

17. Fülöp, V., Moir, J. W. B., Ferguson, S. J. & Hajdu, J. Crystallisation and preliminary crystallographic study of cytochrome *cd*, 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 NO₂ 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.E.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 C₁₂₀

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 (C₆₀)₂²⁻ dianions from the X-ray powder diffraction patterns of the metastable phases of KC₆₀ and RbC₆₀. □

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. Fitzgerald, 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 discovery¹ 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 regimens².

Helicobacter pylori is a micro-aerophilic, Gram-negative, slow-growing, spiral-shaped and flagellated organism. Its most characteristic enzyme is a potent multisubunit urease³ 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 colonize². *H. pylori* is probably the most common chronic bacterial infection of humans, present in almost half of the world population². 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 lymphomas². 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 diseases⁶.

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

Table 1. Genomic 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; <i>virB4</i>
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 (<i>fusA</i>)
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 transfer¹¹.

RNA and repeat elements. Thirty-six tRNA species were identified using tRNAscan-SE¹². 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 identified¹⁴, 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 Riley¹⁵ (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 (pI) 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.

Paralogous families. Ninety-five paralogous gene families comprising 266 gene products (16% of the total) were identified (www.tigr.org/tdb/mdb/hpdbh/hpdbh.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 complex¹⁶. 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 island¹¹ and the other is the flagellar export pathway which is assembled from orthologues of FliH, FliI, FliP, FliB, FliQ, FliR and FliP¹⁷. 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 structures¹⁸ and assembly of surface structures involved in virulence and DNA transformation¹⁹.

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 recombination²⁰, 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: ++, Co²⁺, Zn²⁺, Cd²⁺; ?, 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 TopPred⁴⁷.

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 glutamyl-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 glutamyl-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-tRNA^{Asn}* from *Asp-tRNA^{Asn}* 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).

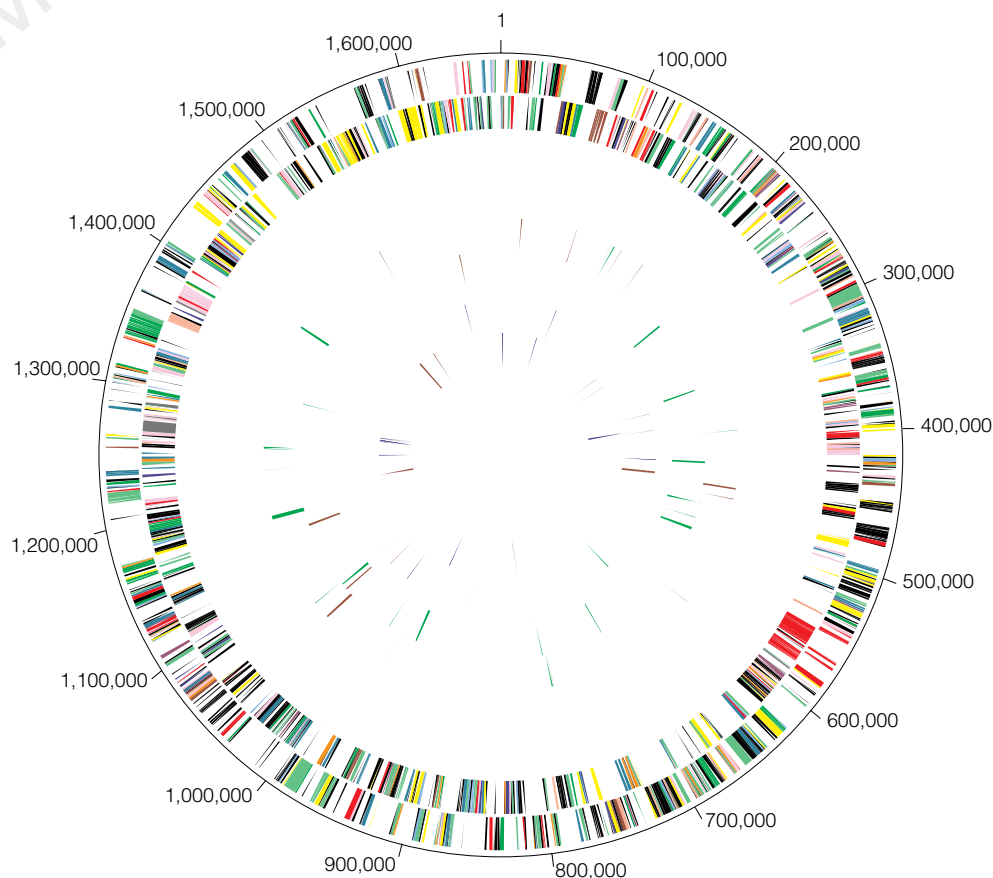
Adhesion and adaptive antigenic variation

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 cells⁵. 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.



HopB EN—DQVPSVQYQIQ—AVQV
HopB EE—DQGVPSVQYQIQ—AVQV
HopC ED—DQGVPSVQYQIQ—AVQV
HopD ED—DQGVPSVQYQIQ—AVQV
HopE EG—DQGVPSVQYQIQ—AVQV

n-Terminal peptide sequences:

HopB EN—DQVPSVQYQIQ—AVQV
HopB EE—DQGVPSVQYQIQ—AVQV
HopC ED—DQGVPSVQYQIQ—AVQV
HopD ED—DQGVPSVQYQIQ—AVQV
HopE EG—DQGVPSVQYQIQ—AVQV

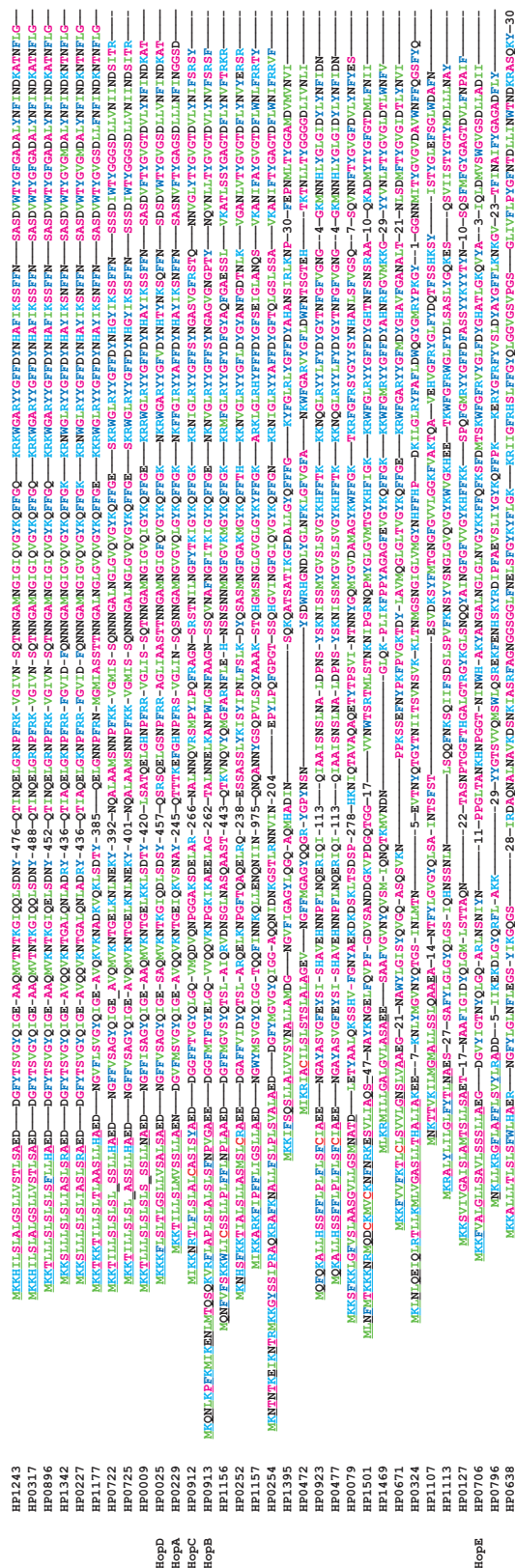
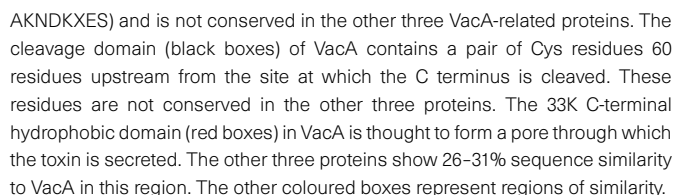
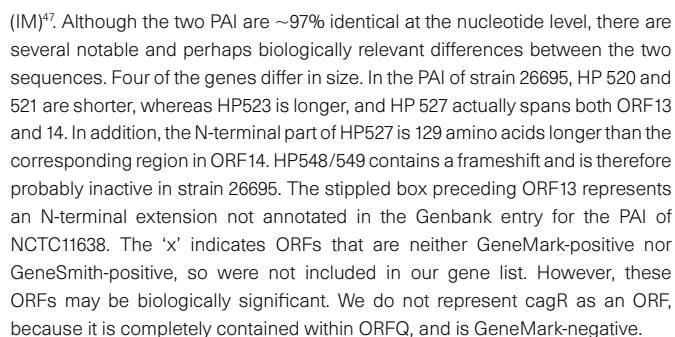


Figure 3 Multiple sequence alignment of members of the outer membrane protein family of *H. pylori*. These proteins were identified as OMPs based on the characteristic alternating hydrophobic residues at their carboxy termini. All members of this family have one domain of similarity at the amino-terminal 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 alignment⁵⁰. The most likely

candidate for HopD is HP0913, which has 15 matches to the first 20-residue N-terminal peptide sequence⁵⁰. These differences may be due to strain variability. The program Signal-P⁴⁸ 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 frameshifts are associated with the presence of dinucleotide repeats (Table 3).

VacA induces the formation of acidic vacuoles in host epithelial cells, and its presence is associated epidemiologically with tissue damage and disease³¹. 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 regions are noted among VacA proteins derived from Tox⁺ and Tox⁻ strains³¹. 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



As with other pathogens, *H. pylori* probably requires an iron-scavenging system for survival in the host⁵. 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 bacterioferritin³⁴, and Pfr, a non-haem cytoplasmic iron-containing ferritin used for storage of iron³⁵. The global ferric uptake regulator (Fur) characterized in other bacteria is also present in *H. pylori*. Consensus

H. pylori motility is essential for colonization³⁶. 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 been 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 regulon⁵.

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 potential³⁷ and subsequently to modify its microenvironment through the action of urease and the release of factors that inhibit acid production by parietal cells⁵. 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 cytoplasm³⁸, this idea is not supported by the identified transporter

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

The designated monomers used to synthesize the oligonucleotides, and the oligonucleotide sequence of the oligonucleotide are as follows:
 CCAGAAATCTCTTTTTTTTTTTGAAATCCCAATTAATGGTAAAG-37bp-1TTACATAAAAAAATTAATCTTTAAGGACATTT
TATGAAAAAGACAATCTCTCTCTCTCTCTCTCTCGCTTCATCGCTCTTGACGCTGAAGACAACGGCTTTTTGTGAGCGCCGGCT
 Y E K D N S T 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 T L S S S S H R S C T T K T T A F I *

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

HP no.	ID	Tract	HP no.	ID	Tract	HP no.	ID	Tract
9	OMP	A14	25	OMP	T15	208	<i>rfaJ</i>	A11
227	OMP	T14	228	IMP	A14	349	<i>pyrG</i>	T15
350	IMP	A15	547	<i>cagA</i>	A14	629	Hypo	T15
722	OMP	T16	725	OMP	T14	733	Hypo	T13
876	<i>frpB</i>	T16	896	OMP	A14	912	OMP	T13
1342	OMP	A14	1400	<i>fecA</i>	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 activity³⁹. 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 response⁴⁰. 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

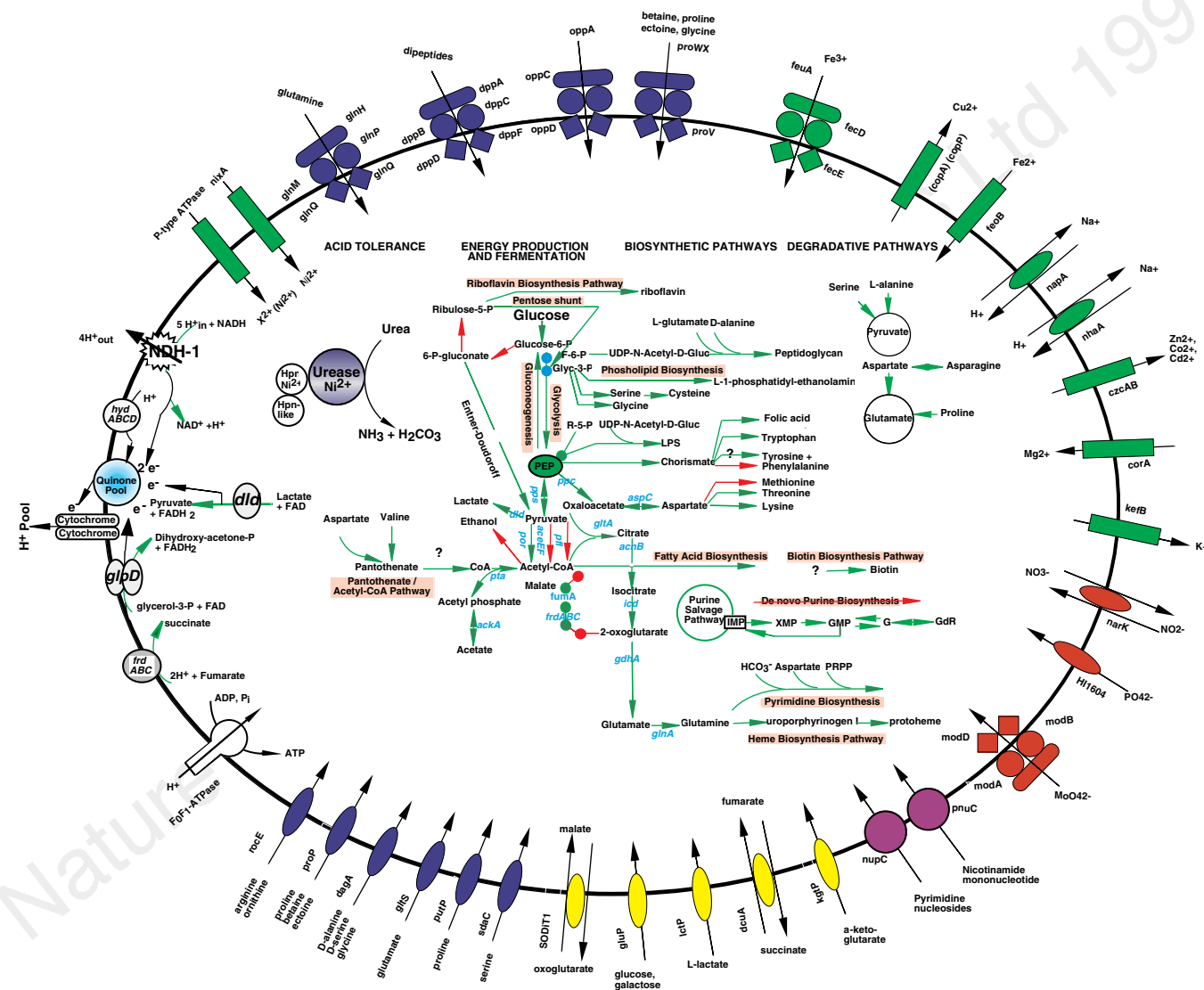


Figure 6 Solute transport and metabolic pathways of *Helicobacter pylori*. Transporters identified by sequence comparisons are characteristic of Gram-negative bacteria. Colours correspond to transport role categories defined by Riley¹⁵: blue, amino acids, peptides and amines; red, anions; yellow, carbohydrates, 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-} (*H1604*), NO_3^- (*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 (*dppABCD*), proline (*proWX*), 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 ferredoxin oxidoreductase (POR), a four-subunit enzyme thus far only described in hyperthermophilic organisms⁴¹. 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 dehydrogenases 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 cerevisiae*. 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 explored⁴³. 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 GeneMark⁴⁵ 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 described⁹. ORFs were also analysed using 175 hidden Markov models constructed for a number of conserved protein families (pfam v1.0) using hmmer⁴³. In addition, all ORFs were searched against the prosite motif database using MacPattern⁴⁶. 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 TopPred⁴⁷ 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 bacteria⁴⁸. 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 EcoCyc⁴⁹.

Received 16 May; accepted 1 July 1997.

- Warren, J. R. & Marshall, B. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. *Lancet* **1**, 1273–1275 (1983).
- 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).
- Mobley, H. L. T., Island, M. D. & Hausinger, R. P. Molecular Biology of Microbial Ureases. *Microbiol. Rev.* **59**, 451–480 (1995).
- Go, M. F. & Graham, D. Y. How does *Helicobacter pylori* cause duodenal ulcer disease: The bug, the host, or both? *J. Gastroenterol. Hepatol.* (suppl.) **9**, 8–12 (1994).
- Labigne, A. & de Reuse, H. Determinants of *Helicobacter pylori* pathogenicity. *Infect. Agents Disease* **5**, 191–202 (1996).
- 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).
- Fleischmann, R. D. et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512 (1995).
- Fraser, C. M. et al. The *Mycoplasma genitalium* genome sequence reveals a minimal gene complement. *Science* **270**, 397–403 (1995).
- Bult, C. J. et al. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058–1073 (1996).
- 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).
- Censini, S. et al. Cag, a pathogenicity island of *Helicobacter pylori*, encodes type-I-specific and disease-associated virulence factors. *Proc. Natl Acad. Sci. USA* **93**, 14648–14653 (1996).
- <http://genome.wustl.edu/eddy/low/rNAScan-SE-Manual/Manual.html>
- 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).
- Marczynski, G. T. & Shapiro, L. Bacterial chromosome origins of replication. *Curr. Opin. Gen. Dev.* **3**, 775–782 (1993).
- Riley, M. Functions of gene products of *Escherichia coli*. *Microbiol. Rev.* **57**, 862–952 (1993).
- Kornberg, A. & Baker, T. A. Replication mechanisms and operations in DNA replication. (ed. Kornberg, A. & Baker, T.) 471–510 (Freeman, New York, 1992).

- Macnab, R. M. in *Escherichia coli and Salmonella Cellular and Molecular Biology* (eds Neidhardt, F. C. et al.) 123–145 (ASM, Washington DC, 1996).
- 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).
- Bardwell, J. C. Building bridges: disulphide bond formation in the cell. *Mol. Microbiol.* **14**, 199–205 (1994).
- Linn, S. in *Escherichia coli and Salmonella Cellular and Molecular Biology* (eds Neidhardt, F. C. et al.) 764–772 (ASM, Washington D.C., 1996).
- 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).
- Curnow, A. W., Ibba, M. & Soll, D. tRNA-dependent asparagine formation. *Nature* **382**, 589–590 (1996).
- 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).
- 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).
- Ilyer, D. et al. The *Helicobacter pylori* blood group antigen binding adhesin. *Gut* **39** (suppl. 2), A55 (1996).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- Moran, A. P. The role of lipopolysaccharide in *Helicobacter pylori* pathogenesis. *Aliment. Pharmacol. Ther.* **10** (suppl. 1), 39–50 (1996).
- 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).
- Earhart, C. F. in *Escherichia coli and Salmonella Cellular and Molecular Biology* (eds Neidhardt, F. C. et al.) 1075–1090 (ASM, Washington DC, 1996).
- Evans, D. J. Jr, Evans, D. G., Lampert, H. C. & Nakano, H. Identification of four new prokaryotic bacterioferritins, from *Helicobacter pylori*, *Anabaena variabilis*, *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).
- Suerbaum, S. The complex flagella of gastric *Helicobacter* species. *Trends Microbiol.* **3**, 168–170 (1995).
- 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).
- Melchers, K. et al. Cloning and membrane topology of a P type ATPase from *Helicobacter pylori*. *J. Biol. Chem.* **271**, 446–457 (1996).
- 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).
- McGowan, C. C., Cover, T. L. & Blaser, M. J. *Helicobacter pylori* and gastric acid: biological and therapeutic implications. *Gastroenterology* **110**, 926–938 (1996).
- 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).
- Mendz, G. L. & Hazell, S. L. Amino acid utilization by *Helicobacter pylori*. *Int. J. Biochem. Cell. Biol.* **27**, 1085–1093 (1995).
- Sonnhammer, E. L. L., Eddy, S. R. & Durbin, R. Pfam: A comprehensive database of protein families based on seed alignments. *Proteins* (in the press).
- 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).
- 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).
- Fuchs, R. MacPattern: protein pattern searching on the Apple Macintosh. *Comput. Appl. Biosci.* **7**, 105–106 (1991).
- Claros, M. G. & von Heijne, G. TopPred II: an improved software for membrane protein structure predictions. *Comput. Appl. Biosci.* **10**, 685–686 (1994).
- 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).
- 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).
- 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/hpdbh/hpdbh.html>. The sequence has been deposited with GenBank under accession number AF000511.

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			CELL ENVELOPE		
<i>General</i>			<i>Membranes, lipoproteins and porins</i>		
HP0695	hydantoin utilization protein A (hyuA)	28.6%	HP0841	pantothenate metabolism flavoprotein (dfp)	31.3%
<i>Aromatic amino-acid family</i>			<i>Pyridoxine</i>		
HP1038	3-dehydroquinate type II (aroC)	99.4%	HP1583	pyridoxal phosphate biosynthetic protein A (pdxA)	34.2%
HP0283	3-dehydroquinate synthase (aroB)	38.1%	HP1582	pyridoxal phosphate biosynthetic protein J (pdxJ)	42.6%
HP0134	3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (dhsl)	54.6%	<i>Riboflavin</i>		
HP0401	3-phosphoshikimate 1-carboxyvinyltransferase (aroA)	53.6%	HP0802	GTP cyclohydrolase II (ribA)	47.2%
HP1279	anthranilate isomerase (trpC)	47.0%	HP0804	GTP cyclohydrolase II/3,4-dihydroxy-2-butanone 4-phosphate synthase (ribA, ribB)	44.0%
HP1282	anthranilate synthase component I (trpE)	47.9%	HP1505	riboflavin biosynthesis protein (ribG)	33.1%
HP1280	anthranilate synthase component II (trpD)	42.5%	HP1087	riboflavin biosynthesis regulatory protein (ribC)	28.9%
HP1281	anthranilate synthase component II (trpD)	40.2%	HP1574	riboflavin synthase alpha subunit (ribC)	32.8%
HP0663	chorismate synthase (aroC)	47.2%	HP0002	riboflavin synthase beta chain (ribE)	52.4%
HP1380	prephenate dehydrogenase (tyrA)	30.2%	<i>Thioresoxin, glutaredoxin and glutathione</i>		
HP1249	shikimate 5-dehydrogenase (aroE)	36.6%	HP1118	gamma-glutamyltranspeptidase (ggT)	53.2%
HP0157	shikimate kinase I (aroK)	36.1%	HP1458	thioredoxin	38.3%
HP1277	tryptophan synthase, alpha subunit (trpA)	46.5%	HP0824	thioredoxin (trxA)	51.5%
HP1278	tryptophan synthase, beta subunit (trpB)	66.1%	HP1164	thioredoxin reductase (trxB)	28.5%
<i>Aspartate family</i>			<i>Thiamine</i>		
HP0649	aspartate ammonia-lyase (aspA)	55.5%	HP0814	thiamin biosynthesis protein (thiF)	34.6%
HP1189	aspartate-semialdehyde dehydrogenase (aspC)	45.7%	HP0843	hydroxyethylthiazole kinase (thiB)	35.7%
HP1229	aspartokinase (lysC)	48.0%	HP0845	thiamin phosphate pyrophosphorylase/hydroxyethylthiazole kinase (thiM)	37.9%
HP0106	cystathionine gamma-synthase (metB)	47.7%	HP0844	thiamine biosynthesis protein (thi)	41.0%
HP0290	diaminopimelate decarboxylase (dap decarboxylase) (lysA)	42.7%	<i>Pyridine nucleotides</i>		
HP0666	diaminopimelate epimerase (dapF)	30.0%	HP0329	NH(3)-dependent NAD+ synthetase (nadE)	37.5%
HP0610	dihydrodipicolinate reductase (dapB)	95.3%	HP1355	nicotinate-nucleotide pyrophosphorylase (nadC)	36.3%
HP1013	dihydrodipicolinate synthetase (dapA)	39.5%	HP1356	quinolinate synthetase A (nadA)	34.2%
HP0822	homoserine dehydrogenase (metL)	37.7%	CELL ENVELOPE		
HP0650	homoserine kinase (thrB)	27.7%	<i>Membranes, lipoproteins and porins</i>		
HP0672	solute-binding signature and mitochondrial signature protein (aspB)	47.3%	HP1450	60 kDa inner-membrane protein	40.0%
HP0212	succinyl-diaminopimelate desuccinylase (dapE)	42.3%	HP0180	apolipoprotein N-acyltransferase (cute)	28.0%
HP0626	tetrahydrodipicolinate N-succinyltransferase (dapD)	36.1%	HP0175	cell binding factor 2	34.9%
HP0098	threonine synthase (thrC)	32.9%	HP0078	Hypothetical protein	28.4%
<i>Glutamate family</i>			HP0567	membrane protein	26.4%
HP0380	glutamate dehydrogenase (gdhA)	58.0%	HP1456	membrane-associated lipoprotein (lpp20)	98.9%
HP0512	glutamine synthetase (glnA)	48.6%	HP1564	outer membrane protein	39.9%
HP1158	pyroline-5-carboxylate reductase (proC)	28.9%	HP0009	outer membrane protein (omp1)	0.0%
<i>Pyruvate family</i>			HP0324	outer membrane protein (omp10)	0.0%
HP0841	alanine racemase, biosynthetic (alr)	32.4%	HP0472	outer membrane protein (omp11)	99.5%
HP1468	branched-chain-amino-acid aminotransferase (ilvE)	63.5%	HP0477	outer membrane protein (omp12)	0.0%
HP0330	ketol-acid reductoisomerase (ilvC)	48.1%	HP0638	outer membrane protein (omp13)	0.0%
<i>Serine family</i>			HP0671	outer membrane protein (omp14)	36.0%
HP0107	cysteine synthetase (cysK)	45.7%	HP0706	outer membrane protein (omp15)	33.5%
HP0096	phosphoglycerate dehydrogenase	31.0%	HP0722	outer membrane protein (omp16)	43.3%
HP0397	phosphoglycerate dehydrogenase (serA)	32.5%	HP0725	outer membrane protein (omp17)	43.3%
HP0736	phosphoserine aminotransferase (serC)	30.7%	HP0798	outer membrane protein (omp18)	0.0%
HP0652	phosphoserine phosphatase (serB)	36.5%	HP0896	outer membrane protein (omp19)	36.6%
HP1210	serine acetyltransferase (cysE)	96.2%	HP0025	outer membrane protein (omp2)	0.0%
HP1033	serine hydroxymethyltransferase (glyA)	54.0%	HP0912	outer membrane protein (omp20)	0.0%
BIOSYNTHESIS OF COFACTORS, PROSTHETIC GROUPS, AND CARRIERS			HP0913	outer membrane protein (omp21)	38.2%
<i>General</i>			HP0923	outer membrane protein (omp22)	0.0%
HP0220	synthesis of [Fe-S] cluster (nifS)	48.0%	HP1107	outer membrane protein (omp23)	0.0%
<i>Biotin</i>			HP1113	outer membrane protein (omp24)	36.0%
HP0598	8-amino-7-oxononanoate synthase (bioF)	34.9%	HP1156	outer membrane protein (omp25)	0.0%
HP0976	adenosylmethionine-8-amino-7-oxononanoate aminotransferase (bioA)	49.2%	HP1157	outer membrane protein (omp26)	23.0%
HP1140	biotin operon repressor/biotin acetyl coenzyme A carboxylase synthetase (birA)	36.9%	HP1177	outer membrane protein (omp27)	37.0%
HP0407	biotin sulfoxide reductase (bioS)	42.7%	HP1243	outer membrane protein (omp28)	0.0%
HP1254	biotin synthetase protein (bioB)	32.1%	HP1342	outer membrane protein (omp29)	0.0%
HP1406	biotin synthetase (bioB)	36.2%	HP0079	outer membrane protein (omp3)	0.0%
HP0029	dethiobiotin synthetase (bioD)	36.0%	HP1395	outer membrane protein (omp30)	0.0%
<i>Folic acid</i>			HP1469	outer membrane protein (omp31)	0.0%
HP1036	7, 8-dihydro-6-hydroxymethylpterin-pyrophosphokinase (folK)	34.6%	HP1501	outer membrane protein (omp32)	0.0%
HP0587	aminodeoxychorismate lyase (pabC)	32.4%	HP0127	outer membrane protein (omp4)	0.0%
HP1232	dihydrodipicolinate synthase (folP)	34.5%	HP0227	outer membrane protein (omp5)	36.8%
HP1545	folypolyglutamate synthase (folC)	35.2%	HP0229	outer membrane protein (omp6)	38.4%
HP0928	GTP cyclohydrolase I (folE)	50.9%	HP0252	outer membrane protein (omp7)	30.6%
HP0577	methylene-tetrahydrofolate dehydrogenase (folD)	48.4%	HP0254	outer membrane protein (omp8)	37.6%
HP0293	para-aminobenzoate synthetase (pabB)	35.1%	HP0317	outer membrane protein (omp9)	36.3%
<i>Haem and porphyrin</i>			HP0339	outer membrane protein P1 (ompP1)	23.3%
HP0163	delta-aminolevulinic acid dehydratase (hemB)	50.5%	HP0955	prolipoprotein diacylglycerol transferase (lgt)	34.4%
HP0376	ferrochelatase (hemH)	33.4%	HP0655	protective surface antigen D15	27.5%
HP0306	glutamate-1-semialdehyde 2,1-aminomutase (hemL)	51.3%	HP1571	rare lipoprotein A (ripA)	37.6%
HP0239	glutamyl-tRNA reductase (hemA)	32.7%	HP0610	toxin-like outer membrane protein	26.3%
HP0665	oxygen-independent coproporphyrinogen III oxidase (hemN)	42.4%	HP0922	toxin-like outer membrane protein	29.5%
HP1228	oxygen-independent coproporphyrinogen III oxidase (hemN)	37.9%	HP0289	toxin-like outer membrane protein	30.6%
HP0237	porphobilinogen deaminase (hemC)	45.7%	<i>Murein sacculus and peptidoglycan</i>		
HP0381	protoporphyrinogen oxidase (hemK)	35.9%	HP0830	amidase	40.6%
HP0604	uroporphyrinogen decarboxylase (hemE)	46.3%	HP0738	D-alanine-D-alanine ligase A (ddlA)	28.5%
HP1224	uroporphyrinogen III cosynthase (hemD)	27.6%	HP0549	glutamate racemase (glr)	36.6%
<i>Menquinone and ubiquinone</i>			HP0772	N-acetylmuramoyl-L-alanine amidase (amiA)	26.8%
HP1360	4-hydroxybenzoate octaprenyltransferase (ubiA)	26.6%	HP0697	penicillin-binding protein 1A (PBP-1A)	33.7%
HP0929	geranyltransferase (ispA)	39.8%	HP1565	penicillin-binding protein 2 (pbp2)	35.0%
HP0240	octaprenyl-diphosphate synthase (ispB)	31.6%	HP1125	peptidoglycan associated lipoprotein precursor (omp18)	42.6%
<i>Molybdopterins</i>			HP0493	phospho-N-acetylmuramoyl-pentapeptide-transferase (mraY)	45.2%
HP0768	molybdenum cofactor biosynthesis protein A (moaA)	31.4%	HP0743	rod shape-determining protein (mreB)	37.7%
HP0798	molybdenum cofactor biosynthesis protein C (moaC)	97.9%	HP1373	rod shape-determining protein (mreB)	51.9%
HP0172	molybdopterins biosynthesis protein (moaE)	36.3%	HP1372	rod shape-determining protein (mreC)	33.6%
HP0756	molybdopterins biosynthesis protein (moaB)	32.2%	HP0645	soluble lytic murein transglycosylase (slt)	32.2%
HP0799	molybdopterins biosynthesis protein (mog)	50.8%	HP1543	toxR-activated gene (tagE)	37.2%
HP0801	molybdopterins converting factor, subunit 1 (moaD)	31.1%	HP1544	toxR-activated gene (tagE)	31.2%
HP0800	molybdopterins converting factor, subunit 2 (moaE)	31.1%	HP1155	transferase, peptidoglycan synthesis (murG)	28.2%
HP0769	molybdopterins-guanine dinucleotide biosynthesis protein A (mobA)	28.3%	HP0740	UDP-MurNac-pentapeptide presynthetase (murF)	25.7%
<i>Pantothenate</i>			HP1494	UDP-MurNac-tripeptide synthetase (murE)	36.0%
HP1058	3-methyl-2-oxobutanate hydroxymethyltransferase (panB)	43.7%	HP1418	UDP-N-acetylenolpyruvylglucosamine reductase (murB)	32.7%
HP0034	aspartate 1-decarboxylase (panD)	50.0%	HP0648	UDP-N-acetylglucosamine enolpyruvyl transferase (murZ)	46.7%
HP0006	pantoate-beta-alanine ligase (panC)	44.2%	HP0623	UDP-N-acetylmuramate-alanine ligase (murC)	37.3%
CELLULAR PROCESSES			HP0494	UDP-N-acetylmuramoylalanine-D-glutamate ligase (murD)	31.1%
<i>General</i>			<i>Surface polysaccharides, lipopolysaccharides and antigens</i>		
HP0019	chemotaxis protein (cheV)	26.8%	HP0003	3-deoxy-D-manno-oculosonic acid 8-phosphate synthetase (kdsA)	53.4%
HP0393	chemotaxis protein (cheV)	31.7%	HP0957	3-deoxy-D-manno-oculosonic-acid transferase (kdsA)	35.9%
HP0616	chemotaxis protein (cheV)	27.9%	HP0858	ADP-heptose synthase (rfcA)	40.6%
HP1067	chemotaxis protein (cheV)	99.2%	HP1191	ADP-heptose-1ps heptosyltransferase II (rfcA)	33.2%
HP0517	GTP-binding protein (era)	95.6%	HP0859	ADP-L-glycero-D-mannoheptose-6-epimerase (rfcD)	32.7%
HP1490	haemolysin	39.2%	CELLULAR PROCESSES		
HP1086	haemolysin (tyl)	40.2%	<i>General</i>		
HP0599	haemolysin secretion protein precursor (hylB)	45.4%	HP0019	chemotaxis protein (cheV)	26.8%
HP0392	histidine kinase (cheA)	41.4%	HP0393	chemotaxis protein (cheV)	31.7%
HP0039	methyl-accepting chemotaxis protein (tipA)	32.8%	HP0616	chemotaxis protein (cheV)	27.9%
HP1013	methyl-accepting chemotaxis protein (tipB)	30.7%	HP1067	chemotaxis protein (cheV)	99.2%
HP0082	methyl-accepting chemotaxis transducer (tipC)	28.2%	HP0517	GTP-binding protein (era)	95.6%
HP0391	purine-binding chemotaxis protein (cheW)	34.3%	HP1490	haemolysin	39.2%
<i>Cell division</i>			HP1086	haemolysin (tyl)	40.2%
HP0331	cell division inhibitor (minD)	50.2%	HP0599	haemolysin secretion protein precursor (hylB)	45.4%
HP0749	cell division membrane protein (ftsX)	25.7%	HP0392	histidine kinase (cheA)	41.4%
HP0978	cell division protein (ftsA) protein	31.9%	HP0039	methyl-accepting chemotaxis protein (tipA)	32.8%
HP0748	cell division protein (ftsE)	37.6%	HP1013	methyl-accepting chemotaxis protein (tipB)	30.7%
HP0286	cell division protein (ftsH)	41.2%	HP0082	methyl-accepting chemotaxis transducer (tipC)	28.2%
HP1089	cell division protein (ftsI)	98.6%	HP0391	purine-binding chemotaxis protein (cheW)	34.3%
HP1556	cell division protein (ftsL)	30.6%	<i>Cell division</i>		
HP1090	cell division protein (ftsK)	39.8%	HP0331	cell division inhibitor (minD)	50.2%
HP1560	cell division protein (ftsW) Escherichia coli	32.7%	HP0749	cell division membrane protein (ftsX)	25.7%
HP0763	cell division protein (ftsY)	46.6%	HP0978	cell division protein (ftsA) protein	31.9%

HP0332	cell division topological specificity factor (mifE)	33.8%	HP1270	subunit (NQO10)	-1.0%	HP1101	(devB)	29.2%
HP0979	cell division protein (ftsZ)	43.3%		NADH-ubiquinone oxidoreductase, NQO11 subunit (NQO11) (Paracoccus denitrificans)	42.6%		glucose-6-phosphate dehydrogenase (g6pD)	36.7%
HP1159	cell filamentation protein (fic)	63.2%	HP1271	NADH-ubiquinone oxidoreductase, NQO12 subunit (NQO12)	43.2%	HP1495	transaldolase (tal)	33.5%
Cell killing			HP1272	NADH-ubiquinone oxidoreductase, NQO13 subunit (NQO13)	40.2%	HP1088	transketolase A (tktA)	46.7%
HP0887	vacuolating cytotoxin	94.7%		NADH-ubiquinone oxidoreductase, NQO14 subunit (NQO14)	31.2%	HP0354	transketolase B (tktB)	39.7%
Chaperones								
HP0010	chaperone and heat shock protein (groEL)	99.6%	HP1273	NADH-ubiquinone oxidoreductase, NQO3 subunit (NQO3)	31.6%	Sugars		
HP0109	chaperone and heat shock protein 70 (dnaK)	63.4%	HP1266	NADH-ubiquinone oxidoreductase, NQO4 subunit (NQO4) (Triticum aestivum)	44.6%	HP0574	galactosidase acetyltransferase (lacA)	41.0%
HP0210	chaperone and heat shock protein C62.5 (htpG)	46.5%	HP1263	NADH-ubiquinone oxidoreductase, NQO5 subunit (NQO5)	-1.0%	HP0360	UDP-glucose 4-epimerase	43.1%
HP0011	co-chaperone (groES)	99.2%	HP1262	NADH-ubiquinone oxidoreductase, NQO6 subunit (NQO6)	62.2%	TCA cycle		
HP1332	co-chaperone and heat-shock protein (dnaJ)	42.7%	HP1261	NADH-ubiquinone oxidoreductase, NQO7 subunit (NQO7)	40.7%	HP0779	aconitase B (acnB)	64.0%
HP0110	co-chaperone and heat-shock protein (grpE)	33.0%	HP1260	NADH-ubiquinone oxidoreductase, NQO8 subunit (NQO8)	42.4%	HP0026	citrate synthase (gltA)	47.8%
HP1024	co-chaperone-curved DNA-binding protein A (CbpA)	37.7%	HP1268	NADH-ubiquinone oxidoreductase, NQO9 subunit (NQO9)	41.2%	HP1325	fumarase (fumC)	63.7%
Chromosome-associated protein								
HP1138	plasmid replication-partition related protein	40.4%	Amino acids and amines			HP0509	glycolate oxidase subunit (gldC)	98.0%
Detoxification								
HP1563	alkyl hydroperoxide reductase (tsaA)	98.5%	HP1398	alanine dehydrogenase (ald)	39.6%	HP0027	isocitrate dehydrogenase (icd)	70.7%
HP0875	catalase	99.4%	HP0294	aliphatic amidase (aimE)	75.4%	FATTY ACID AND PHOSPHOLIPID METABOLISM		
HP0267	chlorohydrolase	42.6%	HP1238	aliphatic amidase (aimE)	37.2%	HP1376	(3R)-hydroxymyristoyl-acyl carrier protein dehydratase (fabZ)	47.4%
HP0243	neutrophil activating protein (napA) (bacterioperitin)	95.8%	HP1399	arginase (rocF)	31.8%	HP1348	1-acyl-glycerol-3-phosphate acyltransferase (plsC) (Escherichia coli)	32.0%
HP0389	superoxide dismutase (sodB)	98.6%	HP0943	D-amino acid dehydrogenase (dadA)	26.2%	HP0561	3-ketoacyl-acyl carrier protein reductase (fabG)	45.7%
HP1452	thiophene and furan oxidizer (tdhF)	37.6%	HP0056	delta-1-pyrroline-5-carboxylate dehydrogenase (Synecocystis sp.)	32.2%	HP0690	acetyl coenzyme A acetyltransferase (thioase) (fadA)	52.0%
Protein and peptide secretion								
HP0355	GTP-binding membrane protein (lepA)	57.3%	HP0723	L-asparaginase II (ansB)	54.1%	HP0950	acetyl-CoA carboxylase beta subunit (accD)	49.4%
HP0074	lipoprotein signal peptidase (lspA)	97.0%	HP0132	L-serine deaminase (sdaA)	45.8%	HP0557	acetyl-CoA synthetase (accE)	52.3%
HP0786	preprotein translocase subunit (secA)	54.0%	Anaerobic			HP0559	acetyl-coenzyme A carboxylase (accA)	50.3%
HP1300	preprotein translocase subunit (secY)	41.2%	HP0666	anaerobic glycerol-3-phosphate dehydrogenase, subunit C (glpC)	27.2%	HP0962	acyl carrier protein (acpP)	55.3%
HP1255	protein translocation protein, low temperature (secG)	30.6%	HP0589	ferredoxin oxidoreductase, alpha subunit	42.7%	HP0568	beta ketoacyl-acyl carrier protein synthase II (fabF)	50.0%
HP1550	protein-export membrane protein (secD)	35.9%	HP0680	ferredoxin oxidoreductase, beta subunit	43.2%	HP0202	beta-ketoacyl-acyl carrier protein synthase III (fabH)	44.4%
HP1549	protein-export membrane protein (secE)	35.1%	HP0591	ferredoxin oxidoreductase, gamma subunit	33.3%	HP0371	biotin carboxyl carrier protein (fabE)	30.8%
HP0576	signal peptidase I (lepB)	40.3%	HP0193	fumarate reductase, cytochrome b subunit (frcD)	58.8%	HP0370	biotin carboxylase (accC)	52.1%
HP1152	signal recognition particle protein (fih)	41.4%	HP0192	fumarate reductase, flavoprotein subunit (frcA)	69.4%	HP0871	CDP-diglyceride hydrolase (cdh)	73.9%
HP0795	trigger factor (tig)	27.6%	HP0191	fumarate reductase, iron-sulfur subunit (frcB)	70.8%	HP0215	CDP-diglyceride synthetase (cdsA)	42.4%
Transformation								
HP0520	cag pathogenicity island protein (cagI)	96.5%	HP1110	pyruvate ferredoxin oxidoreductase, alpha subunit	41.0%	HP0416	cyclopropane fatty acid synthase (cfa)	39.7%
HP0530	cag pathogenicity island protein (cagII)	98.4%	HP1111	pyruvate ferredoxin oxidoreductase, beta subunit	43.7%	HP0700	dialcylglycerol kinase (dglkA)	45.8%
HP0531	cag pathogenicity island protein (cagIII)	97.2%	HP1109	pyruvate ferredoxin oxidoreductase, delta subunit	47.0%	HP0195	enoyl-acyl-carrier-protein reductase (NADH) (fabI)	45.8%
HP0532	cag pathogenicity island protein (cagI2)	98.9%	HP1108	pyruvate ferredoxin oxidoreductase, gamma subunit	37.2%	HP0201	fatty acid/phospholipid synthesis protein (plsX)	37.8%
HP0534	cag pathogenicity island protein (cagI3)	98.0%	ATP-protonmotive force interconversion			HP0808	Holo-acp synthase (acpS)	29.1%
HP0535	cag pathogenicity island protein (cagI4)	97.6%	HP0828	ATP synthase FO, subunit a (atpB)	37.7%	HP0090	malonyl coenzyme A-acyl carrier protein transacylase (fabD)	35.4%
HP0536	cag pathogenicity island protein (cagI5)	96.4%	HP1136	ATP synthase FO, subunit b (atpF)	28.3%	HP1016	phosphatidylglycerophosphate synthase (pgsA)	35.4%
HP0537	cag pathogenicity island protein (cagI6)	98.9%	HP1137	ATP synthase FO, subunit b (atpF)	32.5%	HP1357	phosphatidylserine decarboxylase proenzyme (psd)	33.2%
HP0538	cag pathogenicity island protein (cagI7)	95.3%	HP1212	ATP synthase FO, subunit c (atpE)	41.2%	HP1071	phosphatidylserine synthase (psaA)	99.6%
HP0539	cag pathogenicity island protein (cagI8)	98.7%	HP1134	ATP synthase F1, subunit alpha (atpA)	62.7%	HP0499	phospholipase A1 precursor (DR-phospholipase A)	33.8%
HP0540	cag pathogenicity island protein (cagI9)	99.5%	HP1132	ATP synthase F1, subunit beta (atpD)	85.6%	PURINES, PYRIMIDINES, NUCLEOSIDES AND NUCLEOTIDES		
HP0541	cag pathogenicity island protein (cagI2)	92.5%	HP1135	ATP synthase F1, subunit delta (atpH)	24.6%	General		
HP0542	cag pathogenicity island protein (cagI2)	97.8%	HP1131	ATP synthase F1, subunit epsilon (atpC)	32.7%	HP0757	beta-alanine synthetase homologue	40.0%
HP0543	cag pathogenicity island protein (cagI2)	97.9%	HP1133	ATP synthase F1, subunit gamma (atpG)	37.8%	2'-Deoxyribonucleotide metabolism		
HP0544	cag pathogenicity island protein (cagI2)	99.0%	Electron transport			HP0372	deoxycytidine triphosphate deaminase (ddc)	28.2%
HP0545	cag pathogenicity island protein (cagI2)	98.5%	HP0146	cb3-type cytochrome c oxidase subunit Q (CcoQ)	44.2%	HP0865	deoxyuridine 5'-triphosphate nucleotidohydrolase (dnt)	41.4%
HP0546	cag pathogenicity island protein (cagI2)	95.7%	HP0265	cytochrome c biogenesis protein (ccfA)	35.4%	HP0364	ribonucleoside diphosphate reductase, beta subunit (nrdB)	39.0%
HP0547	cag pathogenicity island protein (cagI2)	92.9%	HP0378	cytochrome c biogenesis protein (ycf5)	37.5%	HP0680	ribonucleoside-diphosphate reductase 1 alpha subunit (nrDA)	28.4%
HP0522	cag pathogenicity island protein (cagI3)	98.1%	HP0147	cytochrome c oxidase, dihemoglobin, membrane-bound (flxR)	33.0%	HP0825	thioredoxin reductase (trxB)	45.9%
HP0523	cag pathogenicity island protein (cagI4)	95.7%	HP0144	cytochrome c oxidase, heme b and copper-binding subunit, membrane-bound (flxN)	43.9%	Purine ribonucleotide biosynthesis		
HP0524	cag pathogenicity island protein (cagI5)	99.1%	HP0145	cytochrome c oxidase, monoheme subunit, membrane-bound (flxO)	45.7%	HP0321	5'-guanylate kinase (gmk)	44.8%
HP0525	cag pathogenicity island protein (cagI6)	97.5%	HP1461	cytochrome c551 peroxidase	48.5%	HP0618	adenylate kinase (ack)	33.3%
HP0527	cag pathogenicity island protein (cagI7)	94.8%	HP0277	ferredoxin	52.5%	HP1112	adenylosuccinate lyase (purB)	49.5%
HP0528	cag pathogenicity island protein (cagI8)	99.0%	HP0588	ferredoxin-like protein	42.6%	HP0255	adenylosuccinate synthetase (purA)	44.6%
HP0529	cag pathogenicity island protein (cagI9)	98.9%	HP1508	ferredoxin-like protein	29.4%	HP1434	formyltetrahydrofolate hydrolase (purU)	49.1%
HP1378	competence lipoprotein (comL)	25.5%	HP1161	flavodoxin (fldA)	47.0%	HP1218	glycinamide ribonucleotide synthetase (purD)	31.8%
HP1361	competence locus E (comE3)	26.7%	HP0642	NAD(P)H-flavin oxidoreductase	46.1%	HP0854	GMP reductase (guaC)	31.8%
HP1008	conjugal transfer protein (traG)	27.3%	HP0954	oxygen-insensitive NAD(P)H nitroreductase	32.7%	HP0409	GMP synthase (guaA)	56.1%
HP1421	conjugative transfer region protein (trbB)	30.7%	HP0634	quinone-reactive Ni/Fe hydrogenase, cytochrome b subunit (hydB)	51.4%	HP0829	inosine-5'-monophosphate dehydrogenase (guaB)	58.5%
HP0533	DNA processing chain A (dprA)	32.9%	HP0632	quinone-reactive Ni/Fe hydrogenase, large subunit (hydA)	68.5%	HP0198	nucleoside diphosphate kinase (ndk)	67.7%
HP0042	trbI protein	31.4%	HP0631	quinone-reactive Ni/Fe hydrogenase, small subunit (hydB)	68.9%	HP0742	phosphoribosylpyrophosphate synthetase (prsA)	56.5%
HP0525	VirB11 homologue	100.0%	HP1539	ubiquinol cytochrome c oxidoreductase, cytochrome b subunit (fbcH)	39.3%	HP1530	purine nucleoside phosphorylase (punB)	20.7%
HP0441	VirB4 homologue	23.5%	HP1538	ubiquinol cytochrome c oxidoreductase, cytochrome c1 subunit (fbcH)	28.8%	Pyrimidine ribonucleotide biosynthesis		
HP0017	virB4 homologue (virB4)	25.2%	HP1540	ubiquinol cytochrome c oxidoreductase, Rieske 2Fe-2S subunit (fbcF)	39.2%	HP1084	aspartate transcarbamoylase (pyrB)	38.7%
HP0459	virB4 homologue (virB4)	25.3%	Entner-Doudorff			HP0919	carbamoyl-phosphate synthase (glutamine-hydrolysing) (pyrAb)	48.6%
CENTRAL INTERMEDIARY METABOLISM								
General								
HP1014	7- α -hydroxysteroid dehydrogenase (hdhA)	33.2%	HP1099	2-keto-3-deoxy-6-phosphogluconate aldolase (eda)	50.3%	HP1237	carbamoyl-phosphate synthetase (pyrAa)	39.7%
HP1186	carbonic anhydrase	37.0%	HP1100	6-phosphogluconate dehydratase	50.7%	HP0349	GTP synthetase (pyrG)	50.7%
HP0004	carbonic anhydrase (icfA)	33.3%	Fermentation			HP0266	dihydroorotase (pyrC)	-1.0%
HP0869	hydrogenase expression/formation protein (hypA)	28.1%	HP0691	3-oxoadipate coA-transferase subunit A (yxiD)	65.5%	HP0681	dihydroorotase dehydrogenase (pyrD)	41.5%
HP0800	hydrogenase expression/formation protein (hypB)	41.4%	HP0692	3-oxoadipate coA-transferase subunit B (yxiE)	73.2%	HP1257	orotate phosphoribosyltransferase (pyrE)	35.5%
HP0899	hydrogenase expression/formation protein (hypC)	38.5%	HP0903	acetate kinase (ackA) (Escherichia coli)	42.3%	HP1474	thymidylate kinase (tmk)	33.9%
HP0898	hydrogenase expression/formation protein (hypD)	47.8%	HP0904	phosphate acetyltransferase (pta)	51.0%	HP0777	uridine 5'-monophosphate (UMP) kinase (pyrH)	50.4%
HP0047	hydrogenase expression/formation protein (hypE)	39.7%	HP0905	phosphotransacetylase (pta)	26.9%	Salvage of nucleosides and nucleotides		
HP0197	S-adenosylmethionine synthetase 2 (metX)	62.1%	HP0357	short-chain alcohol dehydrogenase	57.6%	HP1014	2'-O-cyclic-nucleotide 2'-phosphodiesterase (cptB)	31.8%
Amino sugars								
HP1532	glucosamine fructose-6-phosphate aminotransferase (isomerizing) (glmS)	41.7%	Glucuronogenesis			HP0672	adenine phosphoribosyltransferase (apt)	50.3%
Phosphorus compounds								
HP0620	inorganic pyrophosphatase (ppa)	50.0%	HP1385	fructose-16-bisphosphatase	36.4%	HP1179	phosphopentomutase (deoB)	55.9%
HP0696	N-methylhydantoinase	26.9%	HP0121	phosphoenolpyruvate synthase (ppsA)	52.4%	HP1178	purine-nucleoside phosphorylase (deoD)	55.5%
HP1010	polyphosphate kinase (ppk)	38.5%	HP1345	phosphoglycerate kinase	47.3%	HP0735	xanthine guanine phosphoribosyl transferase (gpt)	27.1%
Polyamine biosynthesis								
HP0422	arginine decarboxylase (speA)	33.3%	HP0176	fructose-bisphosphate aldolase (tsr)	46.0%	Sugar-nucleotide biosynthesis and conversions		
HP0020	carboxynorspermidine decarboxylase (nspC)	45.6%	HP1103	glucokinase (gik)	41.5%	HP0043	mannose-6-phosphate isomerase (pmi) or (algA)	42.8%
HP0832	spermidine synthase (speE)	26.5%	HP1166	glucose-6-phosphate isomerase (pgi)	53.3%	HP0045	modulation protein (nolK)	44.3%
Other								
HP0070	urease accessory protein (ureF)	97.1%	HP0921	glyceraldehyde-3-phosphate dehydrogenase (gap)	46.5%	HP0646	UDP-glucose pyrophosphorylase (galU)	65.6%
HP0069	urease accessory protein (ureG)	94.5%	HP1346	glyceraldehyde-3-phosphate dehydrogenase (gap)	46.7%	HP0683	UDP-N-acetylglucosamine pyrophosphorylase (glmU)	40.0%
HP0068	urease accessory protein (ureG)	95.0%	HP0974	phosphoglycerate mutase (pgm)	46.6%	REGULATORY FUNCTIONS		
HP0067	urease accessory protein (ureH)	96.2%	HP0194	triosephosphate isomerase (tpi)	34.5%	General		
HP0071	urease accessory protein (ureI)	98.5%	Pentose phosphate pathway			HP1032	alternative transcription initiation factor, sigma-F (flaA)	34.6%
HP0073	urease alpha subunit (ureA)	100.0%	HP1386	D-ribulose-5-phosphate 3 epimerase (rpe)	44.2%	HP1168	carbon starvation protein (cstA)	59.8%
HP0072	urease beta subunit (urea amidohydrolase) (ureB)	100.0%	HP1102	glucose-6-phosphate 1-dehydrogenase		HP1442	carbon storage regulator (csrA)	43.3%
HP0075	urease protein (ureC)	98.0%				HP1027	ferric uptake regulation protein (fur)	39.9%
ENERGY METABOLISM								
Aerobic								
HP1222	D-lactate dehydrogenase (ldd)	27.0%				HP0278	guanosine pentaphosphate phosphohydrolase (gppA)	26.4%
HP0861	glycerol-3-phosphate dehydrogenase (NAD(P)H)	36.8%				HP0400	penicillin tolerance protein (lytB)	30.6%
HP0037	NADH-ubiquinone oxidoreductase subunit	19.4%						
HP1269	NADH-ubiquinone oxidoreductase, NQO10							

HP0775	penta-phosphate guanosine-3O-pyrophosphohydrolase (spoT)	36.7%	HP1471	type IIS restriction enzyme R protein (BCOIIR)	28.2%	HP0399	ribosomal protein S1 (rps1)	30.5%
HP0224	peptide methionine sulphoxide reductase (msrA)	66.8%	HP1366	type IIS restriction enzyme R protein (MBOIIR)	37.1%	HP1320	ribosomal protein S10 (rps10)	58.2%
HP1025	putative heat shock protein (hspR)	46.2%	HP1208	ulcer associated adenine specific DNA methyltransferase	93.4%	HP1295	ribosomal protein S11 (rps11)	56.2%
HP1572	regulatory protein DnrR	31.9%	HP1209	ulcer-associated gene restriction endonuclease (iceA)	95.5%	HP1197	ribosomal protein S12 (rps12)	79.0%
HP0703	response regulator	44.2%	HP1347	uracil-DNA glycosylase (ung)	43.1%	HP1296	ribosomal protein S13 (rps13)	55.8%
HP1021	response regulator	28.7%	TRANSCRIPTION		HP1306	ribosomal protein S14 (rps14)	68.3%	
HP1043	response regulator	26.8%	Degradation of RNA		HP1040	ribosomal protein S15 (rps15)	57.8%	
HP1365	response regulator (ompR)	32.4%	HP1213	polynucleotide phosphorylase (pnp)	38.9%	HP1151	ribosomal protein S16 (rps16)	46.8%
HP0166	RNA polymerase sigma-54 factor (rpoN)	51.0%	HP1293	DNA-dependent RNA polymerase		HP1310	ribosomal protein S17 (rps17)	55.4%
HP0714	RNA polymerase sigma-70 factor (rpoD)	37.1%	HP1293	DNA-directed RNA polymerase, alpha subunit (rpoA)	35.3%	HP1244	ribosomal protein S18 (rps18)	55.2%
HP0088	RNA polymerase sigma-70 factor (rpoD)	43.5%	HP1198	DNA-directed RNA polymerase, beta subunit (rpoB)	47.8%	HP1315	ribosomal protein S19 (rps19)	61.1%
HP0792	sigma-54 interacting protein	97.7%	Transcription factors		HP1554	ribosomal protein S2 (rps2)	49.6%	
HP0164	signal-transducing protein, histidine kinase	27.1%	HP0866	transcription elongation factor GreA (greA)	50.3%	HP0076	ribosomal protein S20 (rps20)	41.4%
HP1364	signal-transducing protein, histidine kinase	24.9%	HP1514	transcription termination factor NusA (nusA)	39.1%	HP0662	ribosomal protein S21 (rps21)	42.4%
HP0244	signal-transducing protein, histidine kinase (atoS)	30.0%	HP0001	transcription termination factor NusB (nusB)	30.2%	HP1313	ribosomal protein S3 (rps3)	56.7%
HP0948	transcriptional regulator (hypF)	34.5%	HP1203	transcription termination factor NusG (nusG)	41.0%	HP1294	ribosomal protein S4 (rps4)	51.2%
HP1287	transcriptional regulator (tenA)	34.7%	HP0550	transcription termination factor Rho (rho)	56.6%	HP1302	ribosomal protein S5 (rps5)	65.5%
HP0727	transcriptional regulator, putative	33.3%	RNA processing		HP1246	ribosomal protein S6 (rps6)	32.1%	
REPLICATION			TRANSLATION		HP1196	ribosomal protein S7 (rps7)	62.2%	
Degradation of DNA			General		HP1305	ribosomal protein S8 (rps8)	45.0%	
HP0275	ATP-dependent nuclease (addB)	27.2%	HP0944	translation initiation inhibitor, putative	45.6%	HP0083	ribosomal protein S9 (rps9)	50.4%
HP0269	exonuclease VII, large subunit (xseA)	37.6%	Aminocyl tRNA synthetases		HP1047	ribosome-binding factor A (rbfA)	26.3%	
DNA replication, restriction, modification, recombination and repair			Aminoacyl tRNA synthetases		tRNA modification			
HP0142	A/G-specific adenine glycosylase (mutY)	38.2%	HP1241	alanine-tRNA synthetase (alaS)	44.9%	HP1141	methionyl-tRNA formyltransferase (fmt)	37.5%
HP0050	adenine specific DNA methyltransferase (dpmA)	37.4%	HP0319	arginine-tRNA synthetase (argS)	35.8%	HP1497	peptidyl-tRNA hydrolase (pth)	46.6%
HP0910	adenine specific DNA methyltransferase (HINDIII)	33.4%	HP0617	aspartyl-tRNA synthetase (aspS)	50.1%	HP0361	pseudouridylyl synthase I (hisT)	32.2%
HP1352	adenine specific DNA methyltransferase (HINFIM)	62.5%	HP0886	cysteinyl-tRNA synthetase (cysS)	97.3%	HP1448	ribonuclease P, protein component (mpaA)	29.3%
HP0263	adenine specific DNA methyltransferase (hpaIm)	33.9%	HP0476	glutamyl-tRNA synthetase (glxT)	43.1%	HP1062	S-adenosylmethionine:tRNA	
HP0481	adenine specific DNA methyltransferase (MFOKI)	29.3%	HP0643	glutamyl-tRNA synthetase (glxX)	39.8%	HP1513	ribosyltransferase-isomerase (queA)	39.3%
HP0260	adenine specific DNA methyltransferase (mod)	33.9%	HP0960	glycyl-tRNA synthetase, alpha subunit (glyO)	60.1%	HP1148	tRNA (guanine-N1)-methyltransferase (trmD)	39.1%
HP0593	adenine specific DNA methyltransferase (mod)	38.5%	HP0972	glycyl-tRNA synthetase, beta subunit (glyS)	33.6%	HP1415	tRNA delta(2)-isopentenylpyrophosphate transferase (miaA)	30.7%
HP1522	adenine specific DNA methyltransferase (mod)	42.2%	HP1190	histidyl-tRNA synthetase (hisS)	32.4%	HP0281	tRNA-guanine transglycosylase (tgt)	45.6%
HP0478	adenine specific DNA methyltransferase (VSPIM)	42.1%	HP1422	isoleucyl-tRNA synthetase (ileS)	49.7%	Translation factors		
HP0054	adenine/cytosine DNA methyltransferase	32.1%	HP1547	leucyl-tRNA synthetase (leuS)	45.9%	HP0247	ATP-dependent RNA helicase, DEAD-box family (deaD)	37.7%
HP0790	anti-codon nuclease masking agent (prfB)	42.9%	HP0182	lysyl-tRNA synthetase (lysS)	58.6%	HP0777	peptide chain release factor RF-1 (prfA)	52.6%
HP1529	chromosomal replication initiator protein (dnaA)	34.9%	HP0417	methionyl-tRNA synthetase (metS)	42.4%	HP1256	peptide chain release factor RF-2 (prfB)	49.6%
HP1121	cytosine specific DNA methyltransferase (BSP6IM)	37.0%	HP0403	phenylalanyl-tRNA synthetase, alpha subunit (pheS)	48.7%	HP1195	ribosome releasing factor (lrr)	43.7%
HP0051	cytosine specific DNA methyltransferase (DDEM)	39.0%	HP0402	phenylalanyl-tRNA synthetase, beta subunit (pheT)	30.0%	HP1095	translation elongation factor EF-G (fusA)	67.5%
HP0483	cytosine specific DNA methyltransferase (HPHIMC)	38.7%	HP0238	prolyl-tRNA synthetase (proS)	39.8%	HP0177	translation elongation factor EF-P (efp)	45.1%
HP0701	DNA gyrase, sub A (gyrA)	97.4%	HP1480	seryl-tRNA synthetase (serS)	48.3%	HP1555	translation elongation factor EF-Ts (tsf)	43.1%
HP0601	DNA gyrase, sub B (gyrB)	46.0%	HP1213	threonyl-tRNA synthetase (thrS)	42.1%	HP1205	translation elongation factor EF-Tu (tufB)	89.5%
HP1478	DNA helicase II (uvrD)	35.3%	HP1253	tryptophanyl-tRNA synthetase (trpS)	52.6%	HP1298	translation initiation factor EF1- α (infA)	65.3%
HP0548	DNA helicase, putative	38.8%	HP0774	tyrosyl-tRNA synthetase (tyrS)	54.7%	HP1048	translation initiation factor IF-2 (infB)	45.4%
HP0615	DNA ligase (lig)	40.1%	HP1153	valyl-tRNA synthetase (valS)	43.7%	HP0124	translation initiation factor IF-3 (infC)	43.4%
HP0621	DNA mismatch repair protein (MutS)	32.6%	Degradation of proteins, peptides and glycopeptides		TRANSPORT AND BINDING PROTEINS			
HP1470	DNA polymerase I (polA)	40.0%	HP0570	aminopeptidase a/i (pepA)	38.5%	General		
HP1460	DNA polymerase III alpha-subunit (dnaE)	42.0%	HP0033	ATP-dependent Clp protease (clpA)	40.3%	HP0179	ABC transporter, ATP-binding protein	66.7%
HP0500	DNA polymerase III beta-subunit (dnaH)	26.0%	HP0794	ATP-dependent clp protease proteolytic component (clpP)	64.6%	HP0613	ABC transporter, ATP-binding protein	31.1%
HP1231	DNA polymerase III delta prime subunit (polB)	48.6%	HP1379	ATP-dependent protease (lon)	43.9%	HP0715	ABC transporter, ATP-binding protein	52.3%
HP1387	DNA polymerase III epsilon subunit (dnaQ)	35.1%	HP0223	ATP-dependent protease (sms)	41.0%	HP1576	ABC transporter, ATP-binding protein (abc)	48.2%
HP0717	DNA polymerase III gamma and tau subunits (dnaX)	39.0%	HP1374	ATP-dependent protease ATPase subunit (clpX)	56.3%	HP1465	ABC transporter, ATP-binding protein (H1087)	37.8%
HP0012	DNA primase (dnaG)	36.6%	HP0264	ATP-dependent protease binding subunit (clpB)	97.7%	HP1220	ABC transporter, ATP-binding protein (yhocG)	31.5%
HP1523	DNA recombinase (recG)	32.7%	HP0169	collagenase (prtC)	40.1%	HP0853	ABC transporter, ATP-binding protein (yheS)	36.3%
HP1393	DNA repair protein (recN)	28.3%	HP0516	heat-shock protein (hspU) ORF1	98.4%	HP1577	ABC transporter, permease protein (yaeE)	43.1%
HP0716	DNA topoisomerase I (topA)	45.1%	HP0057	heat-shock protein (hspU)	57.1%	HP0607	acylfavine resistance protein (acrB)	29.7%
HP0440	DNA topoisomerase I (topA)	31.7%	HP0470	oligodeoxyphosphatase F (pepF)	97.9%	HP1432	histidine and glutamine-rich protein	50.0%
HP0602	endonuclease III	36.6%	HP0657	processing protease (ymxG)	24.2%	HP1427	histidine-rich, metal binding polypeptide (hpn)	100.0%
HP0585	endonuclease III (nth)	40.1%	HP1485	proline dipeptidase (pepG)	35.2%	HP1206	multidrug-resistance protein (hetA)	26.2%
HP0705	exonuclease ABC subunit A (uvrA)	53.4%	HP1350	protease	40.6%	HP1082	multidrug-resistance protein (msbA)	32.4%
HP1114	exonuclease ABC subunit B (uvrB)	53.1%	HP1012	protease (pqeA)	29.6%	HP0600	multidrug-resistance protein (spaB)	29.7%
HP0821	exonuclease ABC subunit C (uvrC)	31.5%	HP1435	protease IV (PspA)	41.7%	HP1181	multidrug-efflux transporter	29.1%
HP1526	exodeoxyribonuclease (lexA)	58.9%	HP0404	protein kinase C inhibitor (SP-P16436)	40.2%	HP0497	sodium- and chloride-dependent transporter	31.7%
HP0213	glucose inhibited division protein (gidA)	48.5%	HP1019	serine protease (htrA)	52.9%	HP0498	sodium- and chloride-dependent transporter	30.8%
HP1063	glucose-inhibited division protein (gidB)	32.9%	HP1584	sioligocycloprotease (gcp)	35.7%	HP0214	sodium-dependent transporter (huNadC-1)	36.6%
HP1563	Holliday junction DNA helicase (ruvA)	39.0%	HP0382	zinc-metalloprotease (YJR117W)	36.2%	Amino acids, peptides and amines		
HP0893	Holliday junction DNA helicase (ruvA)	39.0%	Nucleoproteins		HP0940	amino acid ABC transporter, periplasmic binding protein (yckK)	41.5%	
HP1059	Holliday junction DNA helicase (ruvB)	54.6%	HP0835	histone-like DNA-binding protein HU (hup)	44.6%	HP0939	amino acid ABC transporter, permease protein (yckL)	46.9%
HP0877	Holliday junction endodeoxyribonuclease (ruvC)	34.7%	Protein modification		HP1017	amino acid permease (rocE)	41.7%	
HP0675	integrase/recombinase (xerC)	31.8%	HP0383	L-isocaparyl-protein carboxyl methyltransferase (pcm)	43.0%	HP0942	D-alanine glycine permease (dagA)	44.5%
HP0995	integrase/recombinase (xerD)	27.8%	HP1299	methionine amino peptidase (map)	43.0%	HP0301	dipeptide ABC transporter, ATP-binding protein (dppD)	59.5%
HP0323	membrane bound endonuclease (nuc)	31.1%	HP1441	peptidyl-prolyl cis-trans isomerase B, cyclosporin-type rotamase (ipi)	58.1%	HP0302	dipeptide ABC transporter, ATP-binding protein (dppF)	54.8%
HP0678	methylated DNA-protein-cysteine methyltransferase (dat1)	41.0%	HP1123	peptidyl-prolyl cis-trans isomerase, FKBP-type rotamase (lydC)	40.4%	HP0298	dipeptide ABC transporter, periplasmic dipeptide-binding protein (dppA)	39.8%
HP0387	primosomal protein replication factor (priA)	36.3%	HP0793	polypeptide deformylase (def)	41.8%	HP0299	dipeptide ABC transporter, permease protein (dppB)	49.3%
HP0153	recombinase (recA)	99.1%	Ribosomal proteins: synthesis and modification		HP0300	dipeptide ABC transporter, permease protein (dppC)	52.5%	
HP0925	recombinational DNA repair protein (recR)	36.5%	HP1201	ribosomal protein L1 (rpl1)	52.0%	HP1506	glutamate permease (gltS)	56.9%
HP0911	rep helicase, single-stranded DNA-dependent ATPase (rep)	33.8%	HP1200	ribosomal protein L10 (rpl10)	30.4%	HP1171	glutamine ABC transporter, ATP-binding protein (glnQ)	51.9%
HP1362	replicative DNA helicase (dnaB)	39.4%	HP1202	ribosomal protein L11 (rpl11)	63.8%	HP1172	glutamine ABC transporter, periplasmic glutamine-binding protein (glnH)	32.2%
HP1383	restriction modification system S subunit	38.1%	HP1088	ribosomal protein L11 methyltransferase (prmA)	38.4%	HP1169	glutamine ABC transporter, permease protein (glnP)	27.6%
HP0661	ribonuclease H (rnhA)	58.4%	HP0084	ribosomal protein L13 (rpl13)	50.0%	HP1170	glutamine ABC transporter, permease protein (glnP)	30.9%
HP1323	ribonuclease HII (rnhB)	36.3%	HP1309	ribosomal protein L14 (rpl14)	65.9%	HP0250	oligopeptide ABC transporter, ATP-binding protein (oppD)	39.1%
HP1245	single-strand DNA-binding protein (ssb)	32.6%	HP1301	ribosomal protein L15 (rpl15)	42.5%	HP1252	oligopeptide ABC transporter, periplasmic oligopeptide-binding protein (oppA)	28.7%
HP0348	single-stranded-DNA-specific exonuclease (recI)	33.6%	HP1312	ribosomal protein L16 (rpl16)	62.4%	HP1251	oligopeptide ABC transporter, permease protein (oppB)	59.6%
HP1009	site-specific recombinase	21.3%	HP1232	ribosomal protein L17 (rpl17)	48.3%	HP0251	oligopeptide ABC transporter, permease protein (oppC)	31.4%
HP1541	transcription-repair coupling factor (trcF)	37.7%	HP1303	ribosomal protein L18 (rpl18)	45.5%	HP0819	osmoprotection protein (proV)	38.3%
HP0462	type I restriction enzyme S protein (hsdS)	37.0%	HP1147	ribosomal protein L19 (rpl19)	50.9%	HP0818	osmoprotection protein (proWX)	30.4%
HP0463	type I restriction enzyme M protein (hsdM)	29.4%	HP1316	ribosomal protein L2 (rpl2)	58.9%	HP0055	proline permease (putP)	51.4%
HP0464	type I restriction enzyme R protein (hsdR)	31.7%	HP0126	ribosomal protein L20 (rpl20)	54.8%	HP0936	proline/betaine transporter (proP)	29.1%
HP0846	type I restriction enzyme R protein (hsdR)	48.0%	HP0296	ribosomal protein L21 (rpl21)	46.1%	HP0133	serine transporter (sdcA)	44.6%
HP0848	type I restriction enzyme S protein (hsdS)	37.0%	HP1314	ribosomal protein L22 (rpl22)	44.9%	Anions		
HP0850	type I restriction enzyme M protein (hsdM)	54.4%	HP1317	ribosomal protein L23 (rpl23)	31.7%	HP0475	molybdenum ABC transporter, ATP-binding protein (modD)	38.4%
HP1402	type I restriction enzyme R protein (hsdR)	26.6%	HP1308	ribosomal protein L24 (rpl24)	52.2%	HP0473	molybdenum ABC transporter, periplasmic molybdate-binding protein (modA)	95.9%
HP1403	type I restriction enzyme M protein (hsdM)	37.1%	HP0297	ribosomal protein L27 (rpl27)	64.7%	HP0474	molybdenum ABC transporter, permease protein (modB)	28.7%
HP0902	type II restriction enzyme M protein (hsdM)	55.3%	HP0491	ribosomal protein L28 (rpl28)	41.7%	HP0313	nitrite extrusion protein (narK)	23.6%
HP0091	type II restriction enzyme R protein (hsdR)	50.7%	HP1311	ribosomal protein L29 (rpl29)	45.6%	HP1491	phosphate permease	34.8%
HP1369	type III restriction enzyme M protein (mod)	45.6%	HP1319	ribosomal protein L3 (rpl3)	41.8%	Carbohydrates, organic alcohols and acids		
HP1370	type III restriction enzyme M protein (mod)	37.0%	HP0551	ribosomal protein L31 (rpl31)	49.3%	HP1043	2-oxoglutarate/malate translocator (SODT1)	37.0%
HP1371	type III restriction enzyme R protein	26.2%	HP0200	ribosomal protein L32 (rpl32)	41.7%	HP1091	alpha-ketoglutarate permease (kgfP)	45.9%
HP0592	type III restriction enzyme R protein (res)	30.6%	HP1204	ribosomal protein L33 (rpl33)	55.1%	HP0724	anaerobic C4-dicarboxylate transport protein (dcaA)	53.8%
HP1521	type III restriction enzyme R protein (res)	33.1%	HP1447	ribosomal protein L34 (rpl34)	70.5%	HP1174	glucose/galactose transporter (gluP)	53.8%
HP1472	type IIS restriction enzyme M protein (mod)	32.4%	HP0125	ribosomal protein L35 (rpl35)	50.8%	HP0141	L-lactate permease (lcpP)	55.5%
HP1367	type IIS restriction enzyme M1 protein (mod) (Moraxella bovis)	59.3%	HP1297	ribosomal protein L36 (rpl36)	81.6%	HP0140	L-lactate permease (lcpP)	58.7%
HP1368	type IIS restriction enzyme M2 protein (mod)	33.0%	HP1318	ribosomal protein L4 (rpl4)	40.6%			
HP1517	type IIS restriction enzyme R and M protein (ECO57IR)	26.7%	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%			

Cations			HP0258	conserved hypothetical integral membrane protein	32.7%	HP0728	conserved hypothetical protein	29.3%
HP0791	cadmium-transporting ATPase, P-type (cdaA)		HP0284	conserved hypothetical integral membrane protein	29.2%	HP0734	conserved hypothetical protein	31.0%
HP0969	cation efflux system protein (czcA)	97.5%	HP0362	conserved hypothetical integral membrane protein	28.8%	HP0745	conserved hypothetical protein	30.2%
HP1328	cation efflux system protein (czcA)	28.9%	HP0415	conserved hypothetical integral membrane protein	44.4%	HP0747	conserved hypothetical protein	32.4%
HP1329	cation efflux system protein (czcA)	31.3%	HP0467	conserved hypothetical integral membrane protein	100.0%	HP0760	conserved hypothetical protein	36.1%
HP1503	cation-transporting ATPase, P-type (copA)	30.3%	HP0571	conserved hypothetical integral membrane protein	29.5%	HP0810	conserved hypothetical protein	31.0%
HP1073	copper ion binding protein (copP)	92.4%	HP0644	conserved hypothetical integral membrane protein	30.3%	HP0813	conserved hypothetical protein	32.5%
HP1072	copper-transporting ATPase, P-type (copA)	93.9%	HP0677	conserved hypothetical integral membrane protein	28.5%	HP0823	conserved hypothetical protein	27.8%
HP0471	glutathione-regulated potassium-efflux system protein (ketB)	99.3%	HP0693	conserved hypothetical integral membrane protein	46.7%	HP0860	conserved hypothetical protein	52.1%
HP0687	iron(II) transport protein (fecB)	33.6%	HP0718	conserved hypothetical integral membrane protein	33.5%	HP0890	conserved hypothetical protein	32.2%
HP1561	iron(III) ABC transporter, periplasmic iron-binding protein (cseE)	27.5%	HP0737	conserved hypothetical integral membrane protein	33.3%	HP0891	conserved hypothetical protein	33.8%
HP1562	iron(III) ABC transporter, periplasmic iron-binding protein (cseE)	28.2%	HP0758	conserved hypothetical integral membrane protein	47.6%	HP0892	conserved hypothetical protein	39.1%
HP0888	iron(III) dicitrate ABC transporter, ATP-binding protein (fecC)	34.4%	HP0759	conserved hypothetical integral membrane protein	31.1%	HP0894	conserved hypothetical protein	39.8%
HP0889	iron(III) dicitrate ABC transporter, permease protein (fecD)	38.3%	HP0787	conserved hypothetical integral membrane protein	25.2%	HP0926	conserved hypothetical protein	30.7%
HP0686	iron(III) dicitrate transport protein (fecA)	29.7%	HP0851	conserved hypothetical integral membrane protein	37.3%	HP0934	conserved hypothetical protein	33.6%
HP0807	iron(III) dicitrate transport protein (fecA)	28.5%	HP0920	conserved hypothetical integral membrane protein	36.3%	HP0956	conserved hypothetical protein	36.2%
HP1400	iron(III) dicitrate transport protein (fecA)	26.3%	HP0946	conserved hypothetical integral membrane protein	35.9%	HP0959	conserved hypothetical protein	31.1%
HP1344	magnesium and cobalt transport protein (corA)	26.3%	HP0962	conserved hypothetical integral membrane protein	38.5%	HP0966	conserved hypothetical protein	29.1%
HP1183	Na ⁺ /H ⁺ antiporter (napA)	26.6%	HP0983	conserved hypothetical integral membrane protein	32.8%	HP0975	conserved hypothetical protein	25.0%
HP1552	Na ⁺ /H ⁺ antiporter (nhaA)	49.2%	HP1044	conserved hypothetical integral membrane protein	30.6%	HP1020	conserved hypothetical protein	31.5%
HP1077	nickel transport protein (nixA)	98.7%	HP1061	conserved hypothetical integral membrane protein	35.0%	HP1037	conserved hypothetical protein	96.9%
HP0490	putative potassium channel protein, putative	25.7%	HP1080	conserved hypothetical integral membrane protein	44.0%	HP1046	conserved hypothetical protein	32.6%
Nucleosides, purines and pyrimidines			HP1162	conserved hypothetical integral membrane protein	27.6%	HP1049	conserved hypothetical protein	39.7%
HP1290	nicotinamide mononucleotide transporter (pnuC)	28.0%	HP1175	conserved hypothetical integral membrane protein	40.6%	HP1066	conserved hypothetical protein	41.3%
HP1180	pyrimidine nucleoside transport protein (nupC)	32.9%	HP1184	conserved hypothetical integral membrane protein	23.5%	HP1149	conserved hypothetical protein	24.7%
Other			HP1185	conserved hypothetical integral membrane protein	55.5%	HP1160	conserved hypothetical protein	34.7%
HP0876	iron-regulated outer membrane protein (frpB)	27.6%	HP1225	conserved hypothetical integral membrane protein	31.6%	HP1182	conserved hypothetical protein	34.6%
HP0915	iron-regulated outer membrane protein (frpB)	28.1%	HP1234	conserved hypothetical integral membrane protein	29.0%	HP1214	conserved hypothetical protein	21.5%
HP0916	iron-regulated outer membrane protein (frpB)	28.8%	HP1235	conserved hypothetical integral membrane protein	29.0%	HP1221	conserved hypothetical protein	42.4%
HP1129	biopolymer transport protein (exbD)	29.7%	HP1330	conserved hypothetical integral membrane protein	41.7%	HP1240	conserved hypothetical protein	22.5%
HP1130	biopolymer transport protein (exbD)	33.5%	HP1331	conserved hypothetical integral membrane protein	33.6%	HP1242	conserved hypothetical protein	42.3%
HP1339	biopolymer transport protein (exbD)	46.8%	HP1343	conserved hypothetical integral membrane protein	49.1%	HP1259	conserved hypothetical protein	44.6%
HP1340	biopolymer transport protein (exbD)	35.8%	HP1363	conserved hypothetical integral membrane protein	33.1%	HP1284	conserved hypothetical protein	36.8%
HP1445	biopolymer transport protein (exbD)	45.5%	HP1407	conserved hypothetical integral membrane protein	22.4%	HP1291	conserved hypothetical protein	26.3%
HP1446	biopolymer transport protein (exbD)	36.2%	HP1466	conserved hypothetical integral membrane protein	30.9%	HP1335	conserved hypothetical protein	33.9%
HP1512	iron-regulated outer membrane protein (frpB)	26.6%	HP1484	conserved hypothetical integral membrane protein	41.2%	HP1337	conserved hypothetical protein	27.2%
HP0653	nonheme iron-containing ferritin (pfr)	99.4%	HP1509	conserved hypothetical integral membrane protein	34.3%	HP1338	conserved hypothetical protein	36.2%
HP1341	siderophore-mediated iron transport protein (tonB)	37.2%	HP1548	conserved hypothetical integral membrane protein	30.6%	HP1394	conserved hypothetical protein	33.6%
OTHER CATEGORIES			HP1588	conserved hypothetical integral membrane protein	32.0%	HP1401	conserved hypothetical protein	27.5%
General			HP1589	conserved hypothetical integral membrane protein	35.1%	HP1413	conserved hypothetical protein	41.6%
HP0924	4-oxalocrotonate tautomerase (dmpl)	37.7%	HP1590	conserved hypothetical integral membrane protein	30.9%	HP1414	conserved hypothetical protein	27.4%
HP1034	ATP-binding protein (yixH)	36.3%	HP1631	conserved hypothetical integral membrane protein	41.7%	HP1417	conserved hypothetical protein	23.7%
HP1000	PARA protein	29.7%	HP1633	conserved hypothetical integral membrane protein	33.6%	HP1423	conserved hypothetical protein	40.3%
HP1139	SpoQJ regulator (soj)	47.4%	HP1635	conserved hypothetical integral membrane protein	38.8%	HP1426	conserved hypothetical protein	40.0%
HP0827	ss-DNA binding protein 12RNP2 precursor	46.8%	HP1650	conserved hypothetical integral membrane protein	31.6%	HP1428	conserved hypothetical protein	37.8%
Adaptations and atypical conditions			HP1657	conserved hypothetical integral membrane protein	39.0%	HP1443	conserved hypothetical protein	37.9%
HP1496	general stress protein (ctc)	26.5%	HP1689	conserved hypothetical integral membrane protein	37.0%	HP1449	conserved hypothetical protein	39.0%
HP1483	gerC2 protein (gerC2)	33.3%	HP1690	conserved hypothetical integral membrane protein	31.4%	HP1453	conserved hypothetical protein	26.8%
HP0927	heat shock protein (hspX)	32.8%	HP1697	conserved hypothetical integral membrane protein	23.8%	HP1459	conserved hypothetical protein	30.1%
HP0280	heat shock protein B (hspB)	27.2%	HP1705	conserved hypothetical integral membrane protein	42.9%	HP1504	conserved hypothetical protein	23.9%
HP1228	invasion protein (invA)	38.2%	HP1708	conserved hypothetical integral membrane protein	37.2%	HP1510	conserved hypothetical protein	30.6%
HP0970	nickel-cobalt-cadmium resistance protein (nccB)	21.1%	HP1711	conserved hypothetical integral membrane protein	32.3%	HP1533	conserved hypothetical protein	25.4%
HP1444	small protein (smpB)	42.1%	HP1712	conserved hypothetical integral membrane protein	31.9%	HP1570	conserved hypothetical protein	40.5%
HP0930	stationary-phase survival protein (surE)	37.7%	HP1716	conserved hypothetical integral membrane protein	30.6%	HP1573	conserved hypothetical protein	42.2%
HP0315	virulence associated protein D (vapD)	70.2%	HP1718	conserved hypothetical integral membrane protein	30.6%	HP1587	conserved hypothetical protein	39.0%
HP0967	virulence associated protein D (vapD)	28.9%	HP1719	conserved hypothetical integral membrane protein	30.6%	HP1588	conserved hypothetical protein	32.0%
HP1248	virulence associated protein homolog (vacB)	36.0%	HP1720	conserved hypothetical integral membrane protein	30.6%	HP1589	conserved hypothetical protein	35.1%
HP0885	virulence factor mviN protein (mviN)	33.5%	HP1721	conserved hypothetical integral membrane protein	30.6%	HP1713	conserved hypothetical protein	41.8%
Colicin-related functions			HP1722	conserved hypothetical integral membrane protein	30.6%	HP0028	conserved hypothetical secreted protein	42.1%
HP1126	colicin tolerance-like protein (tolB)	25.7%	HP1723	conserved hypothetical integral membrane protein	30.6%	HP0139	conserved hypothetical secreted protein	37.1%
HP0428	phage/colicin/tellurite resistance cluster terY protein	25.6%	HP1724	conserved hypothetical integral membrane protein	30.6%	HP0160	conserved hypothetical secreted protein	30.6%
Drug and analog sensitivity			HP1725	conserved hypothetical integral membrane protein	30.6%	HP0190	conserved hypothetical secreted protein	31.4%
HP1431	16S rRNA (adenosine-N6,N6)-dimethyl-transferase (kgpA)	35.5%	HP1726	conserved hypothetical integral membrane protein	30.6%	HP0211	conserved hypothetical secreted protein	24.3%
HP0606	membrane fusion protein (mtrC)	24.2%	HP1727	conserved hypothetical integral membrane protein	30.6%	HP0235	conserved hypothetical secreted protein	31.5%
HP0630	modulator of drug activity (mda68)	62.3%	HP1728	conserved hypothetical integral membrane protein	30.6%	HP0257	conserved hypothetical secreted protein	29.2%
HP1476	phenylacrylic acid decarboxylase	39.7%	HP1729	conserved hypothetical integral membrane protein	30.6%	HP0320	conserved hypothetical secreted protein	36.4%
HP1165	tetracycline resistance protein tetA(P), putative	27.0%	HP1730	conserved hypothetical integral membrane protein	30.6%	HP0506	conserved hypothetical secreted protein	29.8%
Transposon-related functions			HP1731	conserved hypothetical integral membrane protein	30.6%	HP0518	conserved hypothetical secreted protein	96.9%
HP1008	IS200 insertion sequence from SARA17	33.9%	HP1732	conserved hypothetical integral membrane protein	30.6%	HP0785	conserved hypothetical secreted protein	26.6%
HP0414	IS200 insertion sequence from SARA17	33.9%	HP1733	conserved hypothetical integral membrane protein	30.6%	HP0949	conserved hypothetical secreted protein	39.7%
HP0988	IS605 transposase (tnpA)	97.2%	HP1734	conserved hypothetical integral membrane protein	30.6%	HP0977	conserved hypothetical secreted protein	29.4%
HP0998	IS605 transposase (tnpA)	97.2%	HP1735	conserved hypothetical integral membrane protein	30.6%	HP0980	conserved hypothetical secreted protein	29.4%
HP1096	IS605 transposase (tnpA)	97.2%	HP1736	conserved hypothetical integral membrane protein	30.6%	HP1075	conserved hypothetical secreted protein	42.9%
HP1535	IS605 transposase (tnpA)	97.2%	HP1737	conserved hypothetical integral membrane protein	30.6%	HP1088	conserved hypothetical secreted protein	27.0%
HP0437	IS605 transposase (tnpA)	97.2%	HP1738	conserved hypothetical integral membrane protein	30.6%	HP1117	conserved hypothetical secreted protein	32.3%
HP0989	IS605 transposase (tnpB)	93.4%	HP1739	conserved hypothetical integral membrane protein	30.6%	HP1216	conserved hypothetical secreted protein	31.9%
HP0997	IS605 transposase (tnpB)	93.4%	HP1740	conserved hypothetical integral membrane protein	30.6%	HP1285	conserved hypothetical secreted protein	38.0%
HP1095	IS605 transposase (tnpB)	93.4%	HP1741	conserved hypothetical integral membrane protein	30.6%	HP1288	conserved hypothetical secreted protein	37.5%
HP1534	IS605 transposase (tnpB)	93.4%	HP1742	conserved hypothetical integral membrane protein	30.6%	HP1464	conserved hypothetical secreted protein	27.4%
HP0438	IS605 transposase (tnpB)	93.4%	HP1743	conserved hypothetical integral membrane protein	30.6%	HP1488	conserved hypothetical secreted protein	29.8%
HP0413	transposase-like protein, PS3IS	33.6%	HP1744	conserved hypothetical integral membrane protein	30.6%	HP1551	conserved hypothetical secreted protein	42.7%
HP1007	transposase-like protein, PS3IS	34.3%	HP1745	conserved hypothetical integral membrane protein	30.6%	UNKNOWN		
Other			HP1746	conserved hypothetical integral membrane protein	30.6%	General		
HP0739	2-hydroxy-6-oxohepta-2,4-dienoate hydrolase	30.1%	HP1747	conserved hypothetical integral membrane protein	30.6%	HP0390	adhesin-thiol peroxidase (tagD)	38.3%
HYPOTHETICAL			HP1748	conserved hypothetical integral membrane protein	30.6%	HP1193	aldo-keto reductase, putative	46.6%
General			HP1749	conserved hypothetical integral membrane protein	30.6%	HP0872	alkylphosphonate uptake protein (phnA)	61.1%
HP0831	conserved hypothetical ATP binding protein	32.3%	HP1750	conserved hypothetical integral membrane protein	30.6%	HP0207	ATP-binding protein (mpr)	38.9%
HP0066	conserved hypothetical ATP-binding protein	34.7%	HP1751	conserved hypothetical integral membrane protein	30.6%	HP10136	bacterioferritin comigratory protein (bcp)	35.5%
HP0269	conserved hypothetical ATP-binding protein	37.7%	HP1752	conserved hypothetical integral membrane protein	30.6%	HP0485	catalase-like protein	30.8%
HP0312	conserved hypothetical ATP-binding protein	34.1%	HP1753	conserved hypothetical integral membrane protein	30.6%	HP1104	cinnamyl-alcohol dehydrogenase	
HP1321	conserved hypothetical ATP-binding protein	30.8%	HP1754	conserved hypothetical integral membrane protein	30.6%		EL3-2 (cad)	44.0%
HP1430	conserved hypothetical ATP-binding protein	38.1%	HP1755	conserved hypothetical integral membrane protein	30.6%	HP0981	exonuclease VII-like protein (xseA)	42.5%
HP1507	conserved hypothetical ATP-binding protein	51.6%	HP1756	conserved hypothetical integral membrane protein	30.6%	HP0569	GTP-binding protein (gtp1)	48.1%
HP1567	conserved hypothetical ATP-binding protein	40.9%	HP1757	conserved hypothetical integral membrane protein	30.6%	HP0303	GTP-binding protein (obg)	48.2%
HP1026	conserved hypothetical helicase-like protein	35.2%	HP1758	conserved hypothetical integral membrane protein	30.6%	HP0834	GTP-binding protein homologue (yphC)	36.7%
HP0022	conserved hypothetical integral membrane protein	30.8%	HP1759	conserved hypothetical integral membrane protein	30.6%	HP0480	GTP-binding protein, fusA-homolog (yihK)	54.1%
HP0189	conserved hypothetical integral membrane protein	43.1%	HP1760	conserved hypothetical integral membrane protein	30.6%	HP1489	lipase-like protein	21.7%
HP0226	conserved hypothetical integral membrane protein	27.6%	HP1761	conserved hypothetical integral membrane protein	30.6%	HP0405	nifS-like protein	27.3%
HP0228	conserved hypothetical integral membrane protein	43.2%	HP1762	conserved hypothetical integral membrane protein	30.6%	HP0221	nifU-like protein	37.3%
HP0234	conserved hypothetical integral membrane protein	32.4%	HP1763	conserved hypothetical integral membrane protein	30.2%	HP0658	PET112-like protein	45.4%
General			HP1764	conserved hypothetical integral membrane protein	30.2%	HP0089	pfs protein (pfs)	34.5%
HP0831	conserved hypothetical ATP binding protein	32.3%	HP1765	conserved hypothetical integral membrane protein	30.2%	HP0322	poly E-rich protein	28.7%
HP0066	conserved hypothetical ATP-binding protein	34.7%	HP1766	conserved hypothetical integral membrane protein	30.2%	HP0625	protein E (gcpE)	47.7%
HP0269	conserved hypothetical ATP-binding protein	37.7%	HP1767	conserved hypothetical integral membrane protein	30.2%	HP0431	protein phosphatase 2C homolog (ptc1)	30.7%
HP0312	conserved hypothetical ATP-binding protein	34.1%	HP1768	conserved hypothetical integral membrane protein	30.2%	HP0624	solute-binding signature and mitochondrial signature protein (aspB)	26.4%
HP1321	conserved hypothetical ATP-binding protein	30.8%	HP1769	conserved hypothetical integral membrane protein	30.2%	HP0377	thiol:disulfide interchange protein (dsbC), putative	26.4%
HP1430	conserved hypothetical ATP-binding protein	38.1%	HP1770	conserved hypothetical integral membrane protein				