapB)	95.3%
HP1013	dihydrodipicolinate synthetase (dapA)	39.5%
HP0822	homoserine dehydrogenase (metL)	37.7%
HP1050	homoserine kinase (thrB)	27.7%
HP0672	solute-binding signature and mitochondrial signature protein (aspB)	47.3%
HP0212	succinyl-diaminopimelate desuccinylase (dapE)	42.3%
HP0626	tetrahydrodipicolinate N-succinyltransferase (dapD)	36.1%
HP0098	threonine synthase (thrC)	32.9%
Glutamate family		
HP0380	glutamate dehydrogenase (gdhA)	59.0%
HP0512	glutamine synthetase (glnA)	48.6%
HP1158	pyrroline-5-carboxylate reductase (proC)	28.9%
Pyruvate family		
HP0941	alanine racemase, biosynthetic (alr)	32.4%
HP1468	branched-chain-amino-acid aminotransferase (ilvE)	63.5%
HP0330	ketol-acid reductoisomerase (ilvC)	48.1%
Serine family		
HP0107	cysteine synthetase (cysK)	45.7%
HP0096	phosphoglycerate dehydrogenase	31.0%
HP0397	phosphoglycerate dehydrogenase (serA)	32.5%
HP0736	phosphoserine aminotransferase (serC)	30.7%
HP0652	phosphoserine phosphatase (serB)	36.5%
HP1210	serine acetyltransferase (cysE)	98.2%
HP0183	serine hydroxymethyltransferase (glyA)	54.0%
BIOSYNTHESIS OF COFACTORS, PROSTHETIC GROUPS, AND CARRIERS		
General		
HP0220	synthesis of [Fe-S] cluster (nifS)	48.0%
Biotin		
HP0598	8-amino-7-oxononanoate synthase (bioF)	34.9%
HP0976	adenosylmethionine-8-amino-7-oxononanoate aminotransferase (bioA)	49.2%
HP1140	biotin operon repressor/biotin acetyl coenzyme A carboxylase synthetase (birA)	36.9%
HP0407	biotin sulfoxide reductase (bisC)	42.7%
HP1254	biotin synthesis protein (bioC)	32.1%
HP1406	biotin synthetase (bioB)	36.2%
HP0029	dethiobiotin synthetase (bioD)	36.0%
Folic acid		
HP1036	7, 8-dihydro-6-hydroxymethylpterin-pyrophosphokinase (folK)	34.6%
HP0587	aminodeoxychorismate lyase (pabC)	32.4%
HP1232	dihydropteroate synthase (folP)	34.5%
HP1545	folylpolyglutamate synthase (folC)	35.2%
HP0928	GTP cyclohydrolase I (folE)	50.9%
HP0577	methylene-tetrahydrofolate dehydrogenase (folD)	48.4%
HP0293	para-aminobenzoate synthetase (pabB)	35.1%
Haem and porphyrin		
HP0163	delta-aminolevulinic acid dehydratase (hemB)	50.5%
HP0376	ferrochelatase (hemH)	33.4%
HP0306	glutamate-1-semialdehyde 2,1-aminomutase (hemL)	51.3%
HP0239	glutamyl-tRNA reductase (hemA)	32.7%
HP0665	oxygen-independent coproporphyrinogen III oxidase (hemN)	42.4%
HP1226	oxygen-independent coproporphyrinogen III oxidase (hemN)	37.9%
HP0237	porphobilinogen deaminase (hemC)	45.7%
HP0381	protoporphyrinogen oxidase (hemK)	35.9%
HP0604	uroporphyrinogen decarboxylase (hemE)	46.3%
HP1224	uroporphyrinogen III cosynthase (hemD)	27.6%
Menaquinone and ubiquinone		
HP1360	4-hydroxybenzoate octaprenyltransferase (ubiA)	26.6%
HP0929	geranyltranstransferase (ispA)	39.8%
HP0240	octaprenyl-diphosphate synthase (ispB)	31.6%
Molybdopterin		
HP0768	molybdenum cofactor biosynthesis protein A (moaA)	31.4%
HP0798	molybdenum cofactor biosynthesis protein C (moaC)	97.9%
HP0172	molybdopterin biosynthesis protein (moeA)	36.3%
HP0755	molybdopterin biosynthesis protein (moeB)	32.2%
HP0799	molybdopterin biosynthesis protein (mog)	50.8%
HP0801	molybdopterin converting factor, subunit 1 (moaD)	31.1%
HP0800	molybdopterin converting factor, subunit 2 (moaE)	31.1%
HP0769	molybdopterin-guanine dinucleotide biosynthesis protein A (mobA)	28.3%
Pantothenate		
HP1058	3-methyl-2-oxobutanoate hydroxymethyltransferase (panB)	43.7%
HP0034	aspartate 1-decarboxylase (panD)	50.0%
HP0006	pantoate-beta-alanine ligase (panC)	44.2%
HP0841	pantothenate metabolism flavoprotein (dfp)	31.3%
Pyridoxine		
HP1583	pyridoxal phosphate biosynthetic protein A (pdxA)	34.2%
HP1582	pyridoxal phosphate biosynthetic protein J (pdxJ)	42.6%
Riboflavin		
HP0802	GTP cyclohydrolase II (ribA)	47.2%
HP0804	GTP cyclohydrolase II/3,4-dihydroxy-2-butanone 4-phosphate synthase (ribA, ribB)	44.0%
HP1505	riboflavin biosynthesis protein (ribG)	33.1%
HP1087	riboflavin biosynthesis regulatory protein (ribC)	28.9%
HP1574	riboflavin synthase alpha subunit (ribC)	32.8%
HP0002	riboflavin synthase beta chain (ribE)	52.4%
Thioredoxin, glutaredoxin and glutathione		
HP1118	gamma-glutamyltranspeptidase (ggt)	53.2%
HP1458	thioredoxin	38.3%
HP0824	thioredoxin (trxA)	51.5%
HP1164	thioredoxin reductase (trxB)	28.5%
Thiamine		
HP0814	thiamin biosynthesis protein (thiF)	34.6%
HP0843	thiamin phosphate pyrophosphorylase/hyroxyethylthiazole kinase (thiB)	35.7%
HP0845	thiamin phosphate pyrophosphorylase/hyroxyethylthiazole kinase (thiM)	37.9%
HP0844	thiamine biosynthesis protein (thi)	41.0%
Pyridine nucleotides		
HP0329	NH(3)-dependent NAD+ synthetase (nadE)	37.5%
HP1355	nicotinate-nucleotide pyrophosphorylase (nadC)	36.3%
HP1356	quinolinate synthetase A (nadA)	34.2%
CELL ENVELOPE		
Membranes, lipoproteins and porins		
HP1450	60 kDa inner-membrane protein	40.0%
HP0180	apolipoprotein N-acyltransferase (cute)	28.0%
HP0175	cell binding factor 2	34.9%
HP0078	Hypothetical protein	28.4%
HP0567	membrane protein	26.4%
HP1456	membrane-associated lipoprotein (lpp20)	98.9%
HP1564	outer membrane protein	39.9%
HP0009	outer membrane protein (omp1)	0.0%
HP0324	outer membrane protein (omp10)	0.0%
HP0472	outer membrane protein (omp11)	99.5%
HP0477	outer membrane protein (omp12)	0.0%
HP0638	outer membrane protein (omp13)	0.0%
HP0671	outer membrane protein (omp14)	36.0%
HP0706	outer membrane protein (omp15)	33.5%
HP0722	outer membrane protein (omp16)	43.3%
HP0725	outer membrane protein (omp17)	43.3%
HP0796	outer membrane protein (omp18)	0.0%
HP0896	outer membrane protein (omp19)	36.6%
HP0025	outer membrane protein (omp2)	0.0%
HP0912	outer membrane protein (omp20)	0.0%
HP0913	outer membrane protein (omp21)	38.2%
HP0923	outer membrane protein (omp22)	0.0%
HP1107	outer membrane protein (omp23)	0.0%
HP1113	outer membrane protein (omp24)	36.0%
HP1156	outer membrane protein (omp25)	0.0%
HP1157	outer membrane protein (omp26)	23.0%
HP1177	outer membrane protein (omp27)	37.0%
HP1243	outer membrane protein (omp28)	0.0%
HP1342	outer membrane protein (omp29)	0.0%
HP0079	outer membrane protein (omp3)	0.0%
HP1395	outer membrane protein (omp30)	0.0%
HP1469	outer membrane protein (omp31)	0.0%
HP1501	outer membrane protein (omp32)	0.0%
HP0127	outer membrane protein (omp4)	0.0%
HP0227	outer membrane protein (omp5)	36.8%
HP0229	outer membrane protein (omp6)	38.4%
HP0252	outer membrane protein (omp7)	30.6%
HP0254	outer membrane protein (omp8)	37.6%
HP0317	outer membrane protein (omp9)	36.3%
HP0839	outer membrane protein P1 (ompP1)	23.3%
HP0955	prolipoprotein diacylglyceryl transferase (lgt)	34.4%
HP0655	protective surface antigen D15	27.5%
HP1571	rare lipoprotein A (rlpA)	37.6%
HP0610	toxin-like outer membrane protein	26.3%
HP0922	toxin-like outer membrane protein	29.5%
HP0289	toxin-like outer membrane protein	30.6%
Murein sacculus and peptidoglycan		
HP0830	amidase	40.6%
HP0738	D-alanine:D-alanine ligase A (ddlA)	28.5%
HP0549	glutamate racemase (glr)	36.6%
HP0772	N-acetylmuramoyl-L-alanine amidase (amiA)	26.8%
HP0597	penicillin-binding protein 1A (PBP-1A)	33.7%
HP1565	penicillin-binding protein 2 (pbp2)	35.0%
HP1125	peptidoglycan associated lipoprotein precursor (omp18)	42.6%
HP0493	phospho-N-acetylmuramoyl-pentapeptide-transferase (mraY)	45.2%
HP0743	rod shape-determining protein (mreB)	37.7%
HP1373	rod shape-determining protein (mreB)	51.9%
HP1372	rod shape-determining protein (mreC)	33.6%
HP0645	soluble lytic murein transglycosylase (slt)	32.2%
HP1543	toxR-activated gene (tagE)	37.2%
HP1544	toxR-activated gene (tagE)	31.2%
HP1155	transferase, peptidoglycan synthesis (murG)	28.2%
HP0740	UDP-MurNac-pentapeptide presynthetase (murF)	25.7%
HP1494	UDP-MurNac-tripeptide synthetase (murE)	36.0%
HP1418	UDP-N-acetylenolpyruvoylglucosamine reductase (murB)	32.7%
HP0648	UDP-N-acetylglucosamine enolpyruvyl transferase (murZ)	46.7%
HP0623	UDP-N-acetylmuramate-alanine ligase (murC)	37.3%
HP0494	UDP-N-acetylmuramoylalanine-D-glutamate ligase (murD)	31.1%
Surface polysaccharides, lipopolysaccharides and antigens		
HP0003	3-deoxy-d-manno-octulosonic acid 8-phosphate synthetase (kdsA)	53.4%
HP0957	3-deoxy-d-manno-octulosonic-acid transferase (kdtA)	35.9%
HP0858	ADP-heptose synthase (rfaE)	40.6%
HP1191	ADP-heptose-lps heptosyltransferase II (rfaF)	33.2%
HP0859	ADP-L-glycero-D-mannoheptose-6-epimerase (rfaD)	32.7%
HP0855	alginate O-acetylation protein (algI)	41.8%
HP0326	CMP-N-acetylneuraminic acid synthetase (neuA)	31.9%
HP0230	CTP:CMP-3-deoxy-D-manno-octulosonate-cytidylyl-transferase (kdsB)	36.2%
HP1392	fibronectin/fibrinogen-binding protein	25.7%
HP0379	fucosyltransferase	39.2%
HP0651	fucosyltransferase	39.2%
HP0044	GDP-D-mannose dehydratase (rfbD)	62.1%
HP0867	lipid A disaccharide synthetase (lpxB)	32.0%
HP0159	lipopolysaccharide 1,2-glucosyltransferase (rfaJ)	28.9%
HP0208	lipopolysaccharide 1,2-glucosyltransferase (rfaJ)	26.7%
HP0805	lipooligosaccharide 5G8 epitope biosynthesis-associated protein (lex2B)	36.9%
HP0826	lipooligosaccharide 5G8 epitope biosynthesis-associated protein (lex2B)	39.2%
HP1416	lipopolysaccharide 1,2-glucosyltransferase (rfaJ)	29.2%
HP0679	lipopolysaccharide biosynthesis protein (wbpB)	42.8%
HP1475	lipopolysaccharide core biosynthesis protein (kdtB)	49.0%
HP0279	lipopolysaccharide heptosyltransferase-1 (rfaC)	31.7%
HP0619	lipopolysaccharide biosynthesis glycosyl transferase (lic2B)	37.2%
HP1105	LPS biosynthesis protein	28.7%
HP1578	LPS biosynthesis protein	28.1%
HP1581	methicillin resistance protein (llm)	29.2%
HP0857	phosphoheptose isomerase (gmhA)	44.5%
HP1275	phosphomannomutase (algC) {Pseudomonas aeruginosa}	39.6%
HP1429	polysialic acid capsule expression protein (kpsF)	46.0%
HP0366	spore coat polysaccharide biosynthesis protein C	35.3%
HP0178	spore coat polysaccharide biosynthesis protein E	36.2%
HP0421	type 1 capsular polysaccharide biosynthesis protein J (capJ)	29.0%
HP0196	UDP-3-0-(3-hydroxymyristoyl) glucosamine N-acyltransferase (lpxD)	39.5%
HP1052	UDP-3-0-acyl N-acetylglcosamine deacetylase (envA)	44.6%
HP1375	UDP-N-acetylglucosamine acyltransferase (lpxA)	41.8%
Surface structures		
HP0840	flaA1 protein	60.2%
HP0325	flagellar basal-body L-ring protein (flgH)	32.7%
HP0351	flagellar basal-body M-ring protein (fliF)	34.4%
HP0246	flagellar basal-body P-ring protein (flgI)	37.9%
HP1557	flagellar basal-body protein (fliE)	37.0%
HP1559	flagellar basal-body rod protein (flgB) (proximal rod protein)	31.0%
HP1558	flagellar basal-body rod protein (flgC) (proximal rod protein)	46.0%
HP1092	flagellar basal-body rod protein (flgG)	35.5%
HP1585	flagellar basal-body rod protein (flgG)	47.7%
HP1041	flagellar biosynthesis protein (flhA)	43.1%
HP1035	flagellar biosynthesis protein (flhF)	35.5%
HP0684	flagellar biosynthesis protein (fliP)	43.4%
HP0770	flagellar biosynthetic protein (flhB)	38.7%
HP0685	flagellar biosynthetic protein (fliP)	55.6%
HP1419	flagellar biosynthetic protein (fliQ)	52.3%
HP0173	flagellar biosynthetic protein (fliR)	26.4%
HP0353	flagellar export protein (fliH)	29.1%
HP1420	flagellar export protein ATP synthase (fliI)	47.6%
HP0870	flagellar hook (flgE)	98.9%
HP0908	flagellar hook (flgE)	30.5%
HP1119	flagellar hook-associated protein 1 (HAP1) (flgK)	27.6%
HP0752	flagellar hook-associated protein 2 (fliD)	28.9%
HP0815	flagellar motor rotation protein (motA)	32.9%
HP0816	flagellar motor rotation protein (motB)	29.7%
HP0352	flagellar motor switch protein (fliG)	37.0%
HP1031	flagellar motor switch protein (fliM)	34.4%
HP0753	flagellar protein (fliS)	32.3%
HP0327	flagellar protein G (flaG)	23.3%
HP0797	flagellar sheath adhesin hpaA	98.5%
HP0584	flagellar switch protein (fliN)	39.7%
HP0601	flagellin A (flaA)	99.8%
HP0115	flagellin B (flaB)	99.0%
HP0295	flagellin B homologue (fla)	32.9%
HP1575	flhB protein (flhB)	40.5%
HP1030	fliY protein (fliY)	29.3%
HP0907	Hook assembly protein, flagella (flgD)	25.5%
HP1274	paralysed flagella protein (pflA)	23.9%
HP0751	polar flagellin (flaG)	21.9%
HP0410	putative neuraminyllactose-binding haemagglutinin homologue (hpaA)	24.2%
HP1192	secreted protein involved in flagellar motility	72.5%
HP1462	secreted protein involved in flagellar motility	96.2%
HP0232	secreted protein involved in flagellar motility	99.2%
CELLULAR PROCESSES		
General		
HP0019	chemotaxis protein (cheV)	26.8%
HP0393	chemotaxis protein (cheV)	31.7%
HP0616	chemotaxis protein (cheV)	27.9%
HP1067	chemotaxis protein (cheY)	99.2%
HP0517	GTP-binding protein (era)	95.6%
HP1490	haemolysin	39.2%
HP1086	Haemolysin (tly)	40.2%
HP0599	haemolysin secretion protein precursor (hylB)	45.4%
HP0392	histidine kinase (cheA)	41.4%
HP0099	methyl-accepting chemotaxis protein (tlpA)	32.8%
HP0103	methyl-accepting chemotaxis protein (tlpB)	30.7%
HP0082	methyl-accepting chemotaxis transducer (tlpC)	28.2%
HP0391	purine-binding chemotaxis protein (cheW)	34.3%
Cell division		
HP0331	cell division inhibitor (minD)	50.2%
HP0749	cell division membrane protein (ftsX)	25.7%
HP0978	cell division protein (ftsA) protein	31.9%
HP0748	cell division protein (ftsE)	37.6%
HP0286	cell division protein (ftsH)	41.2%
HP1069	cell division protein (ftsH)	98.6%
HP1556	cell division protein (ftsI)	30.6%
HP1090	cell division protein (ftsK)	39.8%
HP1560	cell division protein (ftsW) Escherichia coli	32.7%
HP0763	cell division protein (ftsY)	46.6%

HP0332	cell division topological specificity factor (minE)	33.8%
HP0979	cell divison protein (ftsZ)	43.3%
HP1159	cell filamentation protein (fic)	63.2%
Cell killing		
HP0887	vacuolating cytotoxin 94.7%	
Chaperones		
HP0010	chaperone and heat shock protein (groEL)	99.6%
HP0109	chaperone and heat shock protein 70 (dnaK)	63.4%
HP0210	chaperone and heat shock protein C62.5 (htpG)	46.5%
HP0011	co-chaperone (groES)	99.2%
HP1332	co-chaperone and heat-shock protein (dnaJ)	42.7%
HP0110	co-chaperone and heat-shock protein (grpE)	33.0%
HP1024	co-chaperone-curved DNA-binding protein A (CbpA)	37.7%
Chromosome-associated protein		
HP1138	plasmid replication-partition related protein	40.4%
Detoxification		
HP1563	alkyl hydroperoxide reductase (tsaA)	98.5%
HP0875	catalase	99.4%
HP0267	chlorohydrolase	42.6%
HP0243	neutrophil activating protein (napA) (bacterioferritin)	95.8%
HP0389	superoxide dismutase (sodB)	98.6%
HP1452	thiophene and furan oxidizer (tdhF)	37.6%
Protein and peptide secretion		
HP0355	GTP-binding membrane protein (lepA)	57.3%
HP0074	lipoprotein signal peptidase (lspA)	97.0%
HP0786	preprotein translocase subunit (secA)	54.0%
HP1300	preprotein translocase subunit (secY)	41.2%
HP1255	protein translocation protein, low temperature (secG)	30.6%
HP1550	protein-export membrane protein (secD)	38.9%
HP1549	protein-export membrane protein (secF)	35.1%
HP0576	signal peptidase I (lepB)	40.3%
HP1152	signal recognition particle protein (ffh)	41.4%
HP0795	trigger factor (tig)	27.6%
Transformation		
HP0520	cag pathogenicity island protein (cag1)	96.5%
HP0530	cag pathogenicity island protein (cag10)	98.4%
HP0531	cag pathogenicity island protein (cag11)	97.2%
HP0532	cag pathogenicity island protein (cag12)	98.9%
HP0534	cag pathogenicity island protein (cag13)	98.0%
HP0535	cag pathogenicity island protein (cag14)	97.6%
HP0536	cag pathogenicity island protein (cag15)	96.4%
HP0537	cag pathogenicity island protein (cag16)	98.9%
HP0538	cag pathogenicity island protein (cag17)	95.3%
HP0539	cag pathogenicity island protein (cag18)	98.7%
HP0540	cag pathogenicity island protein (cag19)	99.5%
HP0521	cag pathogenicity island protein (cag2)	92.5%
HP0541	cag pathogenicity island protein (cag20)	97.8%
HP0542	cag pathogenicity island protein (cag21)	97.9%
HP0543	cag pathogenicity island protein (cag22)	95.5%
HP0544	cag pathogenicity island protein (cag23)	99.0%
HP0545	cag pathogenicity island protein (cag24)	98.5%
HP0546	cag pathogenicity island protein (cag25)	95.7%
HP0547	cag pathogenicity island protein (cag26)	92.9%
HP0522	cag pathogenicity island protein (cag3)	98.1%
HP0523	cag pathogenicity island protein (cag4)	95.7%
HP0524	cag pathogenicity island protein (cag5)	99.1%
HP0526	cag pathogenicity island protein (cag6)	97.5%
HP0527	cag pathogenicity island protein (cag7)	94.6%
HP0528	cag pathogenicity island protein (cag8)	99.0%
HP0529	cag pathogenicity island protein (cag9)	98.9%
HP1378	competence lipoprotein (comL)	25.5%
HP1361	competence locus E (comE3)	26.7%
HP1006	conjugal transfer protein (traG)	27.3%
HP1421	conjugative transfer regulon protein (trbB)	30.7%
HP0333	DNA processing chain A (dprA)	32.9%
HP0042	trbI protein	31.4%
HP0525	virB11 homologue	100.0%
HP0441	VirB4 homologue	23.5%
HP0017	virB4 homologue (virB4)	25.2%
HP0459	virB4 homologue (virB4)	25.3%

CENTRAL INTERMEDIARY METABOLISM

General		
HP1014	7-α-hydroxysteroid dehydrogenase (hdhA)	33.2%
HP1186	carbonic anhydrase	37.0%
HP0004	carbonic anhydrase (icfA)	33.3%
HP0869	hydrogenase expression/formation protein (hypA)	28.1%
HP0900	hydrogenase expression/formation protein (hypB)	41.4%
HP0899	hydrogenase expression/formation protein (hypC)	38.5%
HP0898	hydrogenase expression/formation protein (hypD)	47.8%
HP0047	hydrogenase expression/formation protein (hypE)	39.7%
HP0197	S-adenosylmethionine synthetase 2 (metX)	62.1%
Amino sugars		
HP1532	glucosamine fructose-6-phosphate aminotransferase (isomerizing) (glmS)	41.7%
Phosphorus compounds		
HP0620	inorganic pyrophosphatase (ppa)	50.0%
HP0696	N-methylhydantoinase	26.9%
HP1010	polyphosphate kinase (ppk)	38.5%
Polyamine biosynthesis		
HP0422	arginine decarboxylase (speA)	33.3%
HP0020	carboxynorspermidine decarboxylase (nspC)	45.6%
HP0832	spermidine synthase (speE)	26.5%
Other		
HP0070	urease accessory protein (ureE)	97.1%
HP0069	urease accessory protein (ureF)	94.5%
HP0068	urease accessory protein (ureG)	95.0%
HP0067	urease accessory protein (ureH)	96.2%
HP0071	urease accessory protein (ureI)	98.5%
HP0073	urease alpha subunit (ureA) (urea amidohydrolase)	100.0%
HP0072	urease beta subunit (urea amidohydrolase) (ureB)	100.0%
HP0075	urease protein (ureC)	98.0%

ENERGY METABOLISM

Aerobic		
HP1222	D-lactate dehydrogenase (dld)	27.0%
HP0961	glycerol-3-phosphate dehydrogenase (NAD(P)+)	36.8%
HP0037	NADH-ubiquinone oxidoreductase subunit (NQO10)	19.4%
HP1269	NADH-ubiquinone oxidoreductase, NQO10 subunit (NQO10)	-1.0%
HP1270	NADH-ubiquinone oxidoreductase, NQO11 subunit (NQO11) {[Paracoccus denitrificans}	42.6%
HP1271	NADH-ubiquinone oxidoreductase, NQO12 subunit (NQO12)	43.2%
HP1272	NADH-ubiquinone oxidoreductase, NQO13 subunit (NQO13)	40.2%
HP1273	NADH-ubiquinone oxidoreductase, NQO14 subunit (NQO14)	31.2%
HP1266	NADH-ubiquinone oxidoreductase, NQO3 subunit (NQO3)	31.6%
HP1263	NADH-ubiquinone oxidoreductase, NQO4 subunit (NQO4){Triticum aestivum}	44.6%
HP1262	NADH-ubiquinone oxidoreductase, NQO5 subunit (NQO5)	-1.0%
HP1261	NADH-ubiquinone oxidoreductase, NQO6 subunit (NQO6)	62.2%
HP1260	NADH-ubiquinone oxidoreductase, NQO7 subunit (NQO7)	40.7%
HP1267	NADH-ubiquinone oxidoreductase, NQO8 subunit (NQO8)	42.4%
HP1268	NADH-ubiquinone oxidoreductase, NQO9 subunit (NQO9)	41.2%
Amino acids and amines		
HP1398	alanine dehydrogenase (ald)	39.6%
HP0294	aliphatic amidase (aimE)	75.4%
HP1238	aliphatic amidase (aimE)	37.2%
HP1399	arginase (rocF)	31.8%
HP0943	D-amino acid dehydrogenase (dadA)	26.2%
HP0056	delta-1-pyrroline-5-carboxylate dehydrogenase {Synechocystis sp.)	32.2%
HP0723	L-asparaginase II (ansB)	54.1%
HP0132	L-serine deaminase (sdaA)	45.8%
Anaerobic		
HP0666	anaerobic glycerol-3-phosphate dehydrogenase, subunit C (glpC)	27.2%
HP0589	ferredoxin oxidoreductase, alpha subunit	42.7%
HP0590	ferredoxin oxidoreductase, beta subunit	43.2%
HP0591	ferredoxin oxidoreductase, gamma subunit	33.3%
HP0193	fumarate reductase, cytochrome b subunit (frdC)	58.8%
HP0192	fumarate reductase, flavoprotein subunit (frdA)	69.4%
HP0191	fumarate reductase, iron-sulfur subunit (frdB)	70.8%
HP1110	pyruvate ferredoxin oxidoreductase, alpha subunit	41.0%
HP1111	pyruvate ferredoxin oxidoreductase, beta subunit	43.7%
HP1109	pyruvate ferredoxin oxidoreductase, delta subunit	47.0%
HP1108	pyruvate ferredoxin oxidoreductase, gamma subunit	37.2%
ATP-protonmotive force interconversion		
HP0828	ATP synthase FO, subunit a (atpB)	37.7%
HP1136	ATP synthase FO, subunit b (atpF)	28.3%
HP1137	ATP synthase FO, subunit bÕ (atpFÕ)	32.5%
HP1212	ATP synthase FO, subunit c (atpE)	41.2%
HP1134	ATP synthase F1, subunit alpha (atpA)	62.7%
HP1132	ATP synthase F1, subunit beta (atpD)	85.6%
HP1135	ATP synthase F1, subunit delta (atpH)	24.6%
HP1131	ATP synthase F1, subunit epsilon (atpC)	32.7%
HP1133	ATP synthase F1, subunit gamma (atpG)	37.8%
Electron transport		
HP0146	cbb3-type cytochrome c oxidase subunit Q (CcoQ)	44.2%
HP0265	cytochrome c biogenesis protein (ccdA)	35.4%
HP0378	cytochrome c biogenesis protein (ycf5)	37.5%
HP0147	cytochrome c oxidase, diheme subunit, membrane-bound (fixP)	33.0%
HP0144	cytochrome c oxidase, heme b and copper-binding subunit, membrane-bound (fixN)	43.9%
HP0145	cytochrome c oxidase, monoheme subunit, membrane-bound (fixO)	45.7%
HP1461	cytochrome c551 peroxidase	48.5%
HP1227	cytochrome c553	38.4%
HP0277	ferredoxin	52.5%
HP0588	ferrodoxin-like protein	42.6%
HP1508	ferrodoxin-like protein	29.4%
HP1161	flavodoxin (fldA)	47.0%
HP0642	NAD(P)H-flavin oxidoreductase	46.1%
HP0954	oxygen-insensitive NAD(P)H nitroreductase	32.7%
HP0634	quinone-reactive Ni/Fe hydrogenase (hydD)	54.7%
HP0633	quinone-reactive Ni/Fe hydrogenase, cytochrome b subunit (hydC)	51.4%
HP0632	quinone-reactive Ni/Fe hydrogenase, large subunit (hydB)	68.5%
HP0631	quinone-reactive Ni/Fe hydrogenase, small subunit (hydA)	68.9%
HP1539	ubiquinol cytochrome c oxidoreductase, cytochrome b subunit (fbcH)	39.3%
HP1538	ubiquinol cytochrome c oxidoreductase, cytochrome c1 subunit (fbcH)	28.8%
HP1540	ubiquinol cytochrome c oxidoreductase, Rieske 2Fe-2S subunit (fbcF)	39.2%
Entner-Doudoroff		
HP1099	2-keto-3-deoxy-6-phosphogluconate aldolase (eda)	50.3%
HP1100	6-phosphogluconate dehydratase	50.7%
Fermentation		
HP0691	3-oxoadipate coA-transferase subunit A (yxjD)	65.5%
HP0692	3-oxoadipate coA-transferase subunit B (yxjE)	73.2%
HP0903	acetate kinase (ackA) {Escherichia coli)	42.3%
HP0904	phosphate acetyltransferase (pta)	51.0%
HP0905	phosphotransacetylase (pta)	26.9%
HP0357	short-chain alcohol dehydrogenase	57.6%
Gluconeogenesis		
HP1385	fructose-1,6-bisphosphatase	36.4%
HP0121	phosphoenolpyruvate synthase (ppsA)	52.4%
HP1345	phosphoglycerate kinase	47.3%
Glycolysis		
HP0154	enolase (eno)	56.9%
HP0176	fructose-bisphosphate aldolase (tsr)	46.0%
HP1103	glucokinase (glk)	41.5%
HP1166	glucose-6-phosphate isomerase (pgi)	53.3%
HP0921	glyceraldehyde-3-phosphate dehydrogenase (gap)	46.5%
HP1346	glyceraldehyde-3-phosphate dehydrogenase (gap)	46.7%
HP0974	phosphoglycerate mutase (pgm)	44.6%
HP0194	triosephosphate isomerase (tpi)	34.5%
Pentose phosphate pathway		
HP1386	D-ribulose-5-phosphate 3 epimerase (rpe)	44.2%
HP1102	glucose-6-phosphate 1-dehydrogenase (devB)	29.2%
HP1101	glucose-6-phosphate dehydrogenase (g6pD)	36.7%
HP1495	transaldolase (tal)	33.5%
HP1088	transketolase A (tktA)	46.7%
HP0354	transketolase B (tktB)	39.7%
Sugars		
HP0574	galactosidase acetyltransferase (lacA)	41.0%
HP0360	UDP-glucose 4-epimerase	43.1%
TCA cycle		
HP0779	aconitase B (acnB)	64.0%
HP0026	citrate synthase (gltA)	47.8%
HP1325	fumarase (fumC)	63.7%
HP0509	glycolate oxidase subunit (glcD)	98.0%
HP0027	isocitrate dehydrogenase (icd)	70.7%

FATTY ACID AND PHOSPHOLIPID METABOLISM

General		
HP1376	(3R)-hydroxymyristoyl-[acyl carrier protein] dehydratase (fabZ)	47.4%
HP1348	1-acyl-glycerol-3-phosphate acyltransferase (plsC) {Escherichia coli)	32.0%
HP0561	3-ketoacyl-acyl carrier protein reductase (fabG)	45.7%
HP0690	acetyl coenzyme A acetyltransferase (thiolase) (fadA)	52.0%
HP0950	acetyl-CoA carboxylase beta subunit (accD)	49.4%
HP1045	acetyl-CoA synthetase (acoE)	52.3%
HP0557	acetyl-coenzyme A carboxylase (accA)	50.3%
HP0559	acyl carrier protein (acpP)	55.3%
HP0962	acyl carrier protein (acpP)	56.3%
HP0558	beta ketoacyl-acyl carrier protein synthase II (fabF)	50.0%
HP0202	beta-ketoacyl-acyl carrier protein synthase III (fabH)	44.4%
HP0371	biotin carboxyl carrier protein (fabE)	30.8%
HP0370	biotin carboxylase (accC)	52.1%
HP0871	CDP-diglyceride hydrolase (cdh)	73.9%
HP0215	CDP-diglyceride synthetase (cdsA)	42.4%
HP0416	cyclopropane fatty acid synthase (cfa)	39.7%
HP0700	diacylglycerol kinase (dgkA)	45.8%
HP0195	enoyl-[acyl-carrier-protein) reductase (NADH) (fabI)	45.8%
HP0201	fatty acid/phospholipid synthesis protein (plsX)	37.8%
HP0808	Holo-acp synthase (acpS)	29.1%
HP0090	malonyl coenzyme A-acyl carrier protein transacylase (fabD)	35.4%
HP1016	phosphatidylglycerophosphate synthase (pgsA)	35.4%
HP1357	phosphatidylserine decarboxylase proenzyme (psd)	33.2%
HP1071	phosphatidylserine synthase (pssA)	99.6%
HP0499	phospholipase A1 precursor (DR-phospholipase A)	33.8%

PURINES, PYRIMIDINES, NUCLEOSIDES AND NUCLEOTIDES

General		
HP0757	beta-alanine synthetase homologue	40.0%
2Õ-Deoxyribonucleotide metabolism		
HP0372	deoxycytidine triphosphate deaminase (dcd)	28.2%
HP0865	deoxyuridine 5Õ-triphosphate nucleotidohydrolase (dut)	41.4%
HP0364	ribonucleoside diphosphate reductase, beta subunit (nrdB)	39.0%
HP0680	ribonucleoside-diphosphate reductase 1 alpha subunit (nrdA)	28.4%
HP0825	thioredoxin reductase (trxB)	45.9%
Purine ribonucleotide biosynthesis		
HP0321	5Õ-guanylate kinase (gmk)	44.8%
HP0618	adenylate kinase (adk)	33.3%
HP1112	adenylosuccinate lyase (purB)	49.5%
HP0255	adenylosuccinate synthetase (purA)	44.6%
HP1434	formyltetrahydrofolate hydrolase (purU)	49.1%
HP1218	glycinamide ribonucleotide synthetase (purD)	31.8%
HP0854	GMP reductase (guaC)	31.8%
HP0409	GMP synthase (guaA)	56.1%
HP0829	inosine-5Õ-monophosphate dehydrogenase (guaB)	58.5%
HP0198	nucleoside diphosphate kinase (ndk)	67.7%
HP0742	phosphoribosylpyrophosphate synthetase (prsA)	56.5%
HP1530	purine nucleoside phosphorylase (punB)	20.7%
Pyrimidine ribonucleotide biosynthesis		
HP1084	aspartate transcarbamoylase (pyrB)	38.7%
HP0919	carbamoyl-phosphate synthase (glutamine-hydrolysing) (pyrAb)	48.6%
HP1237	carbamoyl-phosphate synthetase (pyrAa)	39.7%
HP0349	CTP synthetase (pyrG)	50.7%
HP0266	dihydroorotase (pyrC)	-1.0%
HP0581	dihydroorotase (pyrC)	31.5%
HP1011	dihydroorotate dehydrogenase (pyrD)	41.5%
HP1257	orotate phosphoribosyltransferase (pyrE)	35.5%
HP0005	orotidine 5Õ-phosphate decarboxylase (pyrF)	39.0%
HP1474	thymidylate kinase (tmk)	33.9%
HP0777	uridine 5Õ-monophosphate (UMP) kinase (pyrH)	50.4%
Salvage of nucleosides and nucleotides		
HP0104	2Õ,3Õ-cyclic-nucleotide 2Õ-phosphodiesterase (cpdB)	31.8%
HP0572	adenine phosphoribosyltransferase (apt)	50.3%
HP1179	phosphopentomutase (deoB)	55.9%
HP1178	purine-nucleoside phosphorylase (deoD)	55.5%
HP0735	xanthine guanine phosphoribosyl transferase (gpt)	27.1%
Sugar-nucleotide biosynthesis and conversions		
HP0043	mannose-6-phosphate isomerase (pmi) or (algA)	42.8%
HP0045	nodulation protein (nolK)	44.3%
HP0646	UDP-glucose pyrophosphorylase (galU)	65.6%
HP0683	UDP-N-acetylglucosamine pyrophosphorylase (glmU)	40.0%

REGULATORY FUNCTIONS

General		
HP1032	alternative transcription initiation factor, sigma-F (fliA)	34.6%
HP1168	carbon starvation protein (cstA)	59.8%
HP1442	carbon storage regulator (csrA)	43.3%
HP1027	ferric uptake regulation protein (fur)	39.9%
HP0278	guanosine pentaphosphate phosphohydrolase (gppA)	26.4%
HP0400	penicillin tolerance protein (lytB)	30.6%

HP0775	penta-phosphate guanosine-3Ō-pyrophospho-hydrolase (spoT)	36.7%
HP0224	peptide methionine sulphoxide reductase (msrA)	66.8%
HP1025	putative heat shock protein (hspR)	46.2%
HP1572	regulatory protein DniR	31.9%
HP0703	response regulator	44.2%
HP1021	response regulator	28.7%
HP1043	response regulator	26.8%
HP1365	response regulator	32.4%
HP0166	response regulator (ompR)	51.0%
HP0714	RNA polymerase sigma-54 factor (rpoN)	37.1%
HP0088	RNA polymerase sigma-70 factor (rpoD)	43.5%
HP0792	sigma-54 interacting protein	97.7%
HP0164	signal-transducing protein, histidine kinase	27.1%
HP1364	signal-transducing protein, histidine kinase	24.9%
HP0244	signal-transducing protein, histidine kinase (atoS)	30.0%
HP0048	transcriptional regulator (hypF)	34.5%
HP1287	transcriptional regulator (tenA)	34.7%
HP0727	transcriptional regulator, putative	33.3%

REPLICATION

Degradation of DNA

HP0275	ATP-dependent nuclease (addB)	27.2%
HP0259	exonuclease VII, large subunit (xseA)	37.6%

DNA replication, restriction, modification, recombination and repair

HP0142	A/G-specific adenine glycosylase (mutY)	38.2%
HP0050	adenine specific DNA methyltransferase (dpnA)	37.4%
HP0910	adenine specific DNA methyltransferase (HINDIIM)	33.4%
HP1352	adenine specific DNA methyltransferase (HINFIM)	62.5%
HP0263	adenine specific DNA methyltransferase (hpaim)	33.9%
HP0481	adenine specific DNA methyltransferase (MFOKI)	29.3%
HP0260	adenine specific DNA methyltransferase (mod)	33.9%
HP0593	adenine specific DNA methyltransferase (mod)	38.5%
HP1522	adenine specific DNA methyltransferase (mod)	42.2%
HP0478	adenine specific DNA methyltransferase (VSPIM)	42.1%
HP0054	adenine/cytosine DNA methyltransferase	32.1%
HP0790	anti-codon nuclease masking agent (prrB)	42.9%
HP1529	chromosomal replication initiator protein (dnaA)	34.9%
HP1121	cytosine specific DNA methyltransferase (BSP6IM)	37.0%
HP0051	cytosine specific DNA methyltransferase (DDEM)	39.0%
HP0483	cytosine specific DNA methyltransferase (HPHIMC)	38.7%
HP0701	DNA gyrase, sub A (gyrA)	97.4%
HP0501	DNA gyrase, sub B (gyrB)	46.0%
HP1478	DNA helicase II (uvrD)	35.3%
HP0548	DNA helicase, putative	38.8%
HP0615	DNA ligase (lig)	40.1%
HP0621	DNA mismatch repair protein (MutS)	32.6%
HP1470	DNA polymerase I (polA)	40.0%
HP1460	DNA polymerase III alpha-subunit (dnaE)	42.0%
HP0500	DNA polymerase III beta-subunit (dnaN)	26.0%
HP1231	DNA polymerase III delta prime subunit (holB)	48.6%
HP1387	DNA polymerase III epsilon subunit (dnaQ)	35.1%
HP0717	DNA polymerase III gamma and tau subunits (dnaX)	39.0%
HP0012	DNA primase (dnaG)	36.6%
HP1523	DNA recombinase (recG)	32.7%
HP1393	DNA repair protein (recN)	28.3%
HP0116	DNA topoisomerase I (topA)	45.1%
HP0440	DNA topoisomerase I (topA)	31.7%
HP0602	endonuclease III	36.6%
HP0585	endonuclease III (nth)	40.1%
HP0705	excinuclease ABC subunit A (uvrA)	53.4%
HP1114	excinuclease ABC subunit B (uvrB)	53.1%
HP0821	excinuclease ABC subunit C (uvrC)	31.5%
HP1526	exodeoxyribonuclease (lexA)	58.9%
HP0213	glucose inhibited division protein (gidA)	48.5%
HP1063	glucose-inhibited division protein (gidB)	32.9%
HP1553	helicase	33.0%
HP0883	Holliday junction DNA helicase (ruvA)	39.0%
HP1059	Holliday junction DNA helicase (ruvB)	54.6%
HP0877	Holliday junction endodeoxyribonuclease (ruvC)	34.7%
HP0675	integrase/recombinase (xerC)	31.8%
HP0995	integrase/recombinase (xerD)	27.8%
HP0323	membrane bound endonuclease (nuc)	31.1%
HP0676	methylated-DNĀprotein-cysteine methyltransferase (dat1)	41.0%
HP0387	primosomal protein replication factor (priA)	36.3%
HP0153	recombinase (recA)	99.1%
HP0925	recombinational DNA repair protein (recR)	36.5%
HP0911	rep helicase, single-stranded DNA-dependent ATPase (rep)	33.8%
HP1362	replicative DNA helicase (dnaB)	39.4%
HP1383	restriction modification system S subunit	38.1%
HP0661	ribonuclease H (rnhA)	58.4%
HP1323	ribonuclease HII (rnhB)	36.3%
HP1245	single-strand DNA-binding protein (ssb)	32.6%
HP0348	single-stranded-DNA-specific exonuclease (recJ)	33.6%
HP1009	site-specific recombinase	21.3%
HP1541	transcription-repair coupling factor (trcF)	37.7%
HP0462	type I restriction enzyme S protein (hsdS)	37.0%
HP0463	type I restriction enzyme M protein (hsdM)	29.4%
HP0464	type I restriction enzyme R protein (hsdR)	31.7%
HP0846	type I restriction enzyme R protein (hsdR)	48.0%
HP0848	type I restriction enzyme S protein (hsdS)	37.0%
HP0850	type I restriction enzyme M protein (hsdM)	54.4%
HP1402	type I restriction enzyme R protein (hsdR)	26.6%
HP1403	type I restriction enzyme M protein (hsdM)	37.1%
HP1404	type I restriction enzyme S protein (hsdS)	36.0%
HP0092	type II restriction enzyme M protein (hsdM)	55.3%
HP0091	type II restriction enzyme R protein (hsdR)	50.7%
HP1369	type III restriction enzyme M protein (mod)	45.6%
HP1370	type III restriction enzyme M protein (mod)	37.0%
HP1371	type III restriction enzyme R protein	26.2%
HP0592	type III restriction enzyme R protein (res)	30.6%
HP1521	type III restriction enzyme R protein (res)	33.1%
HP1472	type IIS restriction enzyme M protein (mod)	32.4%
HP1367	type IIS restriction enzyme M1 protein (mod) ({Moraxella bovis}	59.3%
HP1368	type IIS restriction enzyme M2 protein (mod)	33.0%
HP1517	type IIS restriction enzyme R and M protein (ECO57IR)	26.7%
HP1471	type IIS restriction enzyme R protein (BCGIB)	28.2%
HP1366	type IIS restriction enzyme R protein (MBOIIR)	37.1%
HP1208	ulcer associated adenine specific DNA methyltransferase	93.4%
HP1209	ulcer-associated gene restriction endonuclease (iceA)	95.5%
HP1347	uracil-DNA glycosylase (ung)	43.1%

TRANSCRIPTION

Degradation of RNA

HP1213	polynucleotide phosphorylase (pnp) DNA-dependent RNA polymerase	38.9%
HP1293	DNA-directed RNA polymerase, alpha subunit (rpoA)	35.3%
HP1198	DNA-directed RNA polymerase, beta subunit (rpoB)	47.8%

Transcription factors

HP0866	transcription elongation factor GreA (greA)	50.3%
HP1514	transcription termination factor NusA (nusA)	39.1%
HP0001	transcription termination factor NusB (nusB)	30.2%
HP1203	transcription termination factor NusG (nusG)	41.0%
HP0550	transcription termination factor Rho (rho)	56.6%

RNA processing

HP0640	poly(A) polymerase (papS)	37.4%
HP0662	ribonuclease III (rnc)	37.3%

TRANSLATION

General

HP0944	translation initiation inhibitor, putative	45.6%

Aminoacyl tRNA synthetases

HP1241	alanyl-tRNA synthetase (alaS)	44.9%
HP0319	arginyl-tRNA synthetase (argS)	35.8%
HP0617	aspartyl-tRNA synthetase (aspS)	50.1%
HP0886	cysteinyl-tRNA synthetase (cysS)	97.3%
HP0476	glutamyl-tRNA synthetase (gltX)	43.1%
HP0643	glutamyl-tRNA synthetase (gltX)	39.8%
HP0960	glycyl-tRNA synthetase, alpha subunit (glyQ)	60.1%
HP0972	glycyl-tRNA synthetase, beta subunit (glyS)	33.6%
HP1190	histidyl-tRNA synthetase (hisS)	32.4%
HP1422	isoleucyl-tRNA synthetase (ileS)	49.7%
HP1547	leucyl-tRNA synthetase (leuS)	45.9%
HP0182	lysyl-tRNA synthetase (lysS)	58.6%
HP0417	methionyl-tRNA synthetase (metS)	42.4%
HP0403	phenylalanyl-tRNA synthetase, alpha subunit (pheS)	48.7%
HP0402	phenylalanyl-tRNA synthetase, beta subunit (pheT)	30.0%
HP0238	prolyl-tRNA synthetase (proS)	39.8%
HP1480	seryl-tRNA synthetase (serS)	48.3%
HP0123	threonyl-tRNA synthetase (thrS)	42.1%
HP1253	tryptophanyl-tRNA synthetase (trpS)	52.6%
HP0774	tyrosyl-tRNA synthetase (tyrS)	54.7%
HP1153	valyl-tRNA synthetase (valS)	43.7%

Degradation of proteins, peptides and glycopeptides

HP0570	aminopeptidase a/i (pepA)	38.5%
HP0033	ATP-dependent C1p protease (clpA)	40.3%
HP0794	ATP-dependent clp protease proteolytic component (clpP)	64.6%
HP1379	ATP-dependent protease (lon)	43.9%
HP0223	ATP-dependent protease (sms)	41.0%
HP1374	ATP-dependent protease ATPase subunit (clpX)	56.3%
HP0264	ATP-dependent protease binding subunit (clpB)	97.7%
HP0169	collagenase (prtC)	40.1%
HP0516	heat-shock protein (hslU) ORF1	98.4%
HP0515	heat-shock protein (hslV)	57.1%
HP0470	oligoendopeptidase F (pepF)	97.9%
HP0657	processing protease (ymxG)	24.2%
HP1485	proline dipeptidase (pepQ)	35.2%
HP1350	protease	40.6%
HP1012	protease (pqqE)	29.6%
HP1435	protease IV (PspA)	41.7%
HP0404	protein kinase C inhibitor (SP:P16436)	40.2%
HP1019	serine protease (htrA)	52.9%
HP1584	sialoglycoprotease (gcp)	35.7%
HP0382	zinc-metalloprotease (YJR117W)	36.2%

Nucleoproteins

HP0835	histone-like DNA-binding protein HU (hup)	44.6%

Protein modification

HP0363	L-isoaspartyl-protein carboxyl methyltransferase (pcm)	43.0%
HP1299	methionine amino peptidase (map)	43.0%
HP1441	peptidyl-prolyl *cis-trans* isomerase B, cyclosporin-type rotamase (ppi)	58.1%
HP1123	peptidyl-prolyl *cis-trans* isomerase, FKBP-type rotamase (slyD)	40.4%
HP0793	polypeptide deformylase (def)	41.8%

Ribosomal proteins: synthesis and modification

HP1201	ribosomal protein L1 (rpl1)	52.0%
HP1200	ribosomal protein L10 (rpl10)	30.4%
HP1202	ribosomal protein L11 (rpl11)	63.8%
HP1068	ribosomal protein L11 methyltransferase (prmA)	38.4%
HP0084	ribosomal protein L13 (rpl13)	50.0%
HP1309	ribosomal protein L14 (rpl14)	65.9%
HP1301	ribosomal protein L15 (rpl15)	42.5%
HP1312	ribosomal protein L16 (rpl16)	62.4%
HP1292	ribosomal protein L17 (rpl17)	48.3%
HP1303	ribosomal protein L18 (rpl18)	45.5%
HP1147	ribosomal protein L19 (rpl19)	50.9%
HP1316	ribosomal protein L2 (rpl2)	58.9%
HP0126	ribosomal protein L20 (rpl20)	54.8%
HP0296	ribosomal protein L21 (rpl21)	46.1%
HP1314	ribosomal protein L22 (rpl22)	44.9%
HP1317	ribosomal protein L23 (rpl23)	31.7%
HP1308	ribosomal protein L24 (rpl24)	52.2%
HP0297	ribosomal protein L27 (rpl27)	64.7%
HP0491	ribosomal protein L28 (rpL28)	41.7%
HP1311	ribosomal protein L29 (rpL29)	45.6%
HP1319	ribosomal protein L3 (rpl3)	41.8%
HP0551	ribosomal protein L31 (rpl31)	49.3%
HP0200	ribosomal protein L32 (rpl32)	41.7%
HP1204	ribosomal protein L33 (rpL33)	55.1%
HP1447	ribosomal protein L34 (rpl34)	70.5%
HP0125	ribosomal protein L35 (rpl35)	50.8%
HP1297	ribosomal protein L36 (rpl36)	81.6%
HP1318	ribosomal protein L4 (rpl4)	40.6%
HP1307	ribosomal protein L5 (rpl5)	53.1%
HP1304	ribosomal protein L6 (rpl6)	44.4%
HP1199	ribosomal protein L7/L12 (rpl7/l12)	65.0%
HP0514	ribosomal protein L9 (rpl9)	39.6%
HP0399	ribosomal protein S1 (rps1)	30.5%
HP1320	ribosomal protein S10 (rps10)	58.2%
HP1295	ribosomal protein S11 (rps11)	56.2%
HP1197	ribosomal protein S12 (rps12)	79.0%
HP1296	ribosomal protein S13 (rps13)	55.8%
HP1306	ribosomal protein S14 (rpS14)	68.3%
HP1040	ribosomal protein S15 (rps15)	57.8%
HP1151	ribosomal protein S16 (rpS16)	46.8%
HP1310	ribosomal protein S17 (rps17)	55.4%
HP1244	ribosomal protein S18 (rps18)	55.2%
HP1315	ribosomal protein S19 (rps19)	61.1%
HP1554	ribosomal protein S2 (rps2)	49.6%
HP0076	ribosomal protein S20 (rps20)	41.4%
HP0562	ribosomal protein S21 (rps21)	42.4%
HP1313	ribosomal protein S3 (rps3)	56.7%
HP1294	ribosomal protein S4 (rps4)	51.2%
HP1302	ribosomal protein S5 (rps5)	65.5%
HP1246	ribosomal protein S6 (rps6)	32.1%
HP1196	ribosomal protein S7 (rps7)	62.2%
HP1305	ribosomal protein S8 (rps8)	45.0%
HP0083	ribosomal protein S9 (rps9)	50.4%
HP1047	ribosome-binding factor A (rbfA)	26.3%

tRNA modification

HP1141	methionyl-tRNA formyltransferase (fmt)	37.5%
HP1497	peptidyl-tRNA hydrolase (pth)	46.6%
HP0361	pseudouridylate synthase I (hisT)	32.2%
HP1448	ribonuclease P, protein component (rnpA)	29.3%
HP1062	S-adenosylmethionine:tRNA ribosyltransferase-isomerase (queA)	39.3%
HP1513	selenocystein synthase (selA)	36.2%
HP1148	tRNA (guanine-N1)-methyltransferase (trmD)	39.1%
HP1415	tRNA delta(2)-isopentenylpyrophosphate transferase (miaA)	30.7%
HP0281	tRNA-guanine transglycosylase (tgt)	45.6%

Translation factors

HP0247	ATP-dependent RNA helicase, DEAD-box family (deaD)	37.7%
HP0077	peptide chain release factor RF-1 (prfA)	52.6%
HP0171	peptide chain release factor RF-2 (prfB)	49.6%
HP1256	ribosome releasing factor (frr)	43.7%
HP1195	translation elongation factor EF-G (fusA)	67.5%
HP0177	translation elongation factor EF-P (efp)	45.1%
HP1555	translation elongation factor EF-Ts (tsf)	43.1%
HP1205	translation elongation factor EF-Tu (tufB)	89.5%
HP1298	translation initiation factor EF-1 (infA)	65.3%
HP1048	translation initiation factor IF-2 (infB)	45.4%
HP0124	translation initiation factor IF-3 (infC)	43.4%

TRANSPORT AND BINDING PROTEINS

General

HP0179	ABC transporter, ATP-binding protein	66.7%
HP0613	ABC transporter, ATP-binding protein	31.1%
HP0715	ABC transporter, ATP-binding protein	52.3%
HP1576	ABC transporter, ATP-binding protein (abc)	48.2%
HP1465	ABC transporter, ATP-binding protein (HI1087)	37.8%
HP1220	ABC transporter, ATP-binding protein (yhcG)	31.5%
HP0853	ABC transporter, ATP-binding protein (yheS)	36.3%
HP1577	ABC transporter, permease protein (yaeE)	43.1%
HP0607	acriflavine resistance protein (acrB)	29.7%
HP1432	histidine and glutamine-rich protein	50.0%
HP1427	histidine-rich, metal binding polypeptide (hpn)	100.0%
HP1206	multidrug-resistance protein (hetA)	26.2%
HP1082	multidrug-resistance protein (msbA)	32.4%
HP0600	multidrug-resistance protein (spaB)	29.7%
HP1181	multidrug-efflux transporter	29.1%
HP0497	sodium- and chloride-dependent transporter	31.7%
HP0498	sodium- and chloride-dependent trans-porter	30.8%
HP0214	sodium-dependent transporter (huNaDC-1)	36.6%

Amino acids, peptides and amines

HP0940	amino acid ABC transporter, periplasmic binding protein (yckK)	41.5%
HP0939	amino acid ABC transporter, permease protein (yckJ)	46.9%
HP1017	amino acid permease (rocE)	41.7%
HP0942	D-alanine glycine permease (dagA)	44.5%
HP0301	dipeptide ABC transporter, ATP-binding protein (dppD)	59.5%
HP0302	dipeptide ABC transporter, ATP-binding protein (dppF)	54.8%
HP0298	dipeptide ABC transporter, periplasmic dipeptide-binding protein (dppA)	39.8%
HP0299	dipeptide ABC transporter, permease protein (dppB)	49.3%
HP0300	dipeptide ABC transporter, permease protein (dppC)	52.5%
HP1506	glutamate permease (gltS)	56.9%
HP1171	glutamine ABC transporter, ATP-binding protein (glnQ)	51.9%
HP1172	glutamine ABC transporter, periplasmic glutamine-binding protein (glnH)	32.2%
HP1169	glutamine ABC transporter, permease protein (glnP)	27.6%
HP1170	glutamine ABC transporter, permease protein (glnP)	30.9%
HP0250	oligopeptide ABC transporter, ATP-binding protein (oppD)	39.1%
HP1252	oligopeptide ABC transporter, periplasmic oligopeptide-binding protein (oppA)	28.7%
HP1251	oligopeptide ABC transporter, permease protein (oppB)	59.6%
HP0251	oligopeptide ABC transporter, permease protein (oppC)	31.4%
HP0819	osmoprotection protein (proV)	38.3%
HP0818	osmoprotection protein (proWX)	30.4%
HP0055	proline permease (putP)	51.4%
HP0936	proline/betaine transporter (proP)	29.1%
HP0133	serine transporter (sdaC)	44.6%

Anions

HP0475	molybdenum ABC transporter, ATP-binding protein (modD)	38.4%
HP0473	molybdenum ABC transporter, periplasmic molybdate-binding protein (modA)	95.9%
HP0474	molybdenum ABC transporter, permease protein (modB)	28.7%
HP0313	nitrite extrusion protein (narK)	23.6%
HP1491	phosphate permease	34.8%

Carbohydrates, organic alcohols and acids

HP0143	2-oxoglutarate/malate translocator (SODiT1)	37.0%
HP1091	alpha-ketoglutarate permease (kgtP)	45.9%
HP0724	anaerobic C4-dicarboxylate transport protein (dcuA)	53.8%
HP1174	glucose/galactose transporter (gluP)	53.6%
HP0141	L-lactate permease (lctP)	55.5%
HP0140	L-lactate permease (lctP)	58.7%

Cations

HP0791	cadmium-transporting ATPase, P-type (cadA)	97.5%
HP0969	cation efflux system protein (czcA)	37.3%
HP1328	cation efflux system protein (czcA)	28.9%
HP1329	cation efflux system protein (czcA)	31.3%
HP1503	cation-transporting ATPase, P-type (copA)	30.3%
HP1073	copper ion binding protein (copP)	92.4%
HP1072	copper-transporting ATPase, P-type (copA)	93.9%
HP0471	glutathione-regulated potassium-efflux system protein (kefB)	99.3%
HP0687	iron(II) transport protein (feoB)	33.6%
HP1561	iron(III) ABC transporter, periplasmic iron-binding protein (ceuE)	27.5%
HP1562	iron(III) ABC transporter, periplasmic iron-binding protein (ceuE)	28.2%
HP0888	iron(III) dicitrate ABC transporter, ATP-binding protein (fecE)	34.4%
HP0889	iron(III) dicitrate ABC transporter, permease protein (fecD)	38.3%
HP0686	iron(III) dicitrate transport protein (fecA)	29.7%
HP0807	iron(III) dicitrate transport protein (fecA)	28.5%
HP1400	iron(III) dicitrate transport protein (fecA)	26.3%
HP1344	magnesium and cobalt transport protein (corA)	26.3%
HP1183	NA+/H+ antiporter (napA)	26.6%
HP1552	Na+/H+ antiporter (nhaA)	49.2%
HP1077	nickel transport protein (nixA)	98.7%
HP0490	putative potassium channel protein, putative	25.7%

Nucleosides, purines and pyrimidines

HP1290	nicotinamide mononucleotide transporter (pnuC)	28.0%
HP1180	pyrimidine nucleoside transport protein (nupC)	32.9%

Other

HP0876	iron-regulated outer membrane protein (frpB)	27.6%
HP0915	iron-regulated outer membrane protein (frpB)	28.1%
HP0916	iron-regulated outer membrane protein (frpB)	28.8%
HP1129	biopolymer transport protein (exbD)	29.7%
HP1130	biopolymer transport protein (exbB)	33.5%
HP1339	biopolymer transport protein (exbB)	46.8%
HP1340	biopolymer transport protein (exbD)	35.8%
HP1445	biopolymer transport protein (exbB)	45.5%
HP1446	biopolymer transport protein (exbD)	36.2%
HP1512	iron-regulated outer membrane protein (frpB)	26.6%
HP0653	nonheme iron-containing ferritin (pfr)	99.4%
HP1341	siderophore-mediated iron transport protein (tonB)	37.2%

OTHER CATEGORIES

General

HP0924	4-oxalocrotonate tautomerase (dmpI)	37.7%
HP1034	ATP-binding protein (ylxH)	36.3%
HP1000	PARA protein	29.7%
HP1139	SpoOJ regulator (soj)	47.4%
HP0827	ss-DNA binding protein 12RNP2 precursor	46.8%

Adaptations and atypical conditions

HP1496	general stress protein (ctc)	26.5%
HP1483	gerC2 protein (gerC2)	33.3%
HP0927	heat shock protein (htpX)	32.8%
HP0280	heat shock protein B (ibpB)	27.2%
HP1228	invasion protein (invA)	38.2%
HP0970	nickel-cobalt-cadmium resistance protein (nccB)	21.1%
HP1444	small protein (smpB)	42.1%
HP0930	stationary-phase survival protein (surE)	37.7%
HP0315	virulence associated protein D (vapD)	70.2%
HP0967	virulence associated protein D (vapD)	28.9%
HP1248	virulence associated protein homolog (vacB)	36.0%
HP0885	virulence factor mviN protein (mviN)	33.5%

Colicin-related functions

HP1126	colicin tolerance-like protein (tolB)	25.7%
HP0428	phage/colicin/tellurite resistance cluster terY protein	25.6%

Drug and analog sensitivity

HP1431	16S rRNA (adenosine-N6,N6-)-dimethyl-transferase (ksgA)	35.5%
HP0606	membrane fusion protein (mtrC)	24.2%
HP0630	modulator of drug activity (mda66)	62.3%
HP1476	phenylacrylic acid decarboxylase	39.7%
HP1165	tetracycline resistance protein tetA(P), putative	27.0%

Transposon-related functions

HP1008	IS200 insertion sequence from SARA17	33.9%
HP0414	IS200 insertion sequence from SARA17	33.9%
HP0988	IS605 transposase (tnpA)	97.2%
HP0998	IS605 transposase (tnpA)	97.2%
HP1096	IS605 transposase (tnpA)	97.2%
HP1535	IS605 transposase (tnpA)	97.2%
HP0437	IS605 transposase (tnpA)	97.2%
HP0989	IS605 transposase (tnpB)	93.4%
HP0997	IS605 transposase (tnpB)	93.4%
HP1095	IS605 transposase (tnpB)	93.4%
HP1534	IS605 transposase (tnpB)	93.4%
HP0438	IS605 transposase (tnpB)	93.4%
HP0413	transposase-like protein, PS3IS	33.6%
HP1007	transposase-like protein, PS3IS	34.3%

Other

HP0739	2-hydroxy-6-oxohepta-2,4-dienoate hydrolase	30.1%

HYPOTHETICAL

General

HP0831	conserved hypothetical ATP binding protein	32.3%
HP0066	conserved hypothetical ATP-binding protein	34.7%
HP0269	conserved hypothetical ATP-binding protein	37.7%
HP0312	conserved hypothetical ATP-binding protein	34.1%
HP1321	conserved hypothetical ATP-binding protein	30.8%
HP1430	conserved hypothetical ATP-binding protein	38.1%
HP1507	conserved hypothetical ATP-binding protein	51.6%
HP1567	conserved hypothetical ATP-binding protein	40.9%
HP1026	conserved hypothetical helicase-like protein	35.2%
HP0022	conserved hypothetical integral membrane protein	30.8%
HP0189	conserved hypothetical integral membrane protein	43.1%
HP0226	conserved hypothetical integral membrane protein	27.6%
HP0228	conserved hypothetical integral membrane protein	43.2%
HP0234	conserved hypothetical integral membrane protein	32.4%
HP0258	conserved hypothetical integral membrane protein	32.7%
HP0284	conserved hypothetical integral membrane protein	29.2%
HP0362	conserved hypothetical integral membrane protein	28.8%
HP0415	conserved hypothetical integral membrane protein	44.4%
HP0467	conserved hypothetical integral membrane protein	100.0%
HP0571	conserved hypothetical integral membrane protein	29.5%
HP0644	conserved hypothetical integral membrane protein	30.3%
HP0677	conserved hypothetical integral membrane protein	28.5%
HP0693	conserved hypothetical integral membrane protein	46.7%
HP0718	conserved hypothetical integral membrane protein	33.5%
HP0737	conserved hypothetical integral membrane protein	33.3%
HP0758	conserved hypothetical integral membrane protein	47.6%
HP0759	conserved hypothetical integral membrane protein	31.1%
HP0787	conserved hypothetical integral membrane protein	25.2%
HP0851	conserved hypothetical integral membrane protein	37.3%
HP0920	conserved hypothetical integral membrane protein	36.3%
HP0946	conserved hypothetical integral membrane protein	35.9%
HP0952	conserved hypothetical integral membrane protein	38.5%
HP0983	conserved hypothetical integral membrane protein	32.8%
HP1044	conserved hypothetical integral membrane protein	30.6%
HP1061	conserved hypothetical integral membrane protein	35.0%
HP1080	conserved hypothetical integral membrane protein	44.0%
HP1162	conserved hypothetical integral membrane protein	27.6%
HP1175	conserved hypothetical integral membrane protein	40.6%
HP1184	conserved hypothetical integral membrane protein	23.5%
HP1185	conserved hypothetical integral membrane protein	55.5%
HP1225	conserved hypothetical integral membrane protein	31.6%
HP1234	conserved hypothetical integral membrane protein	29.0%
HP1235	conserved hypothetical integral membrane protein	30.9%
HP1330	conserved hypothetical integral membrane protein	41.7%
HP1331	conserved hypothetical integral membrane protein	33.6%
HP1343	conserved hypothetical integral membrane protein	49.1%
HP1363	conserved hypothetical integral membrane protein	33.1%
HP1407	conserved hypothetical integral membrane protein	22.4%
HP1466	conserved hypothetical integral membrane protein	30.9%
HP1484	conserved hypothetical integral membrane protein	41.2%
HP1486	conserved hypothetical integral membrane protein	23.8%
HP1487	conserved hypothetical integral membrane protein	30.7%
HP1509	conserved hypothetical integral membrane protein	34.3%
HP1548	conserved hypothetical integral membrane protein	30.6%
HP0138	conserved hypothetical iron-sulfur protein	41.2%
HP1438	conserved hypothetical lipoprotein	32.0%
HP0151	conserved hypothetical membrane protein	21.8%
HP0575	conserved hypothetical membrane protein	38.8%
HP1258	conserved hypothetical mitochondrial protein 4	23.2%
HP1492	conserved hypothetical nifU-like protein	48.2%
HP0032	conserved hypothetical protein	37.0%
HP0035	conserved hypothetical protein	34.1%
HP0086	conserved hypothetical protein	28.7%
HP0094	conserved hypothetical protein	29.8%
HP0100	conserved hypothetical protein	32.0%
HP0102	conserved hypothetical protein	29.3%
HP0105	conserved hypothetical protein	39.7%
HP0117	conserved hypothetical protein	34.2%
HP0162	conserved hypothetical protein	36.7%
HP0216	conserved hypothetical protein	33.9%
HP0233	conserved hypothetical protein	30.5%
HP0248	conserved hypothetical protein	30.7%
HP0274	conserved hypothetical protein	38.5%
HP0285	conserved hypothetical protein	30.8%
HP0309	conserved hypothetical protein	31.3%
HP0310	conserved hypothetical protein	33.7%
HP0318	conserved hypothetical protein	47.2%
HP0328	conserved hypothetical protein	30.7%
HP0334	conserved hypothetical protein	30.8%
HP0347	conserved hypothetical protein	31.8%
HP0373	conserved hypothetical protein	31.4%
HP0374	conserved hypothetical protein	24.7%
HP0388	conserved hypothetical protein	39.8%
HP0395	conserved hypothetical protein	39.9%
HP0396	conserved hypothetical protein	33.7%
HP0419	conserved hypothetical protein	45.6%
HP0447	conserved hypothetical protein	38.2%
HP0465	conserved hypothetical protein	95.5%
HP0466	conserved hypothetical protein	95.7%
HP0468	conserved hypothetical protein	97.1%
HP0469	conserved hypothetical protein	95.1%
HP0496	conserved hypothetical protein	99.2%
HP0507	conserved hypothetical protein	37.2%
HP0519	conserved hypothetical protein	95.3%
HP0552	conserved hypothetical protein	37.8%
HP0553	conserved hypothetical protein	30.0%
HP0639	conserved hypothetical protein	41.0%
HP0654	conserved hypothetical protein	32.0%
HP0656	conserved hypothetical protein	36.0%
HP0707	conserved hypothetical protein	40.1%
HP0709	conserved hypothetical protein	49.6%
HP0710	conserved hypothetical protein	33.7%
HP0716	conserved hypothetical protein	30.2%
HP0728	conserved hypothetical protein	29.3%
HP0734	conserved hypothetical protein	31.0%
HP0741	conserved hypothetical protein	30.2%
HP0745	conserved hypothetical protein	33.7%
HP0747	conserved hypothetical protein	32.4%
HP0760	conserved hypothetical protein	36.1%
HP0810	conserved hypothetical protein	31.0%
HP0813	conserved hypothetical protein	32.5%
HP0823	conserved hypothetical protein	27.8%
HP0860	conserved hypothetical protein	52.1%
HP0890	conserved hypothetical protein	32.2%
HP0891	conserved hypothetical protein	33.8%
HP0892	conserved hypothetical protein	39.1%
HP0894	conserved hypothetical protein	39.8%
HP0926	conserved hypothetical protein	30.7%
HP0934	conserved hypothetical protein	33.6%
HP0956	conserved hypothetical protein	36.2%
HP0959	conserved hypothetical protein	31.1%
HP0966	conserved hypothetical protein	29.1%
HP0975	conserved hypothetical protein	25.0%
HP1020	conserved hypothetical protein	31.5%
HP1037	conserved hypothetical protein	95.9%
HP1046	conserved hypothetical protein	32.6%
HP1049	conserved hypothetical protein	39.7%
HP1066	conserved hypothetical protein	41.3%
HP1149	conserved hypothetical protein	24.7%
HP1160	conserved hypothetical protein	34.7%
HP1182	conserved hypothetical protein	34.6%
HP1214	conserved hypothetical protein	21.5%
HP1221	conserved hypothetical protein	42.4%
HP1240	conserved hypothetical protein	22.5%
HP1242	conserved hypothetical protein	42.3%
HP1259	conserved hypothetical protein	44.6%
HP1284	conserved hypothetical protein	36.8%
HP1291	conserved hypothetical protein	26.3%
HP1335	conserved hypothetical protein	33.9%
HP1337	conserved hypothetical protein	27.2%
HP1338	conserved hypothetical protein	36.2%
HP1394	conserved hypothetical protein	33.6%
HP1401	conserved hypothetical protein	27.5%
HP1413	conserved hypothetical protein	41.6%
HP1414	conserved hypothetical protein	27.4%
HP1417	conserved hypothetical protein	23.7%
HP1423	conserved hypothetical protein	40.3%
HP1426	conserved hypothetical protein	40.0%
HP1428	conserved hypothetical protein	37.8%
HP1443	conserved hypothetical protein	37.9%
HP1449	conserved hypothetical protein	39.0%
HP1453	conserved hypothetical protein	26.8%
HP1459	conserved hypothetical protein	30.1%
HP1504	conserved hypothetical protein	23.9%
HP1510	conserved hypothetical protein	30.6%
HP1533	conserved hypothetical protein	25.4%
HP1570	conserved hypothetical protein	40.5%
HP1573	conserved hypothetical protein	42.2%
HP1587	conserved hypothetical protein	39.0%
HP1588	conserved hypothetical protein	32.0%
HP1589	conserved hypothetical protein	35.1%
HP0713	conserved hypothetical protein (plasmid pHPM180)	41.8%
HP0028	conserved hypothetical secreted protein	42.1%
HP0139	conserved hypothetical secreted protein	37.1%
HP0160	conserved hypothetical secreted protein	30.6%
HP0190	conserved hypothetical secreted protein	31.4%
HP0211	conserved hypothetical secreted protein	24.3%
HP0235	conserved hypothetical secreted protein	31.5%
HP0257	conserved hypothetical secreted protein	29.2%
HP0320	conserved hypothetical secreted protein	36.4%
HP0506	conserved hypothetical secreted protein	29.8%
HP0518	conserved hypothetical secreted protein	96.9%
HP0785	conserved hypothetical secreted protein	26.6%
HP0949	conserved hypothetical secreted protein	39.7%
HP0977	conserved hypothetical secreted protein	29.4%
HP0980	conserved hypothetical secreted protein	57.4%
HP1075	conserved hypothetical secreted protein	42.9%
HP1098	conserved hypothetical secreted protein	27.0%
HP1117	conserved hypothetical secreted protein	32.3%
HP1216	conserved hypothetical secreted protein	31.9%
HP1285	conserved hypothetical secreted protein	38.0%
HP1286	conserved hypothetical secreted protein	37.5%
HP1464	conserved hypothetical secreted protein	27.4%
HP1488	conserved hypothetical secreted protein	29.8%
HP1551	conserved hypothetical secreted protein	42.7%

UNKNOWN

General

HP0390	adhesin-thiol peroxidase (tagD)	38.3%
HP1193	aldo-keto reductase, putative	46.6%
HP0872	alkylphosphonate uptake protein (phnA)	61.1%
HP0207	ATP-binding protein (mpr)	38.9%
HP0136	bacterioferritin comigratory protein (bcp)	35.5%
HP0485	catalase-like protein	30.8%
HP1104	cinnamyl-alcohol dehydrogenase ELI3-2 (cad)	44.0%
HP0981	exonuclease VII-like protein (xseA)	42.5%
HP0569	GTP-binding protein (gtp1)	48.1%
HP0303	GTP-binding protein (obg)	48.2%
HP0834	GTP-binding protein homologue (yphC)	36.7%
HP0480	GTP-binding protein, fusA-homolog (yihK)	54.1%
HP1489	lipase-like protein	21.7%
HP0405	nifS-like protein	27.3%
HP0221	nifU-like protein	37.3%
HP0658	PET112-like protein	45.4%
HP0089	pfs protein (pfs)	34.5%
HP0322	poly E-rich protein	28.7%
HP0625	protein E (gcpE)	47.7%
HP0431	protein phosphatase 2C homolog (ptc1)	30.7%
HP0624	solute-binding signature and mitochondrial signature protein (aspB)	26.4%
HP0377	thiol:disulfide interchange protein (dsbC), putative	26.4%