

## Systematic Identification of Selective Essential Genes in *Helicobacter pylori* by Genome Prioritization and Allelic Replacement Mutagenesis

ALISON F. CHALKER,<sup>1\*</sup> HEATHER W. MINEHART,<sup>1</sup> NICKY J. HUGHES,<sup>1</sup> KRISTIN K. KORETKE,<sup>2</sup>  
MICHAEL A. LONETTO,<sup>2</sup> KERRY K. BRINKMAN,<sup>3</sup> PATRICK V. WARREN,<sup>2</sup> ANDREI LUPAS,<sup>2</sup>  
MICHAEL J. STANHOPE,<sup>2</sup> JAMES R. BROWN,<sup>2</sup> AND PAUL S. HOFFMAN<sup>1</sup>

Department of Anti-Infective Research<sup>1</sup> and Department of Bioinformatics,<sup>2</sup> SmithKline Beecham Pharmaceuticals, Collegeville, Pennsylvania 19426, and Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802<sup>3</sup>

Received 27 September 2000/Accepted 17 November 2000

**A comparative genomic approach was used to identify *Helicobacter pylori* 26695 open reading frames (ORFs) which are conserved in *H. pylori* J99 but highly diverged in other eubacteria. A survey of selected pathways of central intermediary metabolism was also carried out, and genes with a potentially selective role in *H. pylori* were identified. Forty-five ORFs identified in these two analyses were screened using a rapid vector-free allelic replacement mutagenesis technique, and 33 were shown to be essential in vitro. Notably, 13 ORFs gave essentiality results which are unexpected in view of their known or proposed functions, and phylogenetic analysis was used to investigate the annotation of 7 such ORFs which are highly diverged. We propose that the products of a number of these *H. pylori*-specific essential genes may be suitable targets for novel anti-*H. pylori* therapies.**

The gastric pathogen *Helicobacter pylori* causes one of the most common infections of humans. In developed nations up to half of all individuals are infected with *H. pylori*, and this number rises to more than 90% in many developing nations (22). Although infected individuals are often asymptomatic, *H. pylori* is a causal agent of chronic gastritis in humans and is strongly associated with the development of both duodenal and gastric ulcers and the rare cancer gastric mucosa-associated lymphoid tissue lymphoma. It has also been established that there is a connection between *H. pylori* infection and gastric cancer, which is the second most common neoplastic cause of death worldwide. Various studies have even suggested that *H. pylori* infection has a role in nongastric diseases, including coronary heart disease and hypertension, although supporting evidence in these cases is often rather inconclusive. Eradication therapy is recommended for *H. pylori*-infected ulcer patients (39). Antibiotic monotherapies achieve only a low clearance rate, so the usual treatments are combinations of acid suppressors such as proton pump inhibitors with two antibiotics, often from metronidazole, tetracycline, amoxicillin, or clarithromycin. While they can achieve eradication rates of 85 to 90%, the efficacy of these treatments is being eroded by the rise in frequency of antibiotic-resistant isolates. Although resistance to amoxicillin and tetracycline is rare, substantial and increasing rates of resistance to metronidazole and clarithromycin are a significant cause of treatment failure (1). A further concern is that the use of broad-spectrum drug combinations to treat this chronic infection may promote resistance

in other pathogenic bacteria, thereby compromising their utility in the treatment of more serious infections. As *H. pylori* produces an organ-specific infection which is not normally complicated by coinfection with other pathogens, it would seem ideally suited to treatment with a narrow-spectrum therapeutic agent which would not promote resistance in endogenous flora. Hence, there is considerable interest in developing novel therapies which are both more effective and specific for *H. pylori*.

The search for new ways to eradicate *H. pylori* has been greatly facilitated by the publication of the complete genome sequences of two *H. pylori* strains, 26695 (50) and J99 (2). We have used this information to drive two distinct approaches to identifying essential genes which are potential targets for *H. pylori*-specific antibiotics. Although the genome is only 1.65 Mb in size, it encodes many catabolic and anabolic capabilities which are equivalent to those of bacteria with much larger genomes, suggesting that *H. pylori* must have few functional redundancies and limited metabolic diversity. Hence, it is predicted that a relatively high percentage of genes involved in central metabolic pathways will be essential for its survival even if their direct equivalents in other organisms are not. Although the *H. pylori* 26695 and J99 genomes contain just 1,552 and 1,495 open reading frames (ORFs), respectively, the translational products of more than a third of these are only distantly related to their nearest eubacterial homologs or are of completely unknown function. Elucidation of the cellular roles of these unique genes may lead to the identification of completely novel, essential functions which are *H. pylori* specific. We therefore examined pathways of central intermediary metabolism for genes which are uniquely essential to *H. pylori*, and we also used a bioinformatic genome prioritization analysis to identify highly diverged *H. pylori* ORFs. We describe the

\* Corresponding author. Mailing address: Department of Anti-Infective Research, SmithKline Beecham Pharmaceuticals, 1250 South Collegeville Rd., Collegeville, PA 19426-0989. Phone: (610) 917-6366. Fax: (610) 917-7901. E-mail: Alison\_F\_Chalker@sbphrd.com.

allelic replacement mutagenesis of these genes and discuss the biological implications of our findings.

#### MATERIALS AND METHODS

**Bacterial strains and culture conditions.** *H. pylori* 26695 (ATCC 700392) was obtained from the American Type Culture Collection. *H. pylori* cultures were grown under microaerobic conditions (5% [vol/vol] O<sub>2</sub>, 5% [vol/vol] CO<sub>2</sub>, 90% [vol/vol] N<sub>2</sub>) at 37°C on Columbia agar (Oxoid, Basingstoke, United Kingdom) supplemented with 5% (vol/vol) defibrinated horse blood (Becton Dickinson, Sparks, Md.) and containing amphotericin B, vancomycin, and polymyxin B, each at 10 µg/ml. For selection of transformants, 20 µg of chloramphenicol per ml was also added. Supplements were used during the allelic replacement mutagenesis of particular ORFs as follows: 0.2 to 1 mM uracil with and without 0.3 mM orotidine monophosphate (OMP), 0.3 mM UMP, and/or 0.3 mM orotate (HP1257 [*pyrE*]); 0.01 to 1 mM *p*-aminobenzoic acid and 0.1 to 1 mM tyrosine, 0.02 to 1 mM tryptophan, and 1 to 4 mM phenylalanine (HP1038 [*aroQ*]); and 1 mM each adenine, guanine, xanthine, and hypoxanthine (HP0735 [*gpt*]). We found that increasing the concentrations of the medium supplements used during mutagenesis of HP0735 and HP1257 above these levels was toxic to wild-type cells.

**Identification of highly diverged *H. pylori* ORFs.** A previously described relational bacterial genomic database was used to identify high-priority *H. pylori* gene targets in silico (11). The core of this database is an array of ORF-ORF sequence similarity scores (smallest sum probabilities) across multiple genomes as calculated by the Basic Local Alignment Search Tool (BLASTP) version 2.0 (3). Database queries in SQL language allow data mining of this database for various intergenomic protein evolutionary relationships. Predicted protein sequences of *H. pylori* 26695 were compared to those of *H. pylori* J99 (2), *Campylobacter jejuni* NCTC11168 (41), *Escherichia coli* K-12 (8), *Bacillus subtilis* (36), *Haemophilus influenzae* Rd (24), *Enterococcus faecalis* V583 (preliminary sequence data were obtained from The Institute for Genomic Research [TIGR] website at <http://www.tigr.org>), *Yersinia pestis* CO-92 biovar Orientalis (*Y. pestis* Sequencing Group, Sanger Centre, <ftp://ftp.sanger.ac.uk/pub/pathogens/yp/>), *Klebsiella pneumoniae* 573 (Genome Sequencing Center, Washington University, St. Louis, Mo., personal communication), *Saccharomyces cerevisiae* S288-C (28), *Archeoglobus fulgidis* VC-16 (35), *Methanobacterium thermoautotrophicum* ΔH (45), *Methanococcus janaschii* (12), *Pyrococcus abyssi* 29292 (R. Helig, unpublished data, <http://www.genoscope.cns.fr.>), and *Pyrococcus horikoshii* OT3 (34). *H. pylori* 26695 ORFs were selected for further target evaluation if they showed both highly significant homology to the top hit in *H. pylori* J99 ( $P[N] \leq 1.0e-30$ ) and low homology ( $P[N] \leq 1.0e-15$ ) to any ORF from most other bacterial genomes in the comparison profile.

**Membrane protein predictions.** Average hydropathy profiles ( $\langle H \rangle$ ) (51) were generated using GES hydropathy values (21) weighted using a trapezoid window. Using a process similar to the initial steps of the TopPred II algorithm (16), helical transmembrane segments (TMS) were predicted for each peptide sequence by selecting 19 amino acids centered on the highest ( $\langle H \rangle$ ) value (MaxH), masking these from further consideration, and repeating the process until no peaks with a ( $\langle H \rangle$ ) of >0.5 remained. Subcellular locations were assigned based on the peak MaxH value, number of segments with a ( $\langle H \rangle$ ) of >1.0, and distribution and peak ( $\langle H \rangle$ ) values of the putative TMS. A MaxH cutoff of 1.15 was chosen to maximize the discrimination between two SwissProt release 34 test datasets containing transmembrane and cytoplasmic proteins, respectively (10). Proteins with a MaxH of <1.15 were classified as cytoplasmic, while those with a MaxH of >1.15 and at least three possible TMS were classified as membrane proteins. Anchored proteins were defined as having exactly two TMS, one starting before amino acid (aa) 35 and one having a ( $\langle H \rangle$ ) of >1.15 with the other having a ( $\langle H \rangle$ ) not lower than 0.5. Proteins containing a single N-terminal TMS with a ( $\langle H \rangle$ ) of >1.15 were classified as secreted, although no test for a signal sequence cleavage site was performed. Proteins not fitting into any other category were classified as unknown. This method does not detect gram-negative outer membrane proteins, which do not have multiple transmembrane helices, but outer membrane proteins identifiable by homology were eliminated on the grounds of gene prevalence or protein function at other stages of our analysis.

**Phylogenetic analyses.** Publicly available databases including partial genomic sequences were searched for homologs of the highly divergent *H. pylori* ORFs HP1159 (*fic*), HP0785 (*lolA*), HP1553 (*recB*), HP0020 (*nspC*), HP1257 (*pyrE*), HP0735 (*gpt*), and HP1164 (*trxB*) using BLASTP, TBLASTN, and PSI-BLAST (3). Amino acid sequences were aligned using the program CLUSTALW (49) and then manually edited. Nonconserved indels (highly variable regions which constitute insertions or deletions with reference to the conserved blocks of the alignment) were removed from multiple-sequence alignments prior to phyloge-

netic analysis. Maximum-parsimony (MP) analysis was performed using PAUP\* (47), while neighbor-joining (NJ) trees were based on pairwise distances between amino acid sequences calculated by the programs PROTDIST and NEIGHBOR of PHYLIP version 3.57c (23) (<http://evolution.genetics.washington.edu/phylip.html>). For both methods, confidence limits of branch points were estimated by 1,000 bootstrap replications. In PROTDIST the Dayhoff program option, which estimates the expected amino acid replacements per position using a model based on the Dayhoff 120 matrix, was invoked (18). Trees from all analyses were visualized and, where appropriate, converted into figures using TREEVIEW version 1.6.1 (40).

**Annotation of target ORFs.** We utilized ORF annotations assigned by TIGR (further details on TIGR gene annotations can be obtained at <http://www.tigr.org>), with the following exceptions. The *oor* (HP0588-0591) and *por* (HP1108-1111) genes and HP0086 (*mgo*) are annotated as described by subsequent authors (31, 32). HP1138 is annotated as *spoOJ* based on the homology of the translational product to *B. subtilis* SpoJ and its proximity to a *soj* homolog, HP1139. HP1563 is annotated as *ahpC*, which is the more usual nomenclature for bacterial alkyl hydroperoxide reductase enzymes. HP1553 and HP0785 are annotated as *recB* and *lolA*, respectively, based on phylogenetic analyses performed during this work (see Results and Discussion).

**Generation of *H. pylori* allelic replacement mutants.** An invariant resistance cassette containing a *Campylobacter coli* chloramphenicol resistance gene (*cat*) (53) was PCR amplified using primers designated C1 (5'-GATATAGATTGAA AAGTGATGAT-3') and C2 (5'-TTATCAGTGCACAAACTGGG-3'). *H. pylori* 26695 genomic DNA was isolated by washing cells scraped from five 48-h confluent plates, resuspending them in 3.5 ml of Qiagen (Valencia, Calif.) buffer B1, and then using Qiagen Genomic DNA kits as per the manufacturer's instructions. Two pairs of gene-specific primers designated P1-P2 and P3-P4 were used to PCR amplify the up- and downstream regions, respectively, of each target gene using *Pwo* DNA polymerase (Roche, Mannheim, Germany), in order to produce fragments on the order of 500 bp flanking the region to be deleted. P2 and P3 contained 5' leaders complementary to C1 and C2, respectively, followed by 20 bp of gene-specific sequence. Each PCR product was purified using Qiaquick PCR purification cartridges. A template mixture containing 1 µg of each of the three purified products was then PCR amplified using primers P1 and P4 in a single reaction, to generate a linear construct in which the up- and downstream fragments are fused to the central *cat* gene. In order to maximize product recovery, the PCR conditions used, following a standard hot start, were 2 cycles of 94°C for 1 min, 45°C for 1 min, and 72°C for 5 min, followed by 40 cycles of 94°C for 45 s, 40°C for 1 min, and 72°C for 3 min, and finally 70°C for 10 min. The resulting allelic replacement construct was confirmed by sequencing and introduced into *H. pylori* 26695 by transformation (52), by scraping cells from 24-h confluent plates into 0.2 ml of brucella broth (Becton Dickinson) and spotting 100 µl of this high-density inoculum onto Columbia agar plates made as described above. After incubation with 2 µg of DNA for 12 h, the transformed cell mix was spread over the plates and incubated for a further 24 h. The cells were then scraped onto selective Columbia agar plates containing 20 µg of chloramphenicol per ml and incubated for 3 to 7 days.

For nonessential genes, typical frequencies of mutant recovery (representing the combined frequencies of transformation and double-crossover recombination) were  $2 \times 10^2$  to  $4 \times 10^3$  CFU/µg of transforming DNA. In these cases, the chloramphenicol-resistant (Cm<sup>r</sup>) transformant colonies were restreaked on selective plates. In order to confirm that the correct chromosomal rearrangement had occurred, transformants were initially screened by diagnostic PCR. Extracts were prepared by resuspending a loopful of cells in 20 µl of alkaline sodium dodecyl sulfate (0.05 M NaOH, 0.25% sodium dodecyl sulfate), incubating at 100°C for 10 min, and adding 200 µl of water (based on a method provided by J. Versalovic, personal communication). Two microliters of extract was used as the PCR template. The diagnostic PCRs comprised sets of DNA primers designed to hybridize within the *cat* cassette paired with primers hybridizing to distal chromosomal sequences to generate DNA products of characteristic size. In addition, primers designed to hybridize within the intact target gene were used to confirm that it had been deleted. Chromosomal DNA was isolated from selected clones which appeared to contain the correct mutation and further examined by diagnostic PCR and Southern blot analysis. Genes were described as nonessential if one or more mutants carrying a confirmed allelic replacement mutation of the target gene were isolated using these methods.

If no Cm<sup>r</sup> colonies were obtained in comparison with parallel control allelic replacement experiments targeting nonessential genes, the transformation was repeated at least twice more. Genes were described as essential if all three transformations were negative in comparison to the control.

If Cm<sup>r</sup> colonies were obtained but DNA analysis (see above) demonstrated that they contained an intact copy of the wild-type gene in addition to the correct

mutation, the transformation was repeated at least twice more, and DNA analysis of 20 to 30 putative mutants was performed. In these cases mutant recovery frequency tended to be much lower than for nonessential genes. Genes were described as essential if no mutants with the wild-type gene deleted could be identified.

## RESULTS AND DISCUSSION

**Identification of ORFs unique to *H. pylori*.** In order to identify *H. pylori*-specific ORFs for essentiality testing, we performed a global bioinformatics analysis of the *H. pylori* 26695 genome using a series of prioritization steps based on various biological criteria. *H. pylori* strains are genetically diverse, and a previous comparison of the *H. pylori* 26695 and J99 genomes showed that approximately 7% of the genes are strain specific (2), so a requirement for the chosen ORFs was that they be present in both available genomes at a BLASTP confidence level of  $1.0e-30$  or less (3). For the same reason, genes associated with the variable *cag* pathogenicity island (14) were excluded. The translational products of ORFs common to both strains were screened against 13 other microbial genomes to identify unique and highly diverged ORFs. An ORF was defined as having a homolog in another genome if the top hit had a BLASTP score equal to or less than  $1.0e-15$ . Homology to archaeobacterial or yeast ORFs was recorded for information but was not used as an exclusion criterion. *H. pylori* 26695 ORFs were retained if they were either unique or had homologs only in the related organism *C. jejuni* or in only one or two additional eubacterial genomes. These first two steps reduced the number of target ORFs from 1,552 to 606. TMS predictions using hydropathy profiles showed that *H. pylori* 26695 encodes 1,221 cytoplasmic proteins, while the remaining 21% of ORF translational products are either noncytoplasmic or of unknown subcellular localization (*E. coli* contains 27% noncytoplasmic proteins). Since membrane proteins tend to be poorly soluble outside the membrane and therefore less amenable to in vitro biochemical study, ORFs encoding probable integral membrane proteins were excluded from the gene set (with the exception of HP0740 [*murF*]), further reducing the number of target ORFs to 505. ORFs whose only detectable homologs were genes of unknown or hypothetical function were also excluded, since the absence of clues to function would make biochemical exploitation difficult, leaving 172 ORFs. Finally, ORFs known to encode proteins involved in nonessential processes (such as flagellar biosynthesis, hemolysis, urease production, capsular biosynthesis, and restriction-modification), or whose highest homologs in the public databases were genes encoding such proteins, were excluded. From the remaining 73 highly diverged potential target ORFs, 29 were chosen at random for essentiality testing.

**Identification of key genes in central intermediary metabolism.** Biochemical and in silico pathway reconstruction studies have indicated that there are gaps in some *H. pylori* central intermediary metabolic pathways. We selected 13 ORFs encoding key enzymes in these pathways for further study. These included HP1099 (*eda*) and HP1100 (*edd*), which are thought to encode the Entner-Doudoroff (ED) pathway enzymes 2-keeto-3-deoxy-6-phosphogluconate aldolase (EC 4.1.2.14) and 6-phosphogluconate dehydratase (EC 4.2.1.12), respectively; HP1385 (*fbp*), which is thought to encode the Embden-Meyerhof pathway enzyme fructose-1,6-bisphosphatase (EC 3.1.3.11);

HP0086 (*mgo*), encoding malate quinone oxidoreductase (EC 1.1.99.16); and HP0509 (*glcD*), which is thought to encode glycolate oxidase subunit (EC 1.1.3.15). We also selected genes which encode the Krebs cycle enzymes pyruvate:flavodoxin oxidoreductase (EC 1.2.7.1) (HP1108-1111 [*porCDAB*]) and 2-oxoglutarate oxidoreductase (EC 1.2.7.3) (HP0588-0591 [*oorDABC*]). The *porB* (HP1111) and *oorA* (HP0590) genes have previously been shown to be essential (31). Some of these ORFs (*oorDABC* [HP0588-0591], *porCDAB* [HP1108-1111], and *mgo* [HP0086]) are highly diverged and also ranked highly in our global genome bioinformatic analysis, but HP0509 (*glcD*), HP1099 (*eda*), HP1100 (*edd*), and HP1385 (*fbp*) are moderately or highly conserved in other bacterial species.

**Design of allelic replacement mutagenesis cassettes.** As a result of these analyses, 29 highly diverged ORFs and 13 ORFs encoding key products in central intermediary metabolism were chosen for further study. Three more genes whose essentiality or nonessentiality in *H. pylori* has already been reported in the literature (HP1027 [*fur*], HP1021 [response regulator], and HP1364 [histidine kinase]) were also included as controls. Allelic replacement constructs targeting each of these 45 ORFs were made by PCR amplification. In order to minimize potential polar effects, oligonucleotide primers were chosen so that flanking genes and intergenic regions including potential promoters would remain intact in the deletion mutant. The translation start sites of each ORF assigned by TIGR were checked for optimal spacing from potential ribosome binding sites, and where possible potential promoter sequences conforming to the *H. pylori* -10 region consensus TATAaT were identified (25). Flanking genes were identified and in most cases were consistent with the gene order given by TIGR. However for certain ORFs with several possible start sites and those adjacent to short unannotated ORFs, the most conservative option was used when designing the primers.

Six of the target ORFs (HP0616, HP0372, HP0785, HP1164, HP1563, and HP1027) are monocistronic, making it extremely unlikely that replacement of these genes would have a polar effect on expression of neighboring genes. However, the rest appear to be transcriptionally or, in 16 cases, translationally coupled with one or both of their neighbors. Hence, in addition to the precautions we have described, transcriptional termination signals were removed from the chloramphenicol resistance gene marker (*cat*), and the cassettes were designed to integrate in the same orientation as the target genes in order to ensure transcription of the downstream region. The success of our approach is demonstrated by the fact that we detected essential and nonessential genes of both the monocistronic and polycistronic types. Furthermore, in at least one case we detected both an essential (HP1140 [*birA*]) and a nonessential (HP1138 [*spo0J*]) gene in the same operon (*fmt birA soj spo0J*).

**Allelic replacement mutagenesis of potential target genes.** In order to determine gene essentiality, the allelic replacement constructs were directly transformed into *H. pylori* 26695. Typically, essential genes were so annotated following three independent, controlled transformations in which no mutants were recovered. However, in the case of 14 genes which we have designated essential (HP0020, HP0509, HP0588, HP0589, HP0590, HP0808, HP0975, HP1108, HP1109, HP1418, HP1563, HP1553, HP0740, and HP1159), a small number of chloramphenicol-resistant colonies was recovered following transformation with the al-

elic replacement cassette. In each case PCR and Southern blot analysis of 20 to 30 putative mutants demonstrated that they contained an apparently intact wild-type copy of the target gene in addition to the correct mutation, suggesting that only mutants which contain a functional second gene copy are viable. During allelic replacement mutagenesis studies for a further three ORFs (HP1257 [*pyrE*], HP1038 [*aroQ*], and HP0735 [*gpt*]), the transformed cells were plated onto media with and without nutritional supplements appropriate to the proposed gene function. A few chloramphenicol-resistant *aroQ* and *gpt* putative mutant colonies were obtained on supplemented media, but further subculture demonstrated that they did not require the presence of supplements for growth, and they too were found to contain a copy of the wild-type target gene. This type of gene duplication has been observed during similar allelic replacement mutagenesis studies with *Streptococcus pneumoniae* (15), and it was postulated that this might be explained by the stabilization under selective pressure of spontaneous tandem unstable chromosomal duplications, similar to those which have been reported to occur between repetitive sequences at moderately high frequency in *E. coli* and *Salmonella enterica* serovar Typhimurium (29). The essentiality results for all 45 ORFs tested are shown in Table 1.

**Control allelic replacement studies.** Four of the five genes whose essentiality in *H. pylori* has previously been reported behaved as expected in our allelic mutagenesis system. We confirmed the essentiality of *porB* (HP1111) and *oorA* (HP0589) (31) and additionally demonstrated that each of the six remaining *por* and *oor* subunit genes is individually required for cell viability in vitro. We also recovered and confirmed mutants of the nonessential genes HP1027 (*fur*) (7) and HP1364 (histidine kinase) (6). However, we found that the orphan response regulator gene HP1021 is not essential for viability, in contrast to a previous study in which putative HP1021 mutant colonies were recovered but could not be further passaged (6). It is possible that the discrepancy is due to some difference in our respective protocols, and indeed we note that those authors used clinical isolate strains as mutation hosts and a more oxygen-rich atmosphere (5% CO<sub>2</sub>, 95% air) for cell culture. However we were unable to test the hypothesis that HP1021 is essential only under conditions of higher oxygen tension, as in our hands neither the HP1021 mutants nor wild-type *H. pylori* 26695 grew under these conditions.

**Overview of target gene essentiality data.** Of the remaining 40 ORFs, 31 were essential and 9 were nonessential in vitro. In five cases (HP1140 [*birA*], HP0657 [*ymxG*], HP0830 [*gataA*], HP0658 [*gatB*], and HP0975 [*gatC*]) this analysis represents to our knowledge the first description of essentiality of the gene or its homologs in any bacterial species. In most other cases the essentiality data are consistent with the proposed function of the ORF. The results for these 27 ORFs are listed in Table 1 together with their putative function but are not further discussed. However, the behavior of 13 ORFs in our allelic mutagenesis system (shown in boldface in Table 1) was unexpected based on existing knowledge of the gene products or their orthologs. Six of these ORFs either have been biochemically characterized in *H. pylori* (HP0086 [*mgo*] and HP1038 [*aroQ*]), or are well conserved among eubacteria (HP1099 [*eda*], HP1100 [*edd*], HP1385 [*fbp*], and HP1563 [*ahpC*]). The remaining seven (HP1257 [*pyrE*], HP0735 [*gpt*], HP1553 [*recB*],

HP1159 [*fic*], HP1164 [*trxB*], HP0020 [*nspC*], and HP0785 [*lolA*]) are highly diverged, so their primary structural relationships were investigated using phylogenetic analysis.

**Essentiality of *H. pylori* genes of known function.** HP0086 has been shown in complementation studies to encode malate:quinone oxidoreductase (Mqo) (32). Although malate dehydrogenase activity has been reported in *H. pylori* extracts (43), no *mdh* gene has been detected in the *H. pylori* genomes, so our finding that HP0086 is essential is consistent with the argument that *H. pylori* is solely dependent on Mqo for malate oxidation.

The product of HP1038 (*aroQ*) has been purified and confirmed to be a type II dehydroquinase from the shikimate pathway (9). Disruption of shikimate pathway genes is well documented in other bacteria and has been shown to cause auxotrophy for aromatic amino acids and *p*-aminobenzoic acid, but we were unable to obtain HP1038 mutants on rich media even when these supplements were supplied exogenously. It therefore appears either that *H. pylori* is impaired in its ability to take up these supplements or that elimination of the shikimate pathway is lethal. Chorismate is a precursor for synthesis of the quinone components of the electron transport chain, and our results could be explained if *H. pylori* cannot obtain these molecules from the horse blood-supplemented Columbia plates used here.

**Essentiality of conserved genes.** The ED pathway is responsible for catabolism of glucose, the only carbohydrate utilized by *H. pylori* (37). The pathway is constitutively expressed in *H. pylori*, and genes for key enzymes in the alternative glycolytic and pentose phosphate pathways (phosphofructokinase, pyruvate kinase, and 6-phosphogluconate dehydrogenase) cannot be detected in the genome, suggesting that it has a particular importance in this organism. The ED enzymes 6-phosphogluconate dehydratase and 2-keto-3-deoxy-6-phosphogluconate aldolase have been detected in *H. pylori* extracts (38) and are thought to be encoded by HP1099 (*edd*) and HP1100 (*eda*), respectively. Neither gene is essential in *E. coli*, although the accumulation of 2-keto-3-deoxy-6-phosphogluconate in *eda* mutants has a bacteriostatic effect on growth (26). The fact that we were unable to recover mutants of HP1099 or HP1100 is consistent with the hypothesis that catabolism of glucose in *H. pylori* occurs exclusively via the ED pathway.

The conserved ORF HP1385 is thought to encode fructose-1,6-bisphosphatase (Fbp). The relative activities in *H. pylori* extracts of this enzyme and the phosphofructokinase enzyme, which catalyzes the reverse reaction, suggest that gluconeogenesis is favored over glycolysis (30), and indeed no phosphofructokinase gene has been detected in the genome (37). The *fbp* gene is not essential in *E. coli* or *B. subtilis* (27), but in *Pseudomonas aeruginosa*, where Fbp is additionally required for catabolism of glyceraldehyde-3-phosphate, *fbp* null mutants could not be recovered (5). We found that HP1385 is essential, suggesting that Fbp may have a similarly enhanced cell role in *H. pylori*.

The conserved ORF HP1563 is thought to encode the AhpC subunit of alkylhydroperoxide reductase (AhpR), which catalyzes the reduction of organic hydroperoxides and hydrogen peroxide. *ahpC* is not essential in *C. jejuni* (4), but as *C. jejuni* *ahpC* mutants are hypersusceptible to oxidative stress, it is possible that HP1563 mutants were generated but were not

TABLE 1. *H. pylori* target ORFs and essentiality testing results<sup>a</sup>

Category	HP no. <sup>b</sup>	Gene <sup>c</sup>	Function <sup>d</sup>	Product length (aa)	Spectrum <sup>e</sup>	Localization <sup>f</sup>	Essentiality <sup>g</sup>	
Allelic replacement controls	<b>1021</b>	<b>RR</b>	Response regulator	298	<i>C. jejuni</i>		C	
	1027	<i>fur</i>	Ferri uptake regulator	150	<i>C. jejuni</i> , <i>E. coli</i> , <i>H. influenzae</i>		C	
	1364	HK	Histidine kinase	397	<i>C. jejuni</i>		A	
	1111	<i>porB</i>	Pyruvate oxidoreductase subunit B	314	Archaeobacteria		C	
	589	<i>oorA</i>	2-Oxoglutarate oxidoreductase subunit A	375	<i>C. jejuni</i> , archaeobacteria		E	
	<b>1553</b>	<b>recB*</b>	Helicase	945	<i>C. jejuni</i>		C	
Nucleic acid metabolism	1231	<i>holB</i>	DNA polymerase III δ' subunit	218	—		E	
	1138	<i>spoII*</i>	Partition-related protein	290	<i>C. jejuni</i> , <i>B. subtilis</i>		C	
	1448	<i>mmpA</i>	RNase P protein component	161	—		C	
	440	<i>topA</i>	Topoisomerase I	677	<i>C. jejuni</i> , <i>B. subtilis</i> , <i>H. influenzae</i> , <i>E. coli</i> , archaeobacteria		C	
	<b>785</b>	<b>lalA*</b>	Lipoprotein chaperone	184	—		C	
	830	<i>gatA</i>	Glu-tRNA <sup>Gln</sup> amidotransferase subunit A	453	<i>C. jejuni</i> , <i>B. subtilis</i> , <i>S. cerevisiae</i> , archaeobacteria		C	
Protein synthesis and modification	658	<i>gatB</i>	Glu-tRNA <sup>Gln</sup> amidotransferase subunit B	475	<i>C. jejuni</i> , <i>B. subtilis</i> , <i>S. cerevisiae</i> , archaeobacteria		C	
	975	<i>gatC</i>	Glu-tRNA <sup>Gln</sup> amidotransferase subunit C	93	—		C	
	657	<i>ympG</i>	Putative processing protease	432	<i>C. jejuni</i> , <i>B. subtilis</i>		S	
	470	<i>pepF</i>	Oligoendopeptidase F	578	<i>C. jejuni</i> , <i>B. subtilis</i>		C	
	1140	<i>btzA</i>	Biotin acetyl coenzyme A carboxylase	212	<i>C. jejuni</i>		C	
	74	<i>lspA</i>	Lipoprotein signal peptidase	157	<i>C. jejuni</i>		S	
Cell wall biosynthesis and cell division	<b>1159</b>	<b>fic</b>	Cell filamentation protein	177	<i>H. influenzae</i>		C	
	740	<i>murF</i>	UDP-MurNAc-pentapeptide presynthetase	493	<i>C. jejuni</i> , <i>B. subtilis</i>		M	
	1372	<i>mreC</i>	Rod-shape-determining protein	248	<i>C. jejuni</i>		S	
	1418	<i>murB</i>	UDP-N-acetylenolpyruvyl glucosamine reductase	259	<i>C. jejuni</i> , <i>B. subtilis</i>		C	
Energy metabolism	<b>1099</b>	<b>eda</b>	2-Keto-3-deoxy-6-phosphogluconate aldolase	208	<i>B. subtilis</i> , <i>H. influenzae</i> , <i>E. coli</i>		C	
	<b>1100</b>	<b>edd</b>	6-Phosphogluconate dehydratase	608	<i>C. jejuni</i> , <i>B. subtilis</i> , <i>H. influenzae</i> , <i>E. coli</i> , archaeobacteria, <i>S. cerevisiae</i>		C	
Intermediate metabolism	<b>1385</b>	<b>fbp</b>	Fructose-1,6-bisphosphatase	290	<i>C. jejuni</i> , <i>H. influenzae</i> , <i>E. coli</i> , <i>S. cerevisiae</i>		C	
	<b>86</b>	<b>mgo</b>	Malate:quinone oxidoreductase	450	<i>C. jejuni</i>		C	
	588	<i>oorD</i>	2-Oxoglutarate oxidoreductase subunit D	113	<i>C. jejuni</i>		C	
	590	<i>oorB</i>	2-Oxoglutarate oxidoreductase subunit B	273	<i>C. jejuni</i> , archaeobacteria		C	
	591	<i>oorC</i>	2-Oxoglutarate oxidoreductase subunit C	186	<i>C. jejuni</i> , archaeobacteria		C	
	1108	<i>porC</i>	Pyruvate oxidoreductase subunit C	186	Archaeobacteria		C	
	1109	<i>porD</i>	Pyruvate oxidoreductase subunit D	130	—		C	
	1110	<i>porA</i>	Pyruvate oxidoreductase subunit A	407	<i>C. jejuni</i> , <i>E. coli</i> , archaeobacteria		C	
	509	<i>glcD</i>	Glycolate oxidase	459	<i>C. jejuni</i> , <i>B. subtilis</i> , <i>H. influenzae</i> , <i>E. coli</i> , archaeobacteria, <i>S. cerevisiae</i>		U	
	1495	<i>tal</i>	Transaldolase	316	<i>C. jejuni</i>		C	
	Response to exogenous stimuli	<b>735</b>	<b>spt</b>	Xanthine-guanine PRTase	153	<i>C. jejuni</i>		C
		<b>1257</b>	<b>pyrE</b>	Orotate PRTase	201	<i>C. jejuni</i>		C
<b>1038</b>		<b>aroQ</b>	3-Dehydroquinate	167	<i>C. jejuni</i> , <i>B. subtilis</i> , <i>H. influenzae</i>		U	
<b>20</b>		<b>nspC</b>	Carboxymnospermidine decarboxylase	405	<i>C. jejuni</i>		C	
832		<i>speE</i>	Spermidine synthase	262	—		C	
801		<i>moad</i>	Molybdopterin-converting factor subunit 1	74	<i>C. jejuni</i>		C	
372		<i>dcd</i>	Deoxycytidine triphosphate deaminase	190	<i>C. jejuni</i>		C	
808		<i>acpS</i>	Holo-acyl carrier protein synthase	119	<i>C. jejuni</i>		C	
<b>1164</b>		<b>trxB</b>	Thioredoxin reductase	324	<i>C. jejuni</i>		C	
<b>1563</b>		<b>ahpC*</b>	Alkyl hydroperoxide reductase subunit	198	<i>C. jejuni</i> , <i>E. coli</i> , <i>B. subtilis</i> , <i>S. cerevisiae</i>		C	
616	<i>cheY</i>	Chemotaxis protein	313	<i>C. jejuni</i> , <i>B. subtilis</i>		C		

<sup>a</sup> ORFs in boldface are discussed in more detail in the text.

<sup>b</sup> *H. pylori* 26695 locus number assigned by TTGR.

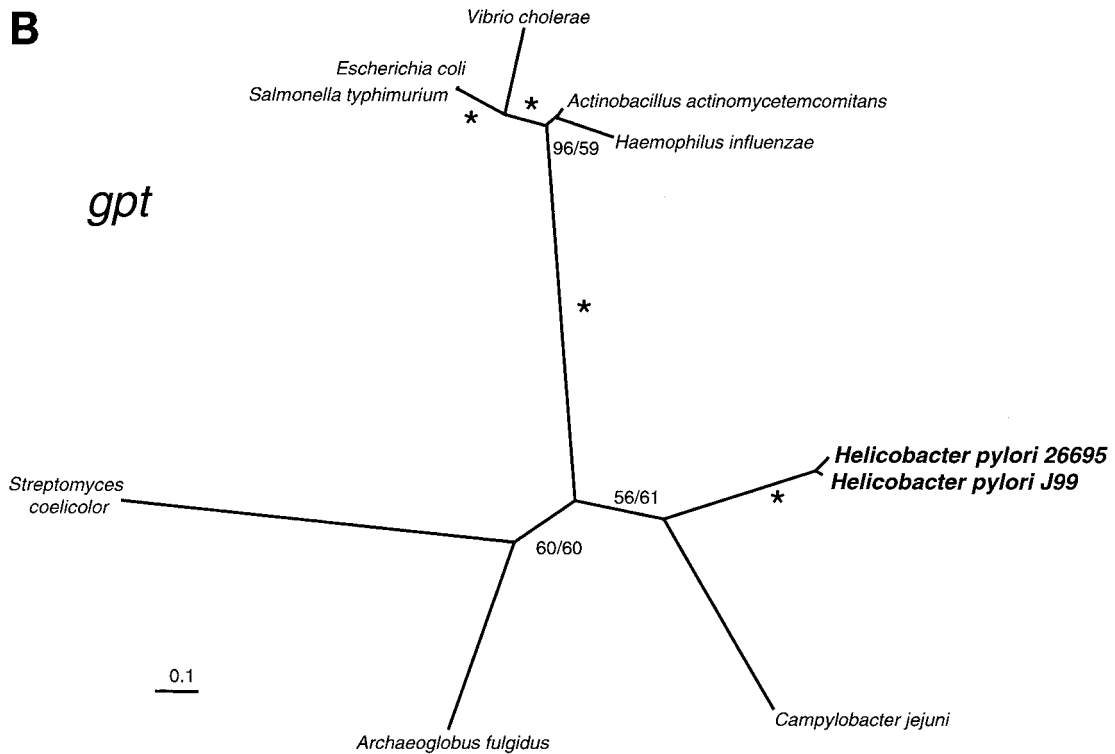
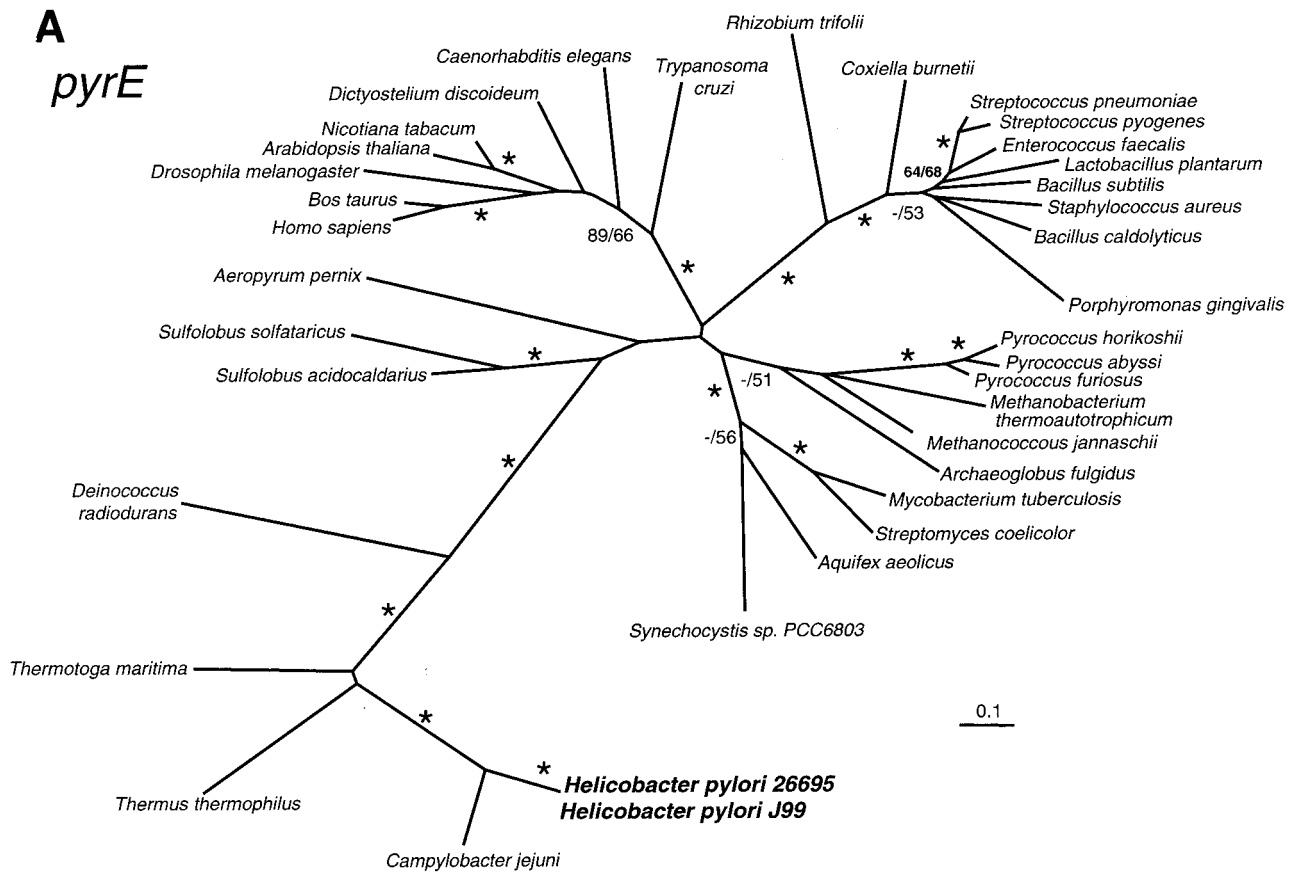
<sup>c</sup> An asterisk indicates that the gene name used in this study differs from previous annotations (see text for details).

<sup>d</sup> Function of predicted gene product according to gene annotation.

<sup>e</sup> Genomes screened which contain homologs of the gene with a confidence score of better than 1.0e-15. Archaeobacteria are listed in the text. —, no hits at this confidence level.

<sup>f</sup> C, cytoplasmic; A, anchored; S, secreted; U, unknown; M, membrane bound.

<sup>g</sup> E, essential in vitro; NE, not essential in vitro.



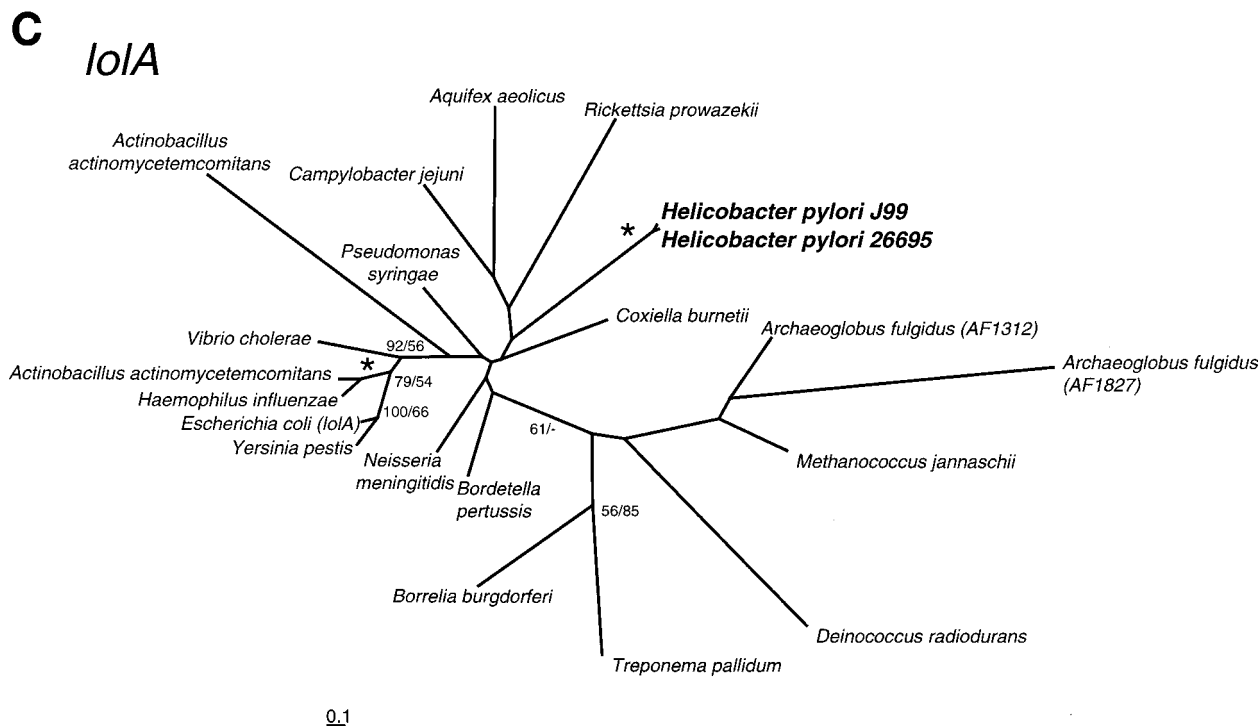


FIG. 1. NJ phylogenetic trees, which show the divergence of *H. pylori* genes HP1257 (*pyrE*) (A), HP0735 (*gpt*) (B), and HP0785 (*lolA*) (C). The numbers at nodes represent the percent occurrence of nodes in 1,000 bootstrap resamplings of the MP and NJ trees. The NJ trees shown here were largely congruent with MP trees constructed using the program PAUP\* (47). Nodes occurring in more than 70% of resamplings by both methods are marked with an asterisk; values of <50% are marked with a dash or not shown. The scale bar represents 0.1 expected amino acid replacement per site as estimated by PROTDIST using the Dayhoff option. The trees were drawn using the program NEIGHBOR from PHYLIP version 3.57c (22).

able to proliferate in the 5% O<sub>2</sub> microaerobic atmosphere used. It therefore appears either that the mechanisms for controlling oxidative stress in *H. pylori* are exquisitely sensitive to oxygen concentration or that they play an unusually critical role in cell viability.

**Essentiality of highly diverged ORFs.** Phylogenetic analysis (Fig. 1A) supports the annotation of HP1257 as *pyrE*, encoding orotate phosphoribosyl transferase (PRTase), although it is very distantly related to other *pyrE* genes. This enzyme catalyzes the production of OMP, which is a precursor for UMP, a central intermediate in pyrimidine nucleotide metabolism. Bacterial *pyrE* mutants are usually uracil auxotrophs (54), but we were unable to obtain HP1257 mutants on rich media in the presence of exogenous uracil, even with supplemental OMP, UMP, and orotate. *H. pylori* has no homologs of the pyrimidine salvage enzymes uracil PRTase (Upp) or uridine kinase (Udk), which convert uracil to UMP (37). Although our results could be explained by poor uptake of medium supplements, they are consistent with recent reports that the conserved gene HP1084 (*pyrB*), encoding aspartate carbamoyltransferase, is also essential (13) and that inhibitors of dihydroorotate dehydrogenase (*PyrD*) are lethal to *H. pylori* (17). Taken together these data appear to demonstrate that de novo pyrimidine biosynthesis is required for survival of *H. pylori*.

HP0735 (*gpt*) is thought to encode xanthine/guanine PRTase, which is involved in the salvage of nucleotides and nucleosides. Phylogenetic analysis (Fig. 1B) shows that the product of

HP0735 is highly diverged from other proteobacterial Gpt enzymes. *H. pylori* does not appear to encode hypoxanthine PRTase (*hpt*), so it is possible that in this organism Gpt has a relaxed substrate specificity (37). *E. coli gpt* and *gpt hpt* mutants are viable (20), but no HP0735 mutants were recovered on rich media even in the presence of exogenous adenine, guanine, xanthine, and hypoxanthine. It therefore appears either that the medium supplements were not taken up or that nucleotide salvage has a particular importance for the survival of *H. pylori*.

Phylogenetic analysis (not shown) demonstrates that the helicase encoded by the essential gene HP1553 is a homolog of RecB, a subunit of the ExoV recombinase. The protein is clearly identifiable by sequence motifs specific to members of the RecB helicase subfamily, including the conserved C-terminal region at aa 876 to 901. This is surprising both because *H. pylori* does not appear to encode the other ExoV subunits RecC and RecD and also because ExoV is not an essential enzyme. Essential helicases have been described for other bacteria, including DnaB in *E. coli* and DnaB and PcrA in *B. subtilis* (42). Our finding suggests that despite its RecB-like structure and the fact that *H. pylori* encodes several other helicases, including the products of HP1478 (*uvrD*) and HP0911, the HP1553 helicase plays a critical role in cell function.

The Fic protein regulates cell division, although its exact function is unknown. Phylogenetic analysis (not shown) demonstrates that the products of HP1159 (*fic*) and *E. coli fic* are

orthologous but highly divergent. *E. coli* mutants are viable, although they form shorter rods than wild-type cells (33), so our finding that HP1159 is essential suggests that there may be less functional redundancy among cell division proteins in *H. pylori* than in *E. coli*.

Reduced thioredoxin is required for cellular processes such as hydrogen peroxide reduction and reduction of catalytic disulfide bonds in cytoplasmic proteins. *H. pylori* 26695 contains two genes annotated as *trxB* (thioredoxin reductase), HP0825 and HP1164. The two proteins are only 16% identical, but HP0825 is more closely related to known thioredoxin reductases and forms an operon with *trxA* (HP0824). In *E. coli* *trxB* mutants, disulfide bond formation is controlled by TrxA and TrxC (46), but we found that HP1164 is essential. Since some thioredoxin reductases show limited homology to AhpF, the electron transfer subunit of AhpR (44), and as *H. pylori* does not encode an AhpF homolog, it is possible that HP1164 fulfills this function. As discussed above, the mechanisms for controlling oxidative stress in *H. pylori* appear to be unusually critical for cell viability.

Carboxynorspermidine decarboxylase (NspC) catalyzes the last step in synthesis of norspermidine. Our finding that HP0020 (*nspC*) is essential suggests that *H. pylori*, like members of the genus *Vibrio*, may utilize norspermidine as its primary polyamine (55). Alternatively, as *H. pylori* does not encode homologs of the SpeB, -C, or -D enzymes, which provide precursors for spermidine synthase SpeE, and as we have shown that HP0832 (*speE*) is not essential, it is possible that NspC is solely responsible for spermidine biosynthesis (19).

All of the ORFs described above are essential in *H. pylori* even though their orthologs in other organisms are nonessential. HP0785 is the only example of a nonessential ORF which might be expected to be essential. The product of HP0785 is homologous to the periplasmic lipoprotein chaperone LolA, although phylogenetic analysis (Fig. 1C) shows that the *H. pylori* protein is highly divergent from its proteobacterial equivalents. LolA localizes outer membrane-specific lipoproteins following their translocation to the periplasmic side of the inner membrane, in conjunction with the LolB receptor protein. The *lolA* gene is essential in *E. coli* (48). Our result that the lipoprotein signal peptidase gene HP0074 is essential demonstrates that lipoproteins are important for *H. pylori* survival, so our finding that HP0785 is not essential suggests that *H. pylori* may encode more than one such chaperone.

**Perspectives.** The data presented here demonstrate that stepwise prioritization of genome ORFs using simple biological criteria, particularly in the case of small genomes such as that of *H. pylori*, can be an effective way of quickly reducing the number of genes of interest to an experimentally manageable number. The essentiality testing of a relatively large number of gene targets was accomplished by the application in *H. pylori* of a rapid and robust vector-free allelic replacement mutagenesis method. When combined with a set of initial selection criteria which are directly relevant to antibacterial drug discovery, this process is an efficient way to enrich for potential target genes and identify those which are critical for normal cell function.

The essentiality results reported for most of the ORFs are consistent with their proposed functions. However, we have identified 12 essential genes which might be expected to be

nonessential by analogy with orthologous genes in other bacteria, including two genes whose products have previously been directly characterized in *H. pylori* (*aroQ* and *mgo*) and one nonessential ORF (*lolA*) whose closest homolog is essential in *E. coli*. It is possible that the most highly diverged of these ORFs have been misannotated, but in general our phylogenetic analyses support the ORF annotation given by TIGR. The exceptions are HP0785 and HP1553, which we have designated *lolA* and *recB*, respectively, but this refinement of the functional annotation has not clarified the reason for their essentiality. These results reflect the fact that *H. pylori* has a relatively small genome with little room for functional redundancy and must therefore contain a higher proportion of essential genes with core functions.

The high degree of sequence divergence of many of the essential gene products described here suggests that it may be possible to design specific chemical inhibitors which do not affect the growth of other organisms. This is recognized as a desirable attribute for any new anti-*H. pylori* therapeutic agent, as it would reduce the severity of gastrointestinal sequelae caused by destruction of the normal flora, and also minimize the spread of drug resistance to other pathogens. Importantly, a further degree of selectivity may be conferred by the fact that the nearest neighbors of some of these genes are not required for viability in other bacteria. The potential of this type of approach has been graphically illustrated by the recent identification of pyrazole-based anti-*H. pylori* compounds which specifically inhibit the pyrimidine biosynthesis enzyme PyrD in *H. pylori* but are inactive against orthologous enzymes in other bacteria and humans (17). We therefore suggest that the products of these highly diverged essential genes merit further investigation as potential targets for novel, highly specific anti-*H. pylori* agents.

#### ACKNOWLEDGMENTS

We are grateful to John Throup for useful discussions, Peter Morrison for assistance with phylogenetic analysis, Thomas Mathie for synthesis of oligonucleotide primers, and Stephanie Van Horn for sequence confirmation. We thank the Genome Sequencing Center, Washington University, St. Louis, Mo., for communication of *K. pneumoniae* DNA sequence data prior to publication.

Sequencing of *H. pylori* 26695 was accomplished with support from TIGR. Sequencing of *E. faecalis* V583 was accomplished with support from NIAID.

#### REFERENCES

- Alarcon, T., D. Domingo, and M. Lopez-Brea. 1999. Antibiotic resistance problems with *Helicobacter pylori*. *Int. J. Antimicrob. Agents* 12:19-26.
- Alm, R. A., L.-S. L. Ling, D. T. Moir, N. L. King, E. D. Brown, P. C. Doig, D. R. Smith, B. Noonan, B. C. Guild, B. L. deJonge, G. Carmel, P. J. Tummino, A. Caruso, M. Uria-Nickelson, D. M. Mills, C. Ives, R. Gibson, D. Merberg, S. D. Mills, Q. Jiang, D. E. Taylor, G. F. Vovis, and T. J. Trust. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 397:176-180.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.
- Baillon, M.-L. A., A. H. M. van Vliet, J. M. Ketley, C. Constantinidou, and C. W. Penn. 1999. An iron-regulated alkyl hydroperoxide reductase (AhpC) confers aerotolerance and oxidative stress resistance to the microaerophilic pathogen *Campylobacter jejuni*. *J. Bacteriol.* 181:4798-4804.
- Banerjee, P. C. 1989. Fructose-bisphosphatase-deficient mutants of mucoid *Pseudomonas aeruginosa*. *Folia Microbiol.* 34:81-86.
- Beier, D., and R. Frank. 2000. Molecular characterization of two-component systems of *Helicobacter pylori*. *J. Bacteriol.* 182:2068-2076.
- Bereswill, S., F. Lichte, S. Greiner, B. Waidner, F. Fassbinder, and M. Kist. 1999. The ferric uptake regulator (Fur) homologue of *Helicobacter pylori*:



- functional analysis of the coding gene and controlled production of the recombinant protein in *Escherichia coli*. *Med. Microbiol. Immunol.* **188**:31–40.
8. Blattner, F. R., G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**:1453–1474.
  9. Bottomley, J. R., C. L. Clayton, P. A. Chalk, and C. Kleanthous. 1996. Cloning, sequencing, expression, purification and preliminary characterization of a type II dehydroquinase from *Helicobacter pylori*. *Biochem. J.* **319**:559–565.
  10. Boyd, D., C. Schierle, and J. Beckwith. 1998. How many membrane proteins are there? *Protein Sci.* **7**:201–205.
  11. Brown, J. R., and P. V. Warren. 1998. Antibiotic discovery: is it in the genes? *Drug Disc. Today* **3**:564–566.
  12. Bult, C. J., O. White, G. J. Olsen, L. Zhou, R. D. Fleischmann, G. G. Sutton, J. A. Blake, L. M. FitzGerald, R. A. Clayton, J. D. Gocayne, A. R. Kerlavage, B. A. Dougherty, J. F. Tomb, M. D. Adams, C. I. Reich, R. Overbeek, E. F. Kirkness, K. G. Weinstock, J. M. Merrick, A. Glodek, J. L. Scott, N. S. M. Geoghagen, and J. C. Venter. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**:1058–1073.
  13. Burns, B. P., S. L. Hazell, G. L. Mendz, T. Kolesnikow, D. Tillet, and B. A. Neilan. 2000. The *Helicobacter pylori* *pyrB* gene encoding aspartate carbamoyltransferase is essential for bacterial survival. *Arch. Biochem. Biophys.* **380**:78–84.
  14. Censini, S., C. Lange, Z. Xiang, J. E. Crabtree, P. Ghiara, M. Borodovsky, R. Rappuoli, and A. Covacci. 1996. *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc. Natl. Acad. Sci. USA* **93**:14648–14653.
  15. Chalker, A. F., A. Lupas, K. Ingraham, C. Y. So, R. D. Lunsford, T. Li, A. Bryant, D. J. Holmes, A. Marra, S. C. Pearson, J. Ray, M. K. R. Burnham, L. M. Palmer, S. Biswas, and M. Zalacain. 2000. Genetic characterization of Gram-positive homologs of the XerCD site-specific recombinases. *J. Mol. Microbiol. Biotechnol.* **2**:225–233.
  16. Claros, M. G., and G. von Heijne. 1994. TopPred II: an improved software for membrane protein structure predictions. *Comput. Appl. Biosci.* **10**:685–686.
  17. Copeland, R. A., J. Marcinkeviciene, T. S. Haque, L. M. Kopcho, W. Jiang, K. Wang, L. D. Ecret, C. Sizemore, K. A. Amsler, L. Foster, S. Tadesse, A. P. Combs, A. M. Stern, G. L. Trainor, A. Slee, M. J. Rogers, and F. Hobbs. 2000. *Helicobacter pylori*-selective antibacterials based on inhibition of pyrimidine biosynthesis. *J. Biol. Chem.* **275**:33373–33378.
  18. Dayhoff, M. O., R. V. Eck, and C. M. Park. 1972. A model of evolutionary change in proteins, p. 89–99. *In* M. O. Dayhoff (ed.), *Atlas of protein sequence and structure*, vol. 5. National Biomedical Research Foundation, Washington, D.C.
  19. Doig, P., B. L. de Jonge, R. A. Alm, E. D. Brown, M. Uria-Nickelsen, B. Noonan, S. D. Mills, P. Tummino, G. Carmel, B. C. Guild, D. T. Moir, G. F. Vovis, and T. J. Trust. 1999. *Helicobacter pylori* physiology predicted from genomic comparison of two strains. *Microbiol. Mol. Biol. Rev.* **63**:675–707.
  20. Dunn, I. S. 1991. *Pseudomonas aeruginosa* plasmids as suicide vectors in *Escherichia coli*: resolution of genomic cointegrates through short regions of homology. *Gene* **108**:109–114.
  21. Engelman, D. M., T. A. Steitz, and A. Goldman. 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* **15**:321–353.
  22. Feldman, R. A., A. J. P. Eccersley, and J. M. Hardie. 1997. Transmission of *Helicobacter pylori*. *Curr. Opin. Gastroenterol.* **13**:8–12.
  23. Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.57c. Department of Genetics, University of Washington, Seattle.
  24. Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**:496–512.
  25. Forsyth, M. H., and T. L. Cover. 1999. Mutational analysis of the *vacA* promoter provides insight into gene transcription in *Helicobacter pylori*. *J. Bacteriol.* **181**:2261–2266.
  26. Fuhrman, L. K., A. Wanken, K. W. Nickerson, and T. Conway. 1998. Rapid accumulation of intracellular 2-keto-3-deoxy-6-phosphogluconate in an Entner-Doudoroff aldolase mutant results in bacteriostasis. *FEMS Microbiol. Lett.* **159**:261–266.
  27. Fujita, Y., K. Yoshida, Y. Miwa, N. Yanai, E. Nagakawa, and Y. Kasahara. 1998. Identification and expression of the *Bacillus subtilis* fructose-1,6-bisphosphatase gene (*fbp*). *J. Bacteriol.* **180**:4309–4313.
  28. Goffeau, A., et al. 1997. The yeast genome directory. *Nature*, vol. 387, suppl. 5.
  29. Haack, K. R., and J. R. Roth. 1995. Recombination between chromosomal IS200 elements supports frequent duplication formation in *Salmonella typhimurium*. *Genetics* **141**:1245–1252.
  30. Hoffman, P. S., A. Goodwin, J. Johnsen, K. Magee, and S. J. O. Veldhuyzen van Zanten. 1996. Metabolic activities of metronidazole-sensitive and -resistant strains of *Helicobacter pylori*: repression of pyruvate oxidoreductase and expression of isocitrate lyase activity correlate with resistance. *J. Bacteriol.* **178**:4822–4829.
  31. Hughes, N. J., C. L. Clayton, P. A. Chalk, and D. J. Kelly. 1998. *Helicobacter pylori* *porCDAB* and *oorDABC* genes encode distinct pyruvate:flavodoxin and 2-oxoglutarate:acceptor oxidoreductases which mediate electron transport to NADP. *J. Bacteriol.* **180**:1119–1128.
  32. Kather, B., K. Stingl, M. E. van der Rest, K. Altendorf, and D. Molenaar. 2000. Another unusual type of citric acid cycle enzyme in *Helicobacter pylori*: the malate:quinone oxidoreductase. *J. Bacteriol.* **182**:3204–3209.
  33. Kawamukai, M., H. Matsuda, W. Fujii, T. Nishida, Y. Izumoto, M. Himeno, R. Utsumi, and T. Komano. 1988. Cloning of the *fic-1* gene involved in cell filamentation induced by cyclic AMP and construction of a  $\Delta$ *fic* *Escherichia coli* strain. *J. Bacteriol.* **170**:3864–3869.
  34. Kawarabayasi, Y., M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, and H. Kikuchi. 1998. Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3. *DNA Res.* **5**:55–76.
  35. Klenk, H.-P., R. A. Clayton, J.-F. Tomb, O. White, K. E. Nelson, K. A. Ketchum, R. J. Dodson, M. Gwinn, E. K. Hickey, J. D. Peterson, D. L. Richardson, A. R. Kerlavage, D. E. Graham, N. C. Kyrpides, R. D. Fleischmann, J. Quackenbush, N. H. Lee, G. G. Sutton, S. Gill, E. F. Kirkness, B. A. Dougherty, K. McKenney, M. D. Adams, B. Loftus, S. Peterson, C. I. Reich, L. K. McNeil, H. H. Badger, A. Glodek, L. Zhou, R. Overbeek, J. D. Gocayne, J. F. Weidman, L. McDonald, T. Utterback, M. D. Cotton, T. Spriggs, P. Artiach, B. P. Kaine, S. M. Sykes, P. W. Sadow, K. P. D'Andrea, C. Bowman, C. Fujii, S. A. Garland, T. M. Mason, G. J. Olsen, C. M. Fraser, H. O. Smith, C. R. Woese, and J. C. Venter. 1997. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390**:364–370.
  36. Kunst, F., N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S.-K. Choi, J.-J. Codani, I. F. Conner-ton, et al. 1997. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* **390**:249–256.
  37. Marais, A., G. L. Mendz, S. L. Hazell, and F. Megraud. 1999. Metabolism and genetics of *Helicobacter pylori*: the genomic era. *Microbiol. Mol. Biol. Rev.* **63**:642–674.
  38. Mendz, G. L., S. L. Hazell, and B. P. Burns. 1994. The Entner-Doudoroff pathway in *Helicobacter pylori*. *Arch. Biochem. Biophys.* **312**:349–356.
  39. NIH Consensus Conference. 1994. *Helicobacter pylori* in peptic ulcer disease. NIH Consensus Development Panel on *Helicobacter pylori* in Peptic Ulcer Disease. *JAMA* **272**:65–69.
  40. Page, R. D. M. 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12**:357–358.
  41. Parkhill, J., B. W. Wren, K. Mungall, J. M. Kettle, C. Churcher, D. Basham, T. Chillingworth, R. M. Davies, T. Feltham, S. Holt, M. H. Park, A. V. Karyshev, S. Moule, M. J. Pallen, C. W. Penn, M. A. Quail, M. A. Rajandream, K. M. Rutherford, A. H. van Vliet, S. Whitehead, and B. G. Barrell. 2000. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* **403**:665–668.
  42. Petit, M. A., E. Dervyn, M. Rose, K. D. Entian, S. McGovern, S. D. Ehrlich, and C. Bruand. 1998. *PcrA* is an essential DNA helicase of *Bacillus subtilis* fulfilling functions both in repair and rolling-circle replication. *Mol. Microbiol.* **29**:261–273.
  43. Pitson, S. M., G. L. Mendz, S. Srinivasan, and S. L. Hazell. 1999. The tricarboxylic acid cycle of *Helicobacter pylori*. *Eur. J. Biochem.* **260**:258–267.
  44. Poole, L. B., A. Godzik, A. Nayeem, and J. D. Schmitt. 2000. AhpF can be dissected into two functional units: tandem repeats of two thioredoxin-like folds in the N-terminus mediate electron transfer from the thioredoxin reductase-like C-terminus to AhpC. *Biochemistry* **39**:6602–6615.
  45. Smith, D. R., L. A. Doucette-Stamm, C. Deloughery, H. Lee, J. Dubois, T. Aldredge, R. Bashirzadeh, D. Blakely, R. Cook, K. Gilbert, D. Harrison, L. Hoang, P. Keagle, W. Lum, B. Pothier, D. Qiu, R. Spadafora, R. Vicaire, Y. Wang, J. Wierzbowski, R. Gibson, N. Jiwan, A. Caruso, D. Bush, and J. N. Reeve. 1997. Complete genome sequence of *Methanobacterium thermoautotrophicum*  $\Delta$ H: functional analysis and comparative genomics. *J. Bacteriol.* **179**:7135–7155.
  46. Stewart, E. J., F. Aslund, and J. Beckwith. 1998. Disulfide bond formation in the *Escherichia coli* cytoplasm: an *in vivo* role reversal for the thioredoxins. *EMBO J.* **17**:5543–5550.
  47. Swofford, D. L. 1999. PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods), version 4. Sinauer Associates, Sunderland, Mass.
  48. Tajima, T., N. Yokota, S. Matsuyama, and H. Tokuda. 1998. Genetic analyses of the *in vivo* function of LolA, a periplasmic chaperone involved in the outer membrane localization of *Escherichia coli* lipoproteins. *FEBS Lett.* **439**:51–54.
  49. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTALW:

- improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
50. Tomb, J.-F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzgerald, N. Lee, M. D. Adams, E. K. Hickey, D. E. Berg, J. D. Gocayne, T. R. Utterback, J. D. Peterson, J. M. Kelley, M. D. Cotton, J. M. Weidman, C. Fujii, C. Bowman, L. Watthey, E. Wallin, W. S. Hayes, M. Borodovsky, P. D. Karp, H. O. Smith, C. M. Fraser, and J. C. Venter. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**: 539–547.
  51. von Heijne, G. 1992. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* **225**:487–494.
  52. Wang, Y., P. Roos, and D. E. Taylor. 1993. Transformation of *Helicobacter pylori* by chromosomal metronidazole resistance and by a plasmid with a selectable chloramphenicol resistance marker. *J. Gen. Microbiol.* **139**:2485–2493.
  53. Wang, Y., and D. E. Taylor. 1990. Chloramphenicol resistance in *Campylobacter coli*: nucleotide sequence, expression, and cloning vector construction. *Gene* **94**:23–28.
  54. Watrin, L., S. Lucas, C. Purcarea, C. Legrain, and D. Prieur. 1999. Isolation and characterization of pyrimidine auxotrophs, and molecular cloning of the *pyrE* gene from the hyperthermophilic archeon *Pyrococcus abyssi*. *Mol. Genet.* **262**:378–381.
  55. Yamamoto, S., T. Sugahara, K. Tougo, and S. Shinoda. 1994. Cloning and nucleotide sequence of the carboxynorspermidine decarboxylase gene from *Vibrio alginolyticus*. *Microbiology* **140**:3117–3124.