INFECTION AND IMMUNITY, July 2000, p. 4155–4168 Vol. 68, No. 7
0019-9567/00/$04.00+0

Comparative Genomics of *Helicobacter pylori*: Analysis of the Outer Membrane Protein Families

RICHARD A. ALM,[1]* JAMES BINA,[2]† BETH M. ANDREWS,[1] PETER DOIG,[1] ROBERT E. W. HANCOCK,[2] AND TREVOR J. TRUST[1]

Infection Discovery AstraZeneca R & D Boston, Waltham, Massachusetts 02451,[1] and Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z3[2]

Received 11 January 2000/Returned for modification 17 March 2000/Accepted 4 April 2000

The two complete genomic sequences of *Helicobacter pylori* J99 and 26695 were used to compare the paralogous families (related genes within one genome, likely to have related function) of genes predicted to encode outer membrane proteins which were present in each strain. We identified five paralogous gene families ranging in size from 3 to 33 members; two of these families contained members specific for either *H. pylori* J99 or *H. pylori* 26695. Most orthologous protein pairs (equivalent genes between two genomes, same function) shared considerable identity between the two strains. The unusual set of outer membrane proteins and the specialized outer membrane may be a reflection of the adaptation of *H. pylori* to the unique gastric environment where it is found. One subfamily of proteins, which contains both channel-forming and adhesin molecules, is extremely highly related at the sequence level and has likely arisen due to ancestral gene duplication. In addition, the largest paralogous family contained two essentially identical pairs of genes in both strains. The presence and genomic organization of these two pairs of duplicated genes were analyzed in a panel of independent *H. pylori* isolates. While one pair was present in every strain examined, one allele of the other pair appeared partially deleted in several isolates.

Helicobacter pylori is a gram-negative bacterial pathogen, and almost 50% of the world's population, approaching 100% in some countries, is infected (43). Infection with *H. pylori* has been associated with chronic gastritis and other severe gastroduodenal diseases such as peptic and gastric ulcers, gastric cancer, and mucosa-associated lymphoid tissue (MALT) lymphoma (14, 29, 35). Several molecular techniques suggest that independent *H. pylori* isolates exhibit extensive genetic diversity (2, 3, 5, 8, 24, 25, 31, 42, 58–60) which has been predicted to be important in pathogenesis, possibly relating to the wide variation in patient symptomology. Comparison of two completely sequenced *H. pylori* isolates, 26695 and J99, showed considerable allelic diversity at the nucleotide level between the gene coding sequences (4). Further, the comparison demonstrated that the chromosomes of these two strains were organized differently in a limited number of discrete regions but the overall gene order was more similar than would have been expected (4).

The gram-negative bacterial outer membrane is an asymmetric bilayer with phospholipids in the inner monolayer and the bulky glycolipid lipopolysaccharide (LPS) in the outer monolayer. Outer membranes constitute a semipermeable, size-dependent permeability barrier representing an effective barrier to hydrolytic enzymes, detergents, dyes, and hydrophobic antimicrobials. Channel-forming proteins, termed porins, also determine the permeability properties of the outer membrane. Porins contain transmembrane diffusion channels that allow small hydrophilic molecules, nutrients, and even small antibiotics to passively diffuse across the outer membrane. Most bacterial species possess only a modest number of different porins that constitute the most abundant species in the outer membrane. Many porins are nonselective and limit substrate diffusion mainly by size, whereas others have been shown to possess a high degree of selectivity for specific substrates (12, 40, 57).

The primary amino acid sequences of porins from different bacterial species generally exhibit little sequence similarity, although all are characterized by a series of amphipathic amino acid sequence motifs (alternating hydrophilic and hydrophobic residues) that form the antiparallel β-sheet structures of the membrane-spanning core region (the β barrel). These β strands are connected on the periplasmic side by short amino acid loops and on the external side of the porin by longer loops. The external loops can then fold back into the core of the β barrel to affect the pore characteristics (size, selectivity) or can function in protein-protein interactions. Many of the porin structures elucidated to date have 16 β strands, although LamB and the iron-regulated gated porins FepA and FhuA have 18 and 22, respectively (12, 23, 51). Some outer membrane proteins, including OmpA, OprH, and several proteins involved in invasion or unknown functions, possess an eight-β-stranded barrel (6). Several porins are also immunologically active and can act as protective antigens, and together with the LPS they often represent the most significant antigenic determinants of a particular bacterial species. In order to evade the host's immune system, many gram-negative bacteria exhibit considerable strain variation, due to either antigenic or phase variation or to antigenic variability, among surface epitopes of their outer membrane proteins.

The outer membrane profile of *H. pylori* on sodium dodecyl sulfate-polyacrylamide gels differs from that of other gram-negative bacteria, as the highly abundant nonselective porins (*Escherichia coli* OmpF and OmpC-like) are absent and a number of less abundant species of proteins are observed (18). A family of five outer membrane proteins from *H. pylori*, termed HopA to HopE, possess N-terminal sequence homology and have been shown to function as porins (17, 22), with

* Corresponding author. Mailing address: Infection Discovery AstraZeneca R & D Boston, 35 Gatehouse Dr., Waltham, MA 02451. Phone: (781) 839-4000. Fax: (781) 839-4570. E-mail: richard.alm@astrazeneca.com.

† Present address: Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA 02115.

two also acting as adhesins for gastric epithelial cells (44). Further, other outer membrane proteins have been identified as gastric epithelial cell or Lewis B binding adhesins (30, 48). The sequence similarity between these characterized outer membrane proteins has been used to define a much larger paralogous family with extensive C-terminal sequence homology (4, 61). We have used the complete genomic sequences of *H. pylori* J99 and 26695 to compare this large family of genes, as well as others that appear to encode outer membrane proteins.

MATERIALS AND METHODS

Computer methods. The nucleotide and amino acid sequence alignments used to produce the identity between orthologs (equivalent genes in *H. pylori* J99 and 26695) shown in Table 1 were generated by ALIGN from version 2.0 of the FASTA program package (47). The phylogeny tree was generated using Felsenstein's PHYLIP (Phylogeny Inference Package), version 3.5c, using the neighbor-joining algorithm. The BLOCKS alignment was created with MACAW (Multiple Alignment Construction and Analysis Workbench) from the National Center for Biotechnology Information (53). Paralogs were identified using BLASTP and TBLASTX algorithms. The output was initially grouped such that all members of a family exhibited homology to at least one other member using a cutoff of $P < 10^{-10}$, and the alignments were then manually inspected.

Bacterial strains. The 19 additional *H. pylori* strains were selected based on diversity of geographical origin and year of isolation (Table 2). All of the *H. pylori* strains were human isolates except ARHp12, which was a natural rhesus monkey isolate provided by S. Drazek. The AH244 and SS1 strains have been passaged in mice. All *H. pylori* strains were grown on blood agar plates for 48 h, and chromosomal DNA was prepared using a modification of the Genomic DNA Wizard Prep kit (Promega, Madison, Wis.).

Primer design and PCR analysis. Primer sequences were selected based on their predicted ability to anneal to both *H. pylori* J99 and 26695 template DNA and are listed in Table 3. All primer combinations yielded PCR products from J99 and 26695 consistent with those expected based on the published sequences (4, 61). PCR assays were performed with 50 ng of template chromosomal DNA and 0.2 μM primer, using *Taq* polymerase (Gibco BRL, Bethesda, Md.) in a Perkin-Elmer 9600 thermocycler under conditions recommended by the manufacturer. Cycling parameters for 35 cycles were as follows: denaturation at 94°C for 20 s, annealing at 55°C for 20 s, and elongation at 72°C with times varied to ensure detection of longer products if present. Products were analyzed on a 1% Tris-acetate-EDTA agarose gel under standard conditions.

RESULTS

Hop group of outer membrane proteins. The HopA-E porin proteins were originally characterized by a highly conserved N-terminal motif (A ↓ EX[D,N]G, where the ↓ represents the cleavage point) (17, 22). Analysis of the *H. pylori* 26695 sequence identified 21 proteins with this characteristic N terminus, 20 of which had orthologous members encoded by the genome of *H. pylori* J99 (Table 1; *H. pylori* J99 and 26695 gene names are preceded by "JHP" and "HP," respectively, and are numbered consecutively around the genome). *H. pylori* 26695 possesses a single strain-specific Hop protein (HP0317) which is located in a strain-specific gene cluster found in a region of organizational difference between the two strains (4). Most of the Hop proteins are predicted to contain antiparallel amphipathic β sheets that can be modeled into β barrels.

Sequence similarity analysis indicated that the Hop group of proteins represents a subfamily of a larger paralogous family of outer membrane proteins encoded by *H. pylori*. There are 12 additional genes in both *H. pylori* J99 and 26695 that encode proteins that display overall sequence similarity to those which contain the Hop N-terminal motif but do not contain the Hop motif. We propose to call these *hor* (*hop* related) genes (Table 1). The total number of members of this paralogous family, consisting of both Hop and Hor proteins, was 33 (see below).

Many gram-negative bacterial outer membrane proteins end with a C-terminal phenylalanine residue, predicted to be important for proper insertion into the lipid bilayer (56). All Hop and Hor proteins have the characteristic C terminus with alternating hydrophobic and hydrophilic residues, with aromatic residues occupying the majority of the alternating positions from −1 to −11 (counting back from the C terminus). This is consistent with the known structure of crystallized porins in which this region represents the last transmembrane β strand that associates with the first β strand to form the β barrel. Phylogenetic analysis indicates that the Hop group of proteins from J99 and 26695 cluster into two major groups (Fig. 1) based almost exclusively on the protein sequence of the C terminus. Eleven and ten members, respectively, of the Hop proteins in *H. pylori* 26695 and J99 ended with tyrosine rather than phenylalanine (Table 1) and are termed here the Y-Hop subgroup. All but two of these proteins are 70 to 80 kDa in size, with the HopA protein from each strain (JHP214/HP0229) being the smallest, having an unprocessed molecular mass of 53 kDa.

The C-terminal domains of the 70- to 80-kDa Y-Hop proteins share remarkable identity, both within an individual *H. pylori* strain and also between strains (Fig. 2A). Of the 10 orthologous pairs of Y-Hop proteins, the C-terminal domain of the Lewis B adhesins BabA (HopS) and BabB (HopT) are the most closely related. Interestingly, the C-terminal domains of the BabA and BabB proteins within each strain are more highly related to each other than the corresponding orthologs (i.e., JHP833 is closer to JHP1164 than HP1243; Fig. 2A). In contrast, the C-terminal domains of the F-Hop family members (ending in the characteristic phenylalanine residue) display less identity, and these proteins are less clustered on the phylogenetic tree (Fig. 1). The Y-Hop proteins are also less divergent at their mature N termini, leaving the central hypervariable domain containing the majority of the member-specific sequences.

One striking feature of the Hop/Hor family of proteins is their great size variation, ranging from 186 to 1,237 amino acids. A BLOCKS alignment analysis on the Hop and Hor proteins demonstrated that the majority of the homology is not at the N terminus, which was used to identify the first five members of the Hop family, but at the C terminus, where there are seven strongly conserved blocks of sequence (Fig. 2B). Interestingly, these conserved blocks of sequence are quite amphipathic and thus are predicted to contain membrane-spanning β strands (9a).

Sequence conservation analysis of the Hop proteins. The regions of outer membrane proteins which are exposed on the surface of a bacterium display a much higher rate of sequence divergence than regions located within the membrane or exposed to the periplasm, and surface-exposed proteins overall vary more than non-surface-exposed proteins (62). This sequence diversity may be driven by the immune system but may also reflect different functional capabilities. We examined the sequence diversity of orthologous pairs of the outer membrane proteins of strains J99 and 26695. These two strains were isolated approximately a decade apart on two different continents from patients presenting different clinical symptoms and thus are unlikely to be directly related. Of the 20 orthologous pairs of Hop proteins, 7 share >95% identity, with 6 having 90 to 95% and 7 having between 80 and 90% identity (Table 1). Furthermore, the distribution of identity in the corresponding genes that encode these proteins is only slightly lower, with 3 having >95% identity and 11 and 6 sharing 90 to 95% and 80 to 90% identity, respectively. This similar distribution of nucleotide and protein similarity between the Hop orthologs was not reflected when all of the orthologs between J99 and 26695 are compared, as the higher drift in the third (wobble) position of the coding triplet results in a higher amino acid identity than nucleotide identity (4, 19).

TABLE 1. Comparison of OM proteins from *H. pylori* J99 (JHP) and 26695 (HP)

Protein group	Gene no.		Length (aa)		% Size variation (aa)	C-terminal residues	% Identity		Gene name[a]
	J99	26695	J99	26695			Protein	Gene	
Family 1 (major outer membrane protein family)									
Hop proteins	7	0009	669[b]	672[b]	0.1 (1[c])	FAY	95.1	94.3	*hopZ*
	21	0025	690	711	3 (21)	FAY	85.2	85.8	*hopD*
	212	0227	696	691	0.7 (5)	FAY	81.6	83.7	*hopM*
	214	0229	483	483	0	LAY	91.5	92.3	*hopA*
	237	0252	479	487	1.6 (8)	VGF	95.1	93.5	*hopF*
	238	0253/0254	471	471[e]	0	IGF	98.1	96.8	*hopG*
	429	0477	371	367	1.1 (4)	YSF	88.1	88.9	*hopJ*
	581	0638	307	305	0[c]	NKH	92.8	94.5	*hopH*
	645	0706	270	273	0[f]	YTF	96.7	94.3	*hopE*
	659	0722	638[b]	644[b]	1.2 (8[c])	FAY	87.8	90.9	*hopO*
	662	0725	651[b]	653[b]	0.6 (4[c])	FAY	91.7	92.3	*hopP*
	833	1243	744	733	1.5 (11)	FAY	92.2	91.0	*babA* (*hopS*)
	848	0912	520	515	1 (5)	YSF	96.5	96.2	*hopC*
	849	0913	527	529	0.4 (2)	YAF	95.3	94.7	*hopB*
	857	0923	366	369	0.8 (3)	YSF	89.2	89.5	*hopK*
	1083	1156	697	696	0.1 (1)	IGF	95.4	95.1	*hopI*
	1084	1157	1,237	1,230	0.6 (7)	MGF	93.2	93.8	*hopL*
	1164	0896	703	708	0.4 (3[c])	FAY	91.8	90.2	*babB* (*hopT*)
	1103	1177	643	641	0.3 (2)	FAY	87.6	89.3	*hopQ*
	1261	1342	696	691	0.7 (5)	FAY	81.6	83.7	*hopN*
	NA[d]	0317	NA[d]	745	NA[d]	FAY	NA[d]	NA[d]	*hopU*
Hor proteins	73	0078/0079	255	684[e]	62.7 (429)	IN(L/F[g])	94.1[h]	96.2[h]	*horA*
	117	0127	286	286	0	VSF	99.0	97.3	*horB*
	307	0324	245	254	0[f]	YHF	90.6	89.3	*horC*
	359	1066	200	200	0	WHF	99.5	95.0	*horD*
	424	0472	186	186	0	FTF	99.5	96.6	*horE*
	614	0671	270	270	0	YNF	98.1	95.7	*horF*
	732	0796	278	278	0	YDF	92.4	92.6	*horG*
	1034	1107	220	230	0[f]	YNF	91.7	90.8	*horH*
	1040	1113	277	277	0	YSF	93.1	94.5	*horI*
	1362	1469	248	248	0	RDF	94.8	94.6	*horJ*
	1394	1501	388	388	0	YTF	97.9	95.6	*horK*
	1432	1395	242	242	0	FTF	90.9	90.4	*horL*
Family 2 (Hof family of outer membrane proteins)	195	0209	438	450	0[f]	YRF	91.3	91.7	*hofA*
	342	1083	479	479	0	AKF	95.6	93.8	*hofB*
	438	0486	528	528	0	YSF	95.1	93.1	*hofC*
	439	0487	465	480	0[f]	RIY	95.4	92.1	*hofD*
	719	0782	455	455	0	FFF	89.7	90.6	*hofE*
	725	0788	499	499	0	WKL	97.2	95.2	*hofF*
	850	0914	514	514	0	LKF	98.1	95.8	*hofG*
	1094	1167	471	471	0	ASF	96.4	94.9	*hofH*
Family 3 (Hom family of outer membrane proteins)	649	0710	657	660	0.5 (3)	WVF	95.2	94.3	*homA*
	870	NA[d]	668	NA[d]	NA[d]	WVF	NA[d]	NA[d]	*homB*
	1008	0373	751	700	6.8 (51)	WVF	75.3	79.3	*homC*
	1346	1453	744	746	0.3 (2)	WIF	94.9	93.0	*homD*
Family 4 (iron-regulated outer membrane proteins)									
FecA-like proteins	626	0686	767	767	0	YEF	93.1	92.6	*fecA-1*
	743	0807	792	787	0.6 (5)[i]	YNF	93.3	93.5	*fecA-2*
	1426	1400	841	842	0[f]	YTF	99.0	97.0	*fecA-3*
FrpB-like proteins	810	0876	791	791	0	YKW	97.6	94.9	*frpB-1*
	851	0915/0916	815	812[e]	0.4 (3)	YKF	96.2	93.7	*frpB-2*
	1405	1512	879	877	0.2 (2)	YQF	97.4	95.3	*frpB-3*

Continued on following page

TABLE 1—*Continued*

Protein group	Gene no.		Length (aa)		% Size variation (aa)	C-terminal residues	% Identity		Gene name[a]
	J99	26695	J99	26695			Protein	Gene	
Family 5 (efflux pump outer membrane proteins)	552	0605	477	477	0	YVH	98.3	96.5	*hefA*
	905	0971	431	413	0.4 (2[f])	VLH	92.3	91.2	*hefD*
	1247	1327	412	412	0	GLE	93.7	93.4	*hefG*
Other outer membrane proteins	456	0506	406	403	0.7 (3)	EGF	96.6	93.9	
	600	0655	906	916	1.1 (10)	TRF	96.4	94.7	
	634	0694	336	257	23.5 (79)[j]	FAF	73.8	73.2	
	663	0726	305	305	0	FLF	93.4	92.8	
	1022	0358	511	511	0	GLF	93.3	93.2	
	1360	1467	231	231	0	YKF	95.2	93.2	
	308	0325	237	237	0	MPY	98.3	96.8	*flgH*
	777	0839	587	587	0	YRW	97.8	94.3	
	1054	1125	179	179	0	LVK	95.5	96.3	*palA*
	1349	1456	175	175	0	VKK	100	96.6	*lpp20*

[a] Those with Hop-like motifs have been named *hop* genes, with the original *hopA-E* gene names being assigned to those previous identified (17, 22). Proteins related to the *hop* family but lacking the N-terminal motif have been called *hor* (hop related) genes. The family of 50-kDa outer membrane protein genes has been called *hof* (*Helicobacter* OMP family) genes. The smaller family of outer membrane protein genes has been called *hom* (*Helicobacter* outer membrane) genes.
[b] Out of frame due to a CT dinucleotide repeat in the signal sequence. Protein size was determined by adjusting the coding sequence by the addition or removal of a single dinucleotide repeat.
[c] The size variation does not include the differences caused by the different numbers of CT dinucleotide repeats in the coding sequence.
[d] NA, not applicable.
[e] The HP0253 and HP0254 genes, the HP0078 and HP0079 genes, and the HP0915 and HP0916 genes were joined by the addition or removal of a single nucleotide.
[f] Size difference due to difference in prediction of initiation codons between *H. pylori* J99 and 26695.
[g] Protein terminates with an F residue in 26695 and an L in J99.
[h] The identity was calculated over the aligned portion of the proteins only.
[i] There is significant difference in the C-terminal 20 amino acids. The C terminus of JHP743 is found in a different reading frame in *H. pylori* 26695 and likely represents a frameshift in HP0807.
[j] The remainder of JHP634 is found in a different reading frame in *H. pylori* 26695 after a frameshift.

This level of conservation for outer membrane proteins was also found in other *H. pylori* strains. The *hopB* genes and their encoded proteins from *H. pylori* J99, 26695, and 17874 share 92% nucleotide identity and 94% amino acid identity (98% similarity) across their entire length. There is a single region between a conserved pair of Cys residues where the three HopB proteins differ substantially, including the insertion of several additional residues in the HopB protein from *H. pylori* 17874 (Fig. 3A). A similar level of identity is found with the *hopC* gene and the encoded protein from these three *H. pylori*

TABLE 2. Strains used in this study

Strain	Country	Isolation yr	Disease state	Reference
J99	United States	1994	Duodenal ulcer	4
26695	United Kingdom	1986[a]	Gastritis	20
AH244	Sweden	1993	Duodenal ulcer	This study
SS1	Australia	1995[a]	Dyspepsia	37
UA861	Canada	1991	Duodenal ulcer	54
ARHp210	Sweden	1997	Asymptomatic	This study
ARHp12	United States	1993[a]	Natural rhesus monkey isolate	This study
ARHp18	Canada	1989[a]	—[b]	This study
ARHp25	Australia	1989[a]	—	33
ARHp64	Argentina	1996[a]	Nonulcer dyspepsia	This study
ARHp65	Argentina	1996[a]	Nonulcer dyspepsia	This study
ARHp55	United States	1996[a]	Duodenal ulcer	This study
ARHp124	Bangladesh	1996[a]	Hiatus hernia and gastritis	This study
ARHp54	United States	1996[a]	Duodenal ulcer	This study
CCUG 17874[c]	Australia	1984	—	41
ARHp221	United States	1998	Cat isolate	26
ARHp246	Kuala Lumpur	1998	Duodenal ulcer, gastritis	This study
ARHp245	France	1998	Pernicious anemia	This study
ARHp241	Kuala Lumpur	1998	Duodenal ulcer, erosive gastritis	This study
ARHp243	France	1998	Duodenal ulcer	This study
ARHp244	France	1998	Nonulcer dyspepsia	This study

[a] Strain isolated prior to this date.
[b] —, Exact clinical presentation was not recorded.
[c] Reported to be identical to the *H. pylori* type strain 11637 (46), although it has been reported that two versions of 11637 exist (1). The strain used here was the same as that used by O'Toole et al. (46).

TABLE 3. Primers used in this study

Name	Sequence (5′-3′)
hopJK	GAAGAAAATGGGGCGTATGCGAGCG
jhp430	TCCAACAGAAAGAGCGTTTGAAGGC
jhp858	GCCAGAAAATGGAGGGCCAACAAACG
jhp428	CGGCTCAAATCCGTGTCTTCAATGCG
jhp856	TGCGGGCATAGGGGCTAGGTTTGGGC
jhp213	ATCACAGAAAGCCCCACCACAAAACC
jhp211	TCGCGCTAGGGACGACAATCTCCC
jhp1260	GGCTTTAGAAGCCATTAAAAGCGCGG
jhp1262	TATTTGTATGCGGGTATTGGTTTTGC
hopMN-1	GAAGATGACGGATTTTACATGAGTGTGGG
hopMN-2	GCGCTAAAGCCACAGCTTGATAGGCC
hopMN-3	TGAAAACACCCAAATCACGCAACC
hopMN-4	TTGGATAGGCCCTTGAATGCTGTGG
hopMN-5	TGAACGGCATCGGCGTGCAAGCGGGC
jhp73F	GAAAAAAGCGGCGCGTTTTTAGGAGGG
jhp73R	GAACACATCTACCGATCCATCTACGCC
jhp73R2	CCCCCAACACAACAAAATAAATATCGC

strains. Significantly, however, there is a single region where the three protein sequences differ significantly (Fig. 3B), including the insertion/deletion of up to six amino acids. Molecular modeling using a strategy described by Huang et al. (28) and hydrophobicity plots suggests that these variable domains are unlikely to be inserted into the membrane. Whether they are located in the periplasmic space as suggested by Odenbreit et al. (44) or exposed on the cell surface, as well as any functional significance, remains to be determined. Sequence alignments of the BabA (HopS) and BabB (HopT) proteins from *H. pylori* 17875 (30) with the corresponding orthologs from *H. pylori* J99 and 26695 demonstrated that these proteins are both 88% identical, with similarity levels being above 92%. Overall there is less variation between these orthologs from different *H. pylori* strains than observed between the outer membrane proteins of other species, e.g., the *Chlamydia trachomatis* major outer membrane protein porin (36).

Duplicated genes encoding Hop outer membrane proteins. *H. pylori* J99 and 26695 contain two pairs of essentially duplicated Hop genes, *hopJ*/*K* and *hopM*/*N*. In both cases the high level of sequence identity of these duplicated genes within a given *H. pylori* strain is not reflected between the strains. While the JHP212 and JHP1261 (*hopM* and *hopN*) genes are 100% identical to each other, they share only 83.7% identity to the corresponding orthologs from *H. pylori* 26695 (HP0227 and HP1342), which themselves are 100% identical. Similarly, the level of identity between the *hopJ* (JHP429/HP0477) and *hopK* (JHP857/HP0923) genes drops to 88.9 and 89.5%, respectively. This is despite the intrastrain sequences sharing extremely high identity. The JHP429 and JHP857 proteins differ by only a single amino acid residue in the mature protein, although three amino acid differences and a five-amino-acid insertion in the predicted signal sequence of JHP429 reduce the overall iden-

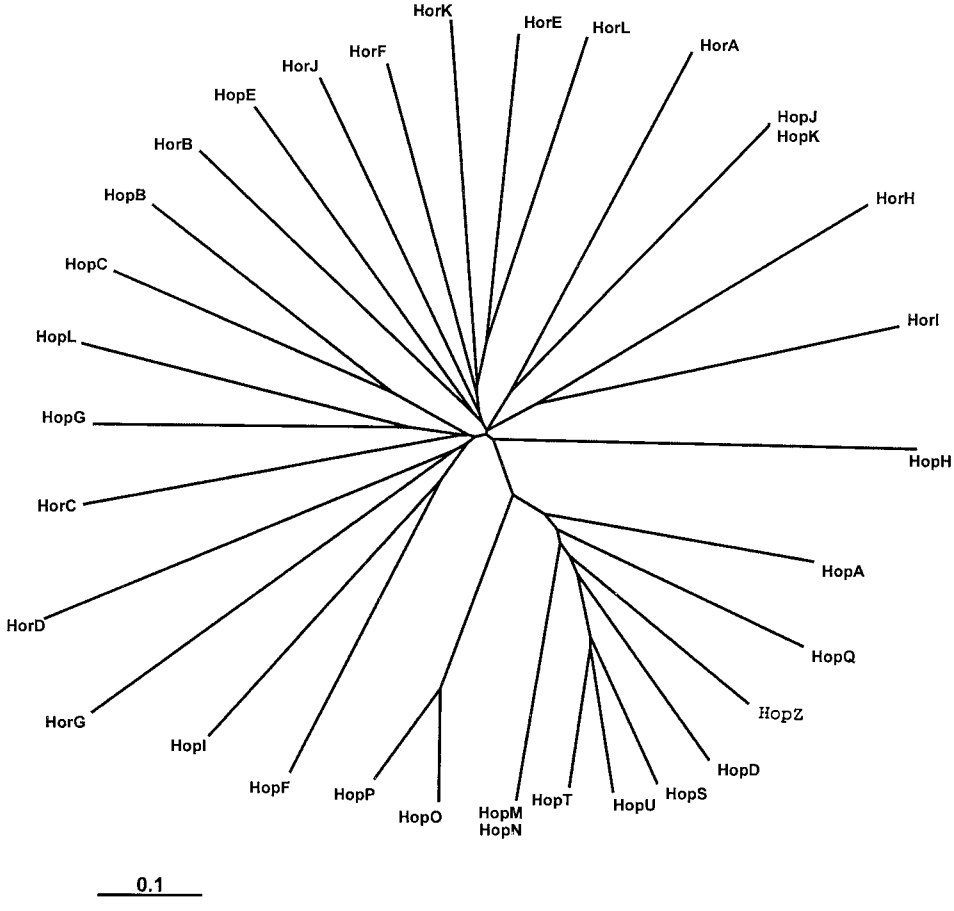
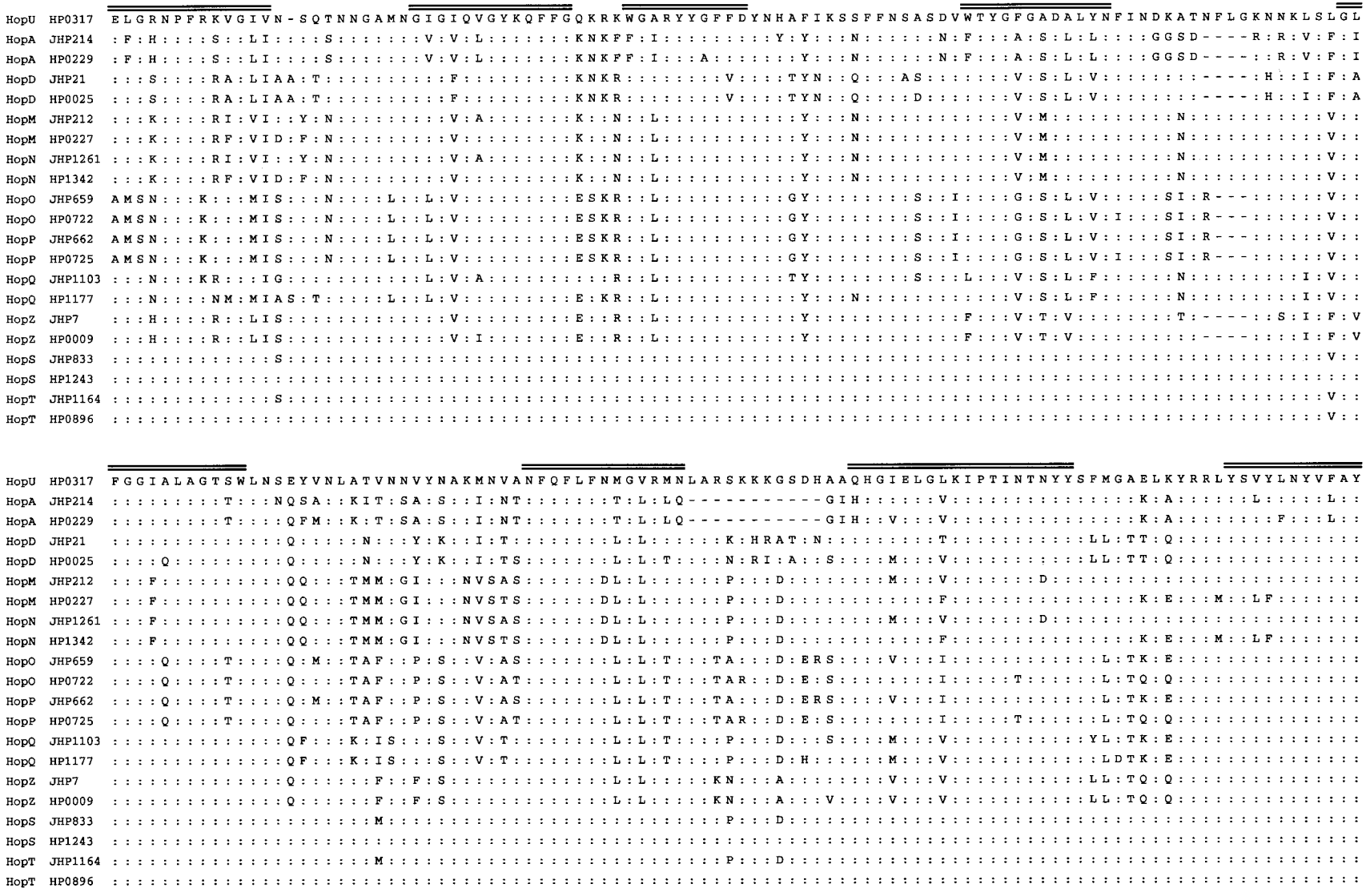


FIG. 1. Phylogenic tree of the large Hop and Hor outer membrane protein family. Protein sequences were analyzed using the PHYLIP program. The two pairs of duplicated Hop proteins (HopJ/K and HopM/N) were not differentiated and are each visualized as one line.



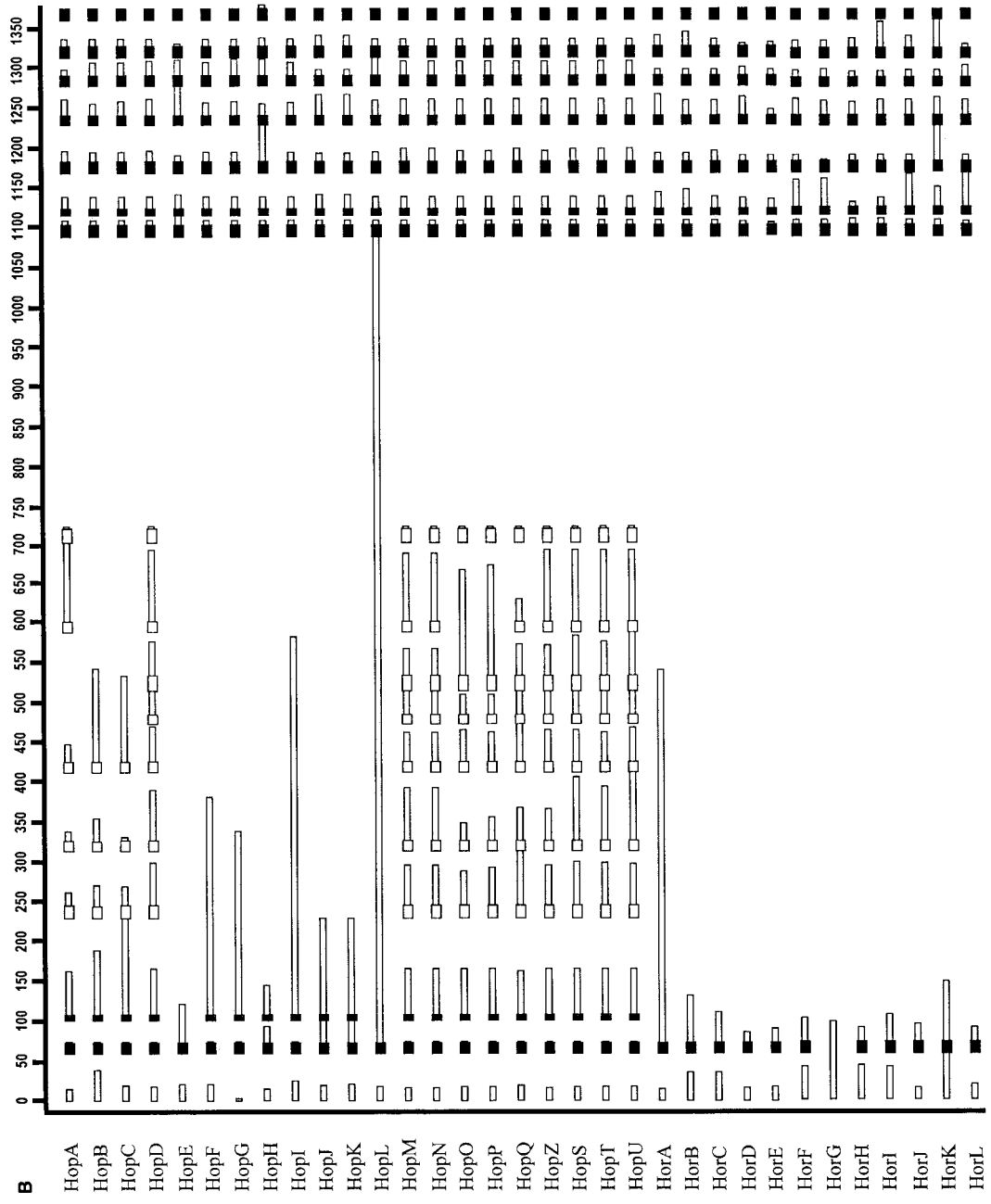


FIG. 2. (A) Alignment of the C-terminal domains of the Y-Hop proteins from *H. pylori* J99 and 26695. The alignment is based on the sequence of HP0317, the strain-specific member from *H. pylori* 26695. The proteins are listed as orthologous pairs from the two strains. Identical residues are indicated by colon; the eight predicted transmembrane sequences are indicated above the sequence. (B) BLOCKS alignment of the Hop and Hor proteins. BLOCKS is a method used to demonstrate similarity among a group of proteins that contain repeated sections of high similarity across the family (filled boxes) or a subset of the family (unfilled boxes) flanked by regions of lesser similarity (empty bars) and variable size (blank regions representing sequence missing from a given protein).

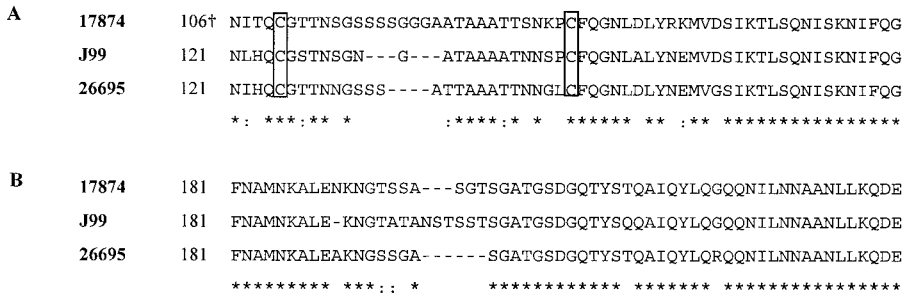


FIG. 3. Alignment of the variable domains of HopB (A) and HopC (B). The *H. pylori* 17874 proteins are found in GenBank (accession number Z82988) and are called AlpB and AlpA, respectively. Positions of the proteins included in the alignment are indicated with numbers; † indicates that the difference in position within the HopB protein represents a difference in the prediction of the initiation codon. The conserved cysteine residues in the HopB proteins are boxed. Identical (*) and conserved (:) residues are indicated.

tity to 97.5%. At the nucleotide level, the portions of the genes encoding the mature JHP429 and JHP857 proteins differ by 3 nucleotides (nt), with two resulting in silent amino acid changes. Similarly, the HP0477 and HP0923 genes in *H. pylori* 26695 are identical except for a 6-bp insertion (encoding two amino acids) in the N-terminal signal sequence of HP0923. In both sequenced *H. pylori* strains, the *hopJ*/*K* and *hopM*/*N* gene duplications are separated by approximately 0.5 Mb.

Ilver et al. (30) identified two copies of the *babA* allele in strain CCUG17875. In contrast, *H. pylori* 26695 and J99 possessed only one *babA* allele (HP1243/JHP833). The genomic location of the J99 *babA* gene is different from that seen in 26695, as its location has been reciprocally exchanged with *babB* (4). Thus, it seems that different *H. pylori* strains may duplicate different genes. Whether this is a random event or whether it confers some biological advantage, such as antigenic or receptor ligand variation, to particular strains in association with the different hosts is unknown, as is the precise mechanism for duplication.

Nineteen additional *H. pylori* strains, representing a variety of geographical sources, clinical spectrums, and isolation dates (Table 2), were examined for the presence of duplicate copies of the *hopJ*/*K* and *hopM*/*N* genes. Specific PCR primers were designed to anneal to both *H. pylori* J99 and 26695 sequences within and flanking these duplicated genes. Using the hopJK primer (Table 3) in conjunction with primers specific for the downstream gene in both genomic locations (jhp430 and jhp858 [Table 3]), all *H. pylori* strains tested were shown to possess both copies of *hopJ* and *hopK* (genes JHP429 and JHP857) (Fig. 4A and B). Further, use of the downstream primers together with primers for the upstream genes in both locations (jhp428 and jhp856 [Table 3]) demonstrated that the *hopJ* and *hopK* genes in all the *H. pylori* strains tested were flanked by the same genes as present in J99 and 26695 (data not shown).

Similar experiments were performed to examine the presence of the duplicated *hopM* and *hopN* genes (JHP212 and JHP1261). Possibly due to the nucleotide variation between *H. pylori* strains (4), not all PCRs generated a specific amplicon. However, using several different primer combinations (jhp213/hopMN-2 [Fig. 4C]; hopMN-3/jhp211, jhp213/hopMN-4, jhp211/jhp213, and jhp211/hopMN-1 [data not shown]), all of the *H. pylori* strains tested were shown to contain a *hopM* (JHP212) ortholog flanked by JHP211 and JHP213 orthologs.

The presence and location of the *hopN* (JHP1261) orthologs were initially analyzed using the primer combinations jhp1260/hopMN-2 and hopMN-3/jhp1262. Specific amplicons were detected with both primer combinations in J99, 26695, and seven additional isolates (ARHp64, ARHp18, ARHp25, ARHp65, ARHp55, AH244,551, and UA861), indicating the presence of an intact *hopN* ortholog in these strains flanked by the same genes found in J99 and 26695 (Fig. 4D and data not shown). Strains ARHp54, ARHp221, ARHp241, and 17874 yielded a specific hopMN-3/JHP1262 amplicon (Fig. 4D), but no product was detected using jhp1260/hopMN-2 (data not shown). Further PCR analysis using the jhp1260/hopMN-4 (Fig. 4D) and jhp1260/jhp1262 (data not shown) primer combinations confirmed that six additional strains (ARHp54, ARHp221, ARHp243, ARHp245, ARHp246, and 17874) contained an intact JHP1261 ortholog at this location. However, these primer combinations yielded products that were ~800 bp shorter in four strains (ARHp12, ARHp124, ARHp241, and ARHp244), suggesting that the N-terminal region of the JHP1261 ortholog had been deleted (Fig. 4D). No products were detected from ARHp210 with any of the primer combinations used, suggesting either an organizational difference at this location, significant sequence divergence causing failure of the primers to anneal, or the absence of the *hopM* and *-N* genes in this strain. Representative PCR products that were generated at both loci from several strains (J99, 26695, AH244, 17874, ARHp25, ARHp65, ARHp241, ARHp243, and ARHp246) were partly sequenced to ensure that the primers were anchoring correctly and that the product represented the *hopM*/*N* gene. In all cases when the sequence generated was translated, the highest similarity in either *H. pylori* genome to the predicted protein was to the HopM and -N proteins.

Analysis of the Hor proteins. The *hor* gene family, which is made up of 11 members previously grouped into the Hop family (61) and JHP359/HP1066 (HorD), are even more highly conserved than the Hop proteins, with five proteins having >95% identity and the remaining 7 being >90% identical between the two sequenced strains. Only one of the orthologous *hor* gene pairs displayed less than 90% identity at the nucleotide level (Table 1). Eleven of the twelve orthologous Hor protein pairs (except JHP73 [HorA] [see below]) are the same size in both *H. pylori* J99 and 26695, which is in contrast to the 20 orthologous Hop protein pairs, where 16 of the pairs differ in size by up to 3% (Table 1).

The JHP73 protein is 255 amino acids in length; although it does not terminate in a hydrophobic residue, it shares significant similarity with the other members of the family and appears to represent a gene fusion between two adjacent genes in *H. pylori* 26695. The N terminus of JHP73 aligns with the N terminus of HP0078, while the C terminus of JHP73 aligns with the C terminus of HP0079 (Fig. 5A). Since HP0078 and HP0079 are 11 nt apart, it is possible that they are the rem-

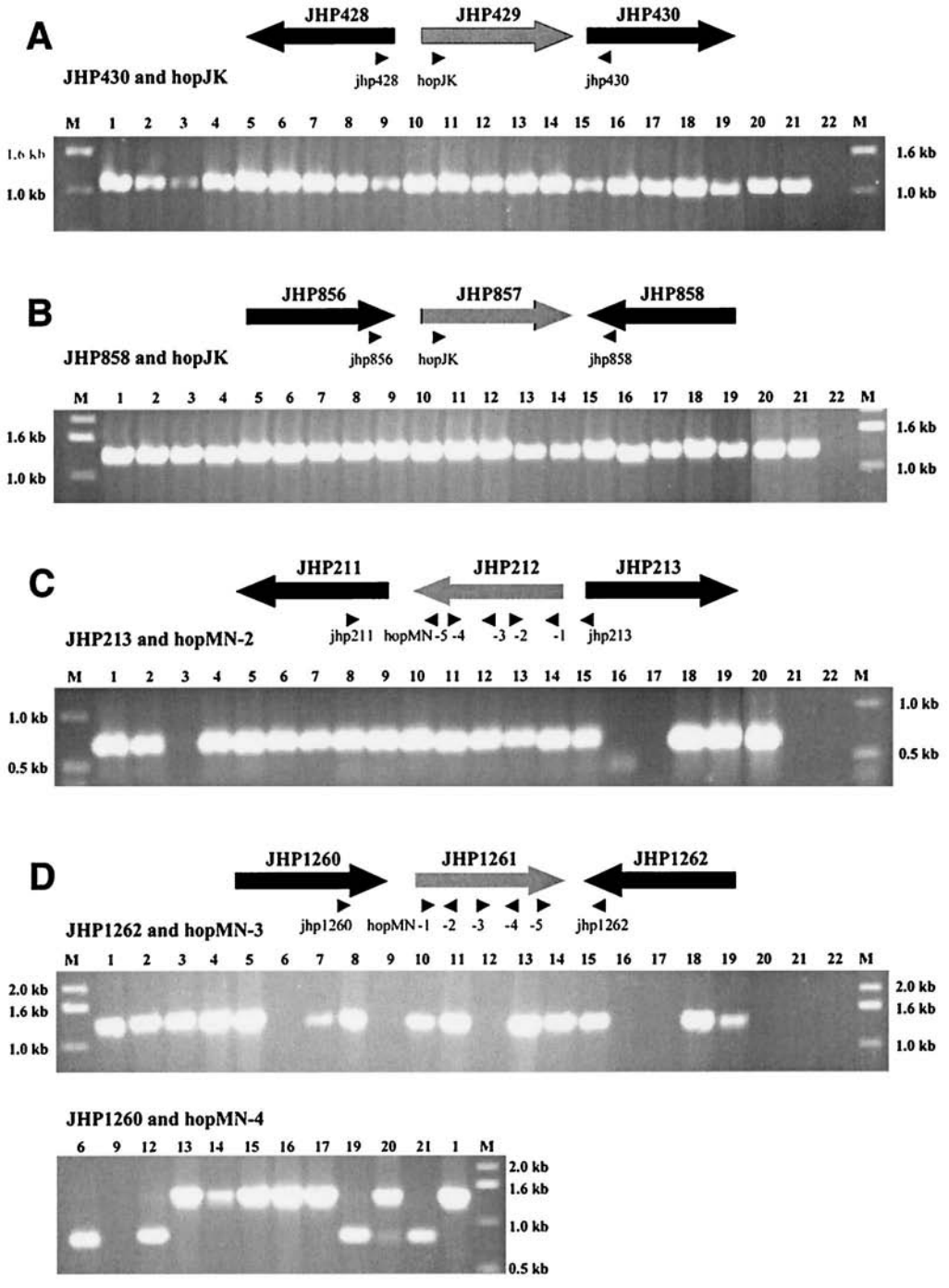


FIG. 4. Examination of *H. pylori* isolates for the duplication of *hop* genes. The genomic organization and primer binding location sites for JHP429 (*hopJ*) (A), JHP857 (*hopK*) (B), JHP212 (*hopM*) (C), and JHP1261 (*hopN*) (D) are shown. Representative PCRs are also shown in each panel, with the primer combinations used indicated. The loading order for each panel is as follows: marker (lane M), J99 (lane 1), 26695 (lane 2), ARHp64 (lane 3), SS1 (lane 4), UA861 (lane 5), ARHp12 (lane 6), ARHp18 (lane 7), ARHp25 (lane 8), ARHp210 (lane 9), ARHp65 (lane 10), ARHp55 (lane 11), ARHp124 (lane 12), ARHp54 (lane 13), CCUG17874 (lane 14), ARHp221 (lane 15), ARHp246 (lane 16), ARHp245 (lane 17), AH244 (lane 18), ARHp241 (lane 19), ARHp243 (lane 20), ARHp244 (lane 21), and no-DNA control (lane 22). The strains shown in the second gel in panel D (primers JHP1260 and hopMN-4) are indicated with the same numbering system. The sizes of the molecular weight markers are indicated.

nants of a single gene and have undergone some genetic decay and thus represent an untranslated pseudogene. Inspection of the breakpoints of the alignment revealed a direct repeat of 9 out of 10 nt at each end, and simple intragenomic recombination within *H. pylori* 26695 could result in an in-frame deletion resulting in the shorter JHP73 protein. Oligonucleotide primers corresponding to the N-terminal and C-terminal coding regions of JHP73 (jhp73F and jhp73R [Table 3]) were designed to examine this area in other *H. pylori* isolates, and PCR analysis confirmed the J99 and 26695 structures (Fig. 5B, lanes

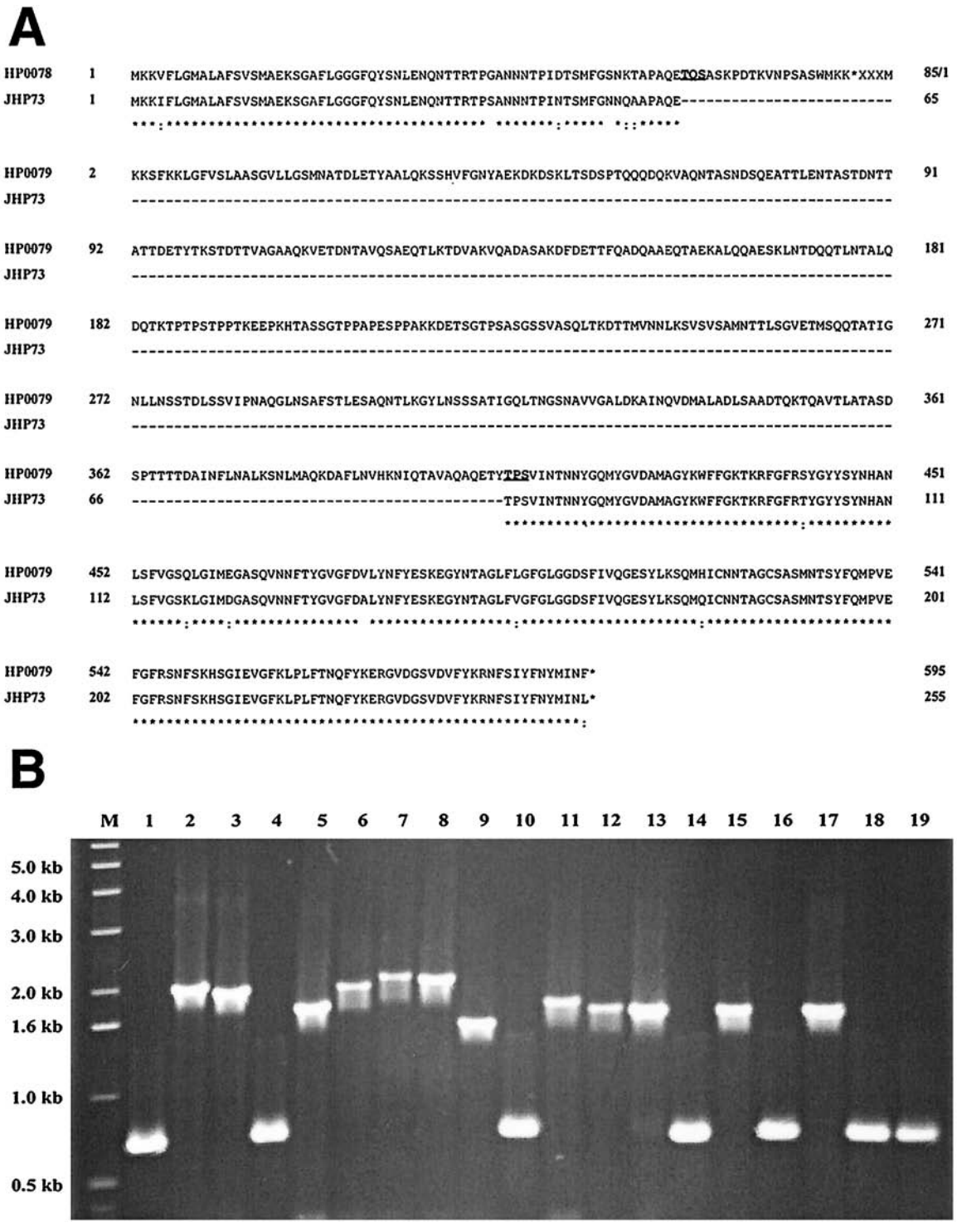


FIG. 5. (A) Alignment of the JHP73 and the HP0078/HP0079 proteins. Amino acid positions of the proteins are indicated by numbers. Identical (*) and conserved (:) residues are indicated. As predicted by Tomb et al. (61), the short HP0078 protein ends after 85 residues and the HP0079 protein begins 11 nt later. (B) PCR analysis of multiple *H. pylori* isolates for the presence of a JHP73 ortholog, using the primer combination jhp73F/jhp73R. Molecular weight markers are shown in lane M, with the sizes indicated on the left. The strains analyzed are J99 (lane 1), 26695 (lane 2), ARHp64 (lane 3), SS1 (lane 4), UA861 (lane 5), ARHp12 (lane 6), ARHp18 (lane 7), ARHp25 (lane 8), ARHp210 (lane 9), ARHp65 (lane 10), ARHp55 (lane 11), ARHp124 (lane 12), ARHp54 (lane 13), CCUG17874 (lane 14), ARHp221 (lane 15), ARHp245 (lane 16), AH244 (lane 17), ARHp243 (lane 18), and ARHp244 (lane 19).

1 and 2). However, there was considerable size heterogeneity in the products generated from the other *H. pylori* isolates, suggesting that this region displays significant variability (Fig. 5B). PCR analysis was also performed using the jhp73F primer in combination with a primer downstream of the termination codon (jhp73R2) to corroborate the specificity of these products. All strains except ARHp18 and ARHp124 yielded a product which was larger by the expected 130 nt (data not shown).

The absence of a product in the ARHp18 and ARHp124 strains is likely due to sequence differences outside the coding region of the JHP73 ortholog. Two strains, ARHp241 and ARHp246, failed to generate a PCR product with either primer pair (data not shown).

Additional paralogous families of outer membrane proteins. The 50-kDa non-heat-modifiable protein located in the outer membrane of strain CCUG17874 (22) was also found in the 26695 and J99 genome sequences. This gene (JHP438/HP0486) encodes a protein of 528 amino acid residues and has a 29-amino-acid residue signal sequence preceding the published N-terminal sequence of the mature protein (22). It possesses the hydrophobic C-terminal sequence motif characteristic of many outer membrane proteins (56). In *H. pylori* J99 and 26695, this protein is a member of a paralogous family which contains eight members (we propose to call this family *hof* genes, for *H. pylori* outer membrane protein family [Table 1]). The molecular masses of the proteins in this family are similar, ranging from 51.2 to 59.7 kDa (predicted sizes of proteins including putative signal sequences), and the predicted mature forms of all the orthologous pairs of proteins are identical in size between *H. pylori* J99 and 26695 (Table 1). The level of amino acid similarity between the orthologs is high, with six of the eight being >95% identical, while the nucleotide identity is characteristically lower, with only two members having >95% identity (Table 1).

There is a smaller paralogous family of proteins that also contain the C-terminal alternating hydrophobic motif and characteristic signal sequences typical of outer membrane proteins which we propose to call the *hom* family (for *H. pylori* outer membrane proteins [Table 1]). *H. pylori* 26695 contains three members of this family, whereas *H. pylori* J99 contains an additional strain-specific member. All of these members have conserved N and C termini, while the central domain of the molecule displays significant variability. The J99 strain-specific member of this paralogous family (JHP870) is 90% identical to JHP649, with all of the differences being confined to the central domain (residues 147 to 344), suggesting that the presence of the JHP870 gene may have resulted from a relatively recent gene duplication. The JHP870 gene is a single insertion in the *H. pylori* J99 genome, with the genes flanking the insertion point in the two genomes being orthologs (JHP869/HP0935 and JHP871/HP0936). Significantly, the intergenic space between HP0935 and HP0936 in the *H. pylori* 26695 genome contains a stretch of 219 nt which displays 96.8% identity (seven mismatches) to the JHP870 gene (the region which encodes residues 496 to 569). The presence of this DNA in *H. pylori* 26695 at this genomic location strongly suggests that a JHP870 ortholog once existed in this strain.

There are two families with homology to the iron-regulated outer membrane proteins from other bacteria. These have been labeled FecA-like and FrpB-like, due to their similarity with the ferric citrate receptor of *E. coli* and to a major iron-regulated outer membrane protein in *Neisseria* spp., respectively. Both of these families contain three paralogous members, although the JHP851 ortholog in *H. pylori* 26695 is split into two ORFs (HP0915/HP0916 [Table 1]). Worst et al. (63) identified iron-repressible outer membrane proteins with molecular sizes of 77, 50, and 48 kDa. The 48- and 50-kDa proteins could represent HopA to -D, which have been shown to be iron repressible (21). The *hopA*, *-B*, *-C*, and *-D* genes all have potential Fur boxes in their upstream regions with 13, 15, and 10 out of 19, respectively, identical residues to the consensus Fur box that binds the *E. coli* iron regulator Fur. This is consistent with the finding that the *H. pylori* Fur protein can partially complement the *fur* mutation in *E. coli* (7).

Both sequenced *H. pylori* strains contain three clusters (*hefA-C*, *hefD-F*, and *hefG-I*) which encode homologs of resistance-nodulation-division efflux pump systems (9). Each system contains an outer membrane component with some homology to the *E. coli* TolC protein, and these proteins (HefA, -D, and -G) share >92% amino acid identity between the J99 and 26695 strains (Table 1). These proposed efflux systems have been shown to be highly conserved in sequence and organization between multiple *H. pylori* strains (9).

The vacuolating cytotoxin (VacA) of *H. pylori* is translated as a preprotein, which is subsequently processed at both the N and C termini to yield an 87-kDa mature toxin (15). Although virtually all strains carry the *vacA* gene, it appears to be expressed in only ~50% of *H. pylori* strains (16). Allelic diversity has been observed in the signal sequence and the central domain between different isolates, but the C terminus which is cleaved as the protoxin traverses the membrane is highly conserved between strains. Both sequenced *H. pylori* genomes contain three large proteins that display similarity to VacA. Although these paralogs have been labeled as outer membrane proteins (61), all three lack the dicysteine cleavage signal as well as recognizable N-terminal signal sequences. Therefore, the cellular localization of these proteins cannot be accurately predicted. Two of the three orthologous pairs differ significantly in size, with JHP856 encoding a protein that is 130 amino acids shorter than that encoded by HP0922, and JHP556 represents a fusion between HP0609 and HP0610 (4).

Another putative outer membrane protein that has been described is a 30-kDa lipoprotein named HpaA (JHP733/HP0797). There have been conflicting reports in the literature regarding the exact localization of this protein (11, 39, 46). This protein possesses similarity with two other similarly sized proteins present in both *H. pylori* genomes. One of these putative paralogs (JHP444/HP0492) contains a consensus type II signal sequence and may also be a lipoprotein, whereas the other (JHP971/HP0410) is predicted to contain a type I signal sequence.

Outer membrane proteins not in paralogous families. Six additional open reading frames whose products are predicted to be located in the outer membrane were identified based on the C-terminal motif characteristic of outer membrane proteins. These outer membrane proteins and the level of identity between the orthologs are listed in Table 1. In addition, four other probable outer membrane proteins that are not part of paralogous families have been included in Table 1. The JHP777/HP0839 protein is a homolog of the *Haemophilus influenzae* P1 protein that is also related to the fatty acid transport protein FadL of *E. coli* (10). The FlgH protein that forms the flagellar L ring serves as a frictionless bearing for the flagellum and is located in the outer membrane. The FlgH homolog of *H. pylori* (JHP308/HP0325) contains a 21-amino-acid signal sequence but is rather small relative to its homologs in other bacterial species. There are also several lipoproteins that may be associated with the outer membrane. Among these was a homolog of the peptidoglycan-associated lipoprotein (PAL) family that includes the *Campylobacter jejuni* PAL protein (13) and the lpp20 lipoprotein that has been localized to the outer membrane, albeit not exclusively (34) (Table 1).

Signal sequences and ribosome binding sites (RBSs). Outer membrane proteins contain signal sequences that are recognized and processed by the Sec pathway secretion machinery during protein translocation to yield the mature product. We examined the putative signal sequences of the larger families of outer membrane proteins (Hop, Hor, and Hof), some members of which have been previously sequenced to yield a cleavage site. This permitted an analysis of the consensus signal

sequences and cleavage regions for *H. pylori* outer membrane proteins. Although substantial variations were permitted, the consensus sequence around the cleavage site (↓) was SLLXA↓n, where X is from the group of amino acids which include L, S, R, Q, H, A, I, Y, P, N, and G and n is the amino-terminal amino acid of the mature protein (most often E). A leucine residue was found in position −3 in 21 of 35 signal sequences analyzed, in contrast to most gram-negative bacterial signal sequences (Signal P Server), where alanine is found 10 times more often than leucine. This difference may reflect small changes in the specificity of signal peptidase I in *H. pylori*. Although some *H. pylori* signal sequences had unusual features (e.g., four Hop proteins had Arg in position −2, and up to four Ser residues were observed in the hydrophobic core region), they all fell within the range of known signal sequences.

Analysis of the RBSs of the *hop* genes from the two sequenced *H. pylori* strains revealed the consensus of AAGGA-(5 to 9 nt)-ATG. While this is generally consistent with sequences observed in other bacteria, 23 of the 41 genes analyzed had the shorter spacing of 5 to 6 nt. This may either contribute to poorer expression for these genes or reflect a minor difference in the translation machinery of *H. pylori*. Of the 20 orthologous pairs of *hop* genes, only 7 have nucleotide differences between the RBS and initiation codon. Of these, three have insertion/deletion of bases that alters the spacing between the RBS and the initiation codon, and such changes in spacing may affect the level of expression of these proteins. Significantly, differences in spacing are observed in both the *babA* and *babB* genes between *H. pylori* J99 and 26695. Several of the *hop* genes from both strains contain a string of A residues between the RBS and the initiation codon, and slipped-strand repair at these locations may play a role in the modulation of expression of these proteins. Consistent with this notion is the difference in the number of A nucleotides from GAAAAC to GAAAAAAC in the *babA* genes from *H. pylori* J99 and 26695, respectively. The initiation codon of all *hop* genes is AUG, with the exception of *hopF* and *hopI*, which have a predicted UUG initiation codon in both strains. The RBSs of the *hor* and *hof* genes all fit within the consensus seen for the *hop* genes except for *horD* and *horF*, which have spacings of 12 and 10 nt to their respective initiation codons. The spacing for the orthologous *hor* and *hof* genes in *H. pylori* J99 and 26695 is identical except for *hofE*, which has a spacing of 7 nt in strain 26695, compared to 8 nt in J99.

DISCUSSION

Gram-negative bacterial outer membranes mediate the interaction with the surrounding environment. For *H. pylori* to survive and persist in the gastric mucosa, adaptation of the outer membrane could be expected. Comparative analysis of two complete *H. pylori* genome sequences has confirmed the presence of large families of integral outer membrane proteins that represent approximately 4% of each strain's coding potential. Members of the Hop outer membrane protein family have been implicated as adhesins (30, 44, 48), including two which also act as porins (22). The use of outer membrane proteins as adhesins may represent an adaptation to the gastric environment, where the acidic conditions would likely depolymerize any polymeric pilus structure. A similar adaptation may be the encasement of the flagellum of *H. pylori* by a sheath with a composition similar to that of the outer membrane, an organization that may also protect the polymeric flagellar structure.

The presence of large paralogous outer membrane protein families may have resulted from gene duplication that produced a repertoire of proteins which may not only be antigenically diverse but also have different functions. Both *H. pylori* J99 and 26695 possessed two *hop* genes in duplicate copies (*hopJ*/*K* and *hopM*/*N*). All other *H. pylori* isolates tested also contained copies of both the *hopJ* and *hopK* genes flanked by the same neighboring genes as in J99 and 26695. Analysis of the *hopM* and *hopN* genes demonstrated that many of the strains tested contained both genes, but several appeared to show deletions at one of the loci, consistent with recent findings (32). The presence of a large number of related proteins suggests the existence of a mechanism for generating chromosomal diversity needed for host defense evasion or determination of host specificity (30, 61). The presence of the same genes being duplicated almost identically in multiple strains but with different sequences between strains is intriguing. It is possible that each strain maintains almost perfect duplication by repeatedly taking up DNA from lysed surrounding cells and integrating this DNA into the two duplicated sites. In this manner, the gene sequences between *H. pylori* strains would be able to diverge significantly while still maintaining a very high intrastrain identity. This model would also explain the higher C-terminal identity of BabA and BabB within a given strain that either protein with its corresponding ortholog from another *H. pylori* strain.

H. pylori strains are likely to also contain strain-specific outer membrane protein genes (similar to *hopU* or *homB*) which may confer an advantage during the evolution of the host-parasite interaction. However, the presence of essentially the same members in each family, together with the conservation of gene duplications in strains with different origins, suggests that the proteins are preserved for a functional reason. We predict that the highly conserved domains of sequence represent conserved scaffolding for a β-barrel pore. Several lines of evidence support this hypothesis: (i) five members of the Hop family form pores in planar bilayer membranes, and all of the porins reported to date have β-barrel structures, (ii) linker insertion mutagenesis studies of HopE were consistent with the conserved regions being largely transmembrane β strands (9a), and (iii) examination of the sequence for amphipathic signatures typical of the transmembrane β-strands of porins (28) revealed that all members of the Hop/Hor family contained such sequences and they were all clustered at the C terminus. The smaller members of the Hop/Hor families (<35 kDa; e.g., HopE) have predicted amphipathic β-strands throughout their sequences, whereas the larger proteins in the family (e.g., HopA to -D, BabA, and BabB) contain large N-terminal segments without sequences predicted to form amphipathic β strands. The conserved N and C termini shared by the Hop proteins may be involved in correct transport and integration into the outer membrane or in protein-protein interactions with either other family members or other self-copies during multimer formation.

There are examples for β-barrel porins associated with various N-terminal regions. The iron-regulated outer membrane proteins FepA (12) and FhuA (23) were recently shown to possess an N terminus containing four additional β strands inserted from the periplasmic side into the center of the barrel, forming a gate to the iron-siderophore binding site. Such gated porins do not demonstrate nonspecific channel-forming (porin) activity without deletion mutations (50). *H. pylori* has six homologs of the iron-regulated outer membrane proteins (Table 1), but none of these contain Hop motifs. A second precedent, although not confirmed at the three-dimensional structure level, are the autotransporters. In this class of protein, a C-terminal β barrel is proposed to mediate the export

(and cleavage) of the N-terminal portion (27, 38). Such autotransporters include proteins as the VacA cytolysin of *H. pylori*, the immunoglobulin A protease of *Neisseria gonorrhoeae*, and the Hsr surface protein of *H. mustelae* (45, 49, 52). Cleavage of the N terminus requires a site-specific proteolytic activity, and loss of the normally cleaved residues prevents release of the N-terminal domain.

Substrate-specific porins are often constituted similarly to the nonspecific porins but are of higher molecular weight and often have smaller channel sizes. The additional residues of substrate-specific porins are found in the surface-exposed loops and fold either into the channel to form binding sites or over the top to constrict the channel entrance (51). It seems possible that the Hop and Hor proteins include nonspecific porins, specific porins, and gated porins. The BabA (HopS), BabB (HopT), and HopZ proteins have been shown to be adhesins, but they have strong C-terminal homology (Fig. 2B) to HopA and HopD, which show porin activity. Thus, it is possible that the adhesins are analogous to uncleaved autotransporters with a C-terminal β-barrel domain and an N-terminal adhesin domain which protrudes through the barrel.

The expression of outer membrane proteins and the subsequent alterations in the bacterial surface may play a role in the colonization or persistence of an *H. pylori* infection or to the severity of disease associated with chronic infection. Analysis of the two genomic sequences identified several methods by which expression of these proteins could be affected. Expression of several genes may be regulated by slipped-strand repair at either mono- or dinucleotide repeats (4, 61) and may play a role in antigenic variation and virulence during infection, similar to the opacity protein of *N. gonorrhoeae* (55). Indeed, this phenomenon was observed in J99 as repetitive sequencing revealed individual clones with different lengths of repeats (4). The same five *hop* orthologs in *H. pylori* J99 and 26695 possess these repeats, and in every case the number of dinucleotide (CT) repeats in their signal sequence differs without affecting the predicted expression status (4). Different spacings between the RBS and the initiation codon in other orthologous genes, including that for the BabA (HopS) adhesin, may also lead to an alteration in the expression level. Almost half of the genes predicted to be regulated by slipped-strand repair would affect the composition of the outer membrane, including outer membrane protein genes and those involved in LPS biosynthesis (4). Indeed, the serotype of several clinical isolates correlated with the varying length of a homopolymeric tract and the resulting expression status of the α-1,3 fucosyltransferase genes (4a).

The alteration in location or transcriptional direction with respect to the origin of replication may also affect the expression level of these proteins. Of the 10 organizational differences observed in the gene order between *H. pylori* J99 and 26695, two involved members of the Y-Hop subfamily (4). One was a simple inversion of 2.5 kb between the inverted repeats which encoded the conserved C terminus of the HopO and HopP proteins, whereas the other was a gene shuffling of the BabA (Hops) and BabB (HopT) adhesin genes (4). Together with the other organizational differences observed between the two strains (4), seven outer membrane protein genes in Table 1 are located in a different transcriptional orientation. There also appears to be a bias to the direction of transcription within several of the families of genes encoding outer membrane proteins. All of the genes which encode the Y-Hop proteins are located on the complementary strand except *hopP* (see above) and *hopN*, which represents a duplicated allele. Conversely, all of the F-Hop proteins are encoded on the plus strand. A similar bias is seen with all of the *hof* genes. In both strains, all are transcribed on the plus strand, except the 26695 *hofB* gene, whose relative location is inverted and translocated due an organizational difference (4). The organizational differences and shuffling of the outer membrane protein genes observed between *H. pylori* J99 and 26695 have also been detected in other *H. pylori* isolates (L. L. Ling, D. T. Moir, R. A. Alm, D. M. Mills, B. M. Andrews, G. F. Vovis, and T. J. Trust, unpublished data), which suggests that a subtle mechanism of regulation may be occurring. The reason(s) for such possible regulatory mechanisms in *H. pylori* is not known. However, this ability to possibly perform phase variation may play a role in evading the host's immune system and could be especially important in light of the limited sequence variation between orthologs.

REFERENCES

1. **Akopyants, N. S., Q. Jiang, D. E. Taylor, and D. E. Berg.** 1997. Corrected identity of isolates of *Helicobacter pylori* reference strain NCTC11637. Helicobacter **2:**48–52.
2. **Akopyanz, N., N. O. Bukanov, T. U. Westblom, and D. E. Berg.** 1992. PCR-based RFLP analysis of DNA sequence diversity in the gastric pathogen *Helicobacter pylori*. Nucleic Acids Res. **20:**6221–6225.
3. **Akopyanz, N., N. O. Bukanov, T. U. Westblom, S. Kresovich, and D. E. Berg.** 1992. DNA diversity among clinical isolates of *Helicobacter pylori* detected by PCR-based RAPD fingerprinting. Nucleic Acids Res. **20:**5137–5142.
4. **Alm, R. A., L. L. Ling, D. T. Moir, B. 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-Nickelsen, 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:**186–190.

4a. **Appelmelk, B. J., S. L. Martin, M. A. Monteiro, C. A. Clayton, A. A. McColm, P. Zheng, T. Verboom, J. J. Maaskant, D. H. van den Eijnden, C. H. Hokke, M. B. Perry, C. M. J. E. Vandenbroucke-Grauls, and J. G. Kusters.** 1999. Phase variation in *Helicobacter pylori* lipopolysaccharide due to changes in the lengths of poly(C) tracts in α3-fucosyltransferase genes. Infect. Immun. **67:**5361–5366.

5. **Atherton, J. C., P. Cao, R. M. Peek Jr., M. K. R. Tummuru, M. J. Blaser, and T. L. Cover.** 1995. Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. J. Biol. Chem. **270:**17771–17777.
6. **Baldermann, C., A. Lupas, J. Lubieniecki, and H. Engelhardt.** 1998. The regulated outer membrane protein Omp21 from *Comamonas acidovorans* is identified as a member of a new family of eight-stranded β-sheet proteins by its sequence and properties. J. Bacteriol. **180:**3741–3749.
7. **Bereswill, S., F. Lichte, T. Vey, F. Fassbinder, and M. Kist.** 1998. Cloning and characterization of the fur gene from *Helicobacter pylori*. FEMS Microbiol. Lett. **159:**193–200.
8. **Berg, D. E., R. H. Gilman, J. Lelwala-Guruge, K. Srivastava, Y. Valdez, J. Watanabe, J. Miyagi, N. S. Akopyants, A. Ramirez-Ramos, T. H. Yoshiwara, S. Recavarren, and R. Leon-Barua.** 1997. *Helicobacter pylori* populations in Peruvian patients. Clin. Infect. Dis. **25:**996–1002.
9. **Bina, J. E., R. A. Alm, M. Uria-Nickelsen, S. R. Thomas, T. J. Trust, and R. E. W. Hancock.** 2000. *Helicobacter pylori* uptake and efflux: basis for intrinsic susceptibility to antibiotics in vitro. Antimicrob. Agents Chemother. **44:**248–254.

9a. **Bina, J., M. Bains, and R. E. W. Hancock.** 2000. Functional expression in *Escherichia coli* and membrane topology of porin HopE, a member of a large family of conserved proteins in *Helicobacter pylori*. J. Bacteriol. **182:**2370–2375.

10. **Black, P. N.** 1991. Primary sequence of the *Escherichia coli fadL* gene encoding an outer membrane protein required for long-chain fatty acid transport. J. Bacteriol. **173:**435–442.
11. **Bölin, I., H. Lönroth, and A. M. Svennerholm.** 1995. Identification of *Helicobacter pylori* by immunological dot blot method based on reaction of a species-specific monoclonal antibody with a surface-exposed protein. J. Clin. Microbiol. **33:**381–384.
12. **Buchanan, S. K., B. S. Smith, L. Venkatramani, D. Xia, L. Esser, M. Palnitkar, R. Chakraborty, D. van der Helm, and J. Diesenhofer.** 1999. Crystal structure of the outer membrane active transporter FepA from *Escherichia coli*. Nat. Struct. Biol. **6:**56–63.
13. **Burnens, A., U. Stucki, J. Nicolet, and J. Frey.** 1995. Identification and characterization of an immunogenic outer membrane protein of *Campylobacter jejuni*. J. Clin. Microbiol. **33:**2826–2832.
14. **Cover, T. L., and M. J. Blaser.** 1992. *Helicobacter pylori* and gastroduodenal disease. Annu. Rev. Med. **42:**135–145.
15. **Cover, T. L., and M. J. Blaser.** 1992. Purification and characterization of the vacuolating toxin from *Helicobacter pylori*. J. Biol. Chem. **267:**10570–10575.
16. **Cover, T. L., M. K. Tummuru, P. Cao, S. A. Thompson, and M. J. Blaser.** 1994. Divergence of genetic sequences for the vacuolating cytotoxin among

Helicobacter pylori strains. J. Biol. Chem. **269:**10566–10573.
17. **Doig, P., M. M. Exner, R. E. W. Hancock, and T. J. Trust.** 1995. Isolation and characterization of a conserved porin protein from *Helicobacter pylori*. J. Bacteriol. **177:**5447–5452.
18. **Doig, P., and T. J. Trust.** 1994. Identification of surface-exposed outer membrane antigens of *Helicobacter pylori*. Infect. Immun. **62:**4526–4533.
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. Guild, D. T. Moir, G. F. Vovis, and T. J. Trust.** 1999. The physiology of *Helicobacter pylori* predicted from a genomic comparison of two strains. Microbiol. Mol. Biol. Rev. **63:** 675–707.
20. **Eaton, K. A., D. R. Morgan, and S. Krakowka.** 1989. *Campylobacter pylori* virulence factors in gnotobiotic piglets. Infect. Immun. **57:**1119–1125.
21. **Exner, M.** 1998. Ph.D. thesis. University of British Columbia, Vancouver, British Columbia, Canada.
22. **Exner, M. M., P. Doig, T. J. Trust, and R. E. W. Hancock.** 1995. Isolation and characterization of a family of porin proteins from *Helicobacter pylori*. Infect. Immun. **63:**1567–1572.
23. **Ferguson, S. D., E. Hofmann, J. W. Coulton, K. Diederichs, and W. Welte.** 1998. Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide. Science **282:**2215–2220.
24. **Go, M. F., V. Kapur, D. Y. Graham, and J. M. Musser.** 1996. Population genetic analysis of *Helicobacter pylori* by multilocus enzyme electrophoresis: extensive allelic diversity and recombinational population structure. J. Bacteriol. **178:**3934–3938.
25. **Han, J., E. Yu, I. Lee, and Y. Lee.** 1997. Diversity among clinical isolates of *Helicobacter pylori* in Korea. Mol. Cells **7:**544–547.
26. **Handt, L. K., J. G. Fox, I. H. Stalis, R. Rufo, G. Lee, J. Linn, X. Li, and H. Kleanthous.** 1995. Characterization of feline *Helicobacter pylori* strains and associated gastritis in a colony of domestic cats. J. Clin. Microbiol. **33:**2280–2289.
27. **Henderson, I. R., F. Navarro-Garcia, and J. P. Nataro.** 1998. The great escape: structure and function of the autotransporter proteins. Trends Microbiol. **6:**370–378.
28. **Huang, H., D. Jeanteur, F. Pattus, and R. E. W. Hancock.** 1995. Membrane topology and site-specific mutagenesis of *Pseudomonas aeruginosa* porin OprD. Mol. Microbiol. **16:**931–941.
29. **Hunt, R. H.** 1996. The role of *Helicobacter pylori* in pathogenesis: the spectrum of clinical outcomes. Scand. J. Gastroenterol. Suppl. **220:**3–9.
30. **Ilver, D., A. Arnqvist, J. Ögren, I. M. Frick, D. Kersulyte, E. T. Incecik, D. E. Berg, A. Covacci, L. Engstrand, and T. Borén.** 1998. *Helicobacter pylori* adhesin binding fucosylated histo-blood group antigens revealed by retagging. Science **279:**373–377.
31. **Jiang, Q., K. Hiratsuka, and D. E. Taylor.** 1996. Variability of gene order in different *Helicobacter pylori* strains contributes to genome diversity. Mol. Microbiol. **20:**833–842.
32. **Kersulyte, D., H. Chalkauskas, and D. E. Berg.** 1999. Emergence of recombinant strains of *Helicobacter pylori* during human infection. Mol. Microbiol. **31:**31–43.
33. **Kostrzynska, M., J. D. Betts, J. W. Austin, and T. J. Trust.** 1991. Identification, characterization, and spatial localization of two flagellin species in *Helicobacter pylori* flagella. J. Bacteriol. **173:**937–946.
34. **Kostrzynska, M., P. W. O'Toole, D. E. Taylor, and T. J. Trust.** 1994. Molecular characterization of a conserved 20-kilodalton membrane-associated lipoprotein antigen of *Helicobacter pylori*. J. Bacteriol. **176:**5938–5948.
35. **Labigne, A., and H. de Reuse.** 1996. Determinants of *Helicobacter pylori* pathogenicity. Infect. Agents Dis. **5:**191–202.
36. **Lampe, M. F., R. J. Suchland, and W. E. Stamm.** 1993. Nucleotide sequence of the variable domains within the major outer membrane protein gene from serovariants of *Chlamydia trachomatis*. Infect. Immun. **61:**213–219.
37. **Lee, A., J. O'Rourke, M. C. DeUngria, B. Robertson, G. Daskalopoulos, and M. F. Dixon.** 1997. A standardized mouse model of *Helicobacter pylori* infection: introducing the Sydney strain. Gastroenterology **112:**1386–1397.
38. **Loveless, B. J., and M. H. Saier.** 1997. A novel family of channel-forming, autotransporting bacterial virulence factors. Mol. Membr. Biol. **14:**113–123.
39. **Luke, C. J., and C. W. Penn.** 1995. Identification of a 29kDa flagellar sheath protein in *Helicobacter pylori* using a murine monoclonal antibody. Microbiology **141:**597–604.
40. **Maier, C., E. Bremer, A. Schmid, and R. Benz.** 1988. Pore-forming activity of the Tsx protein from the outer membrane of *Escherichia coli*. Demonstration of a nucleoside-specific binding site. J. Biol. Chem. **263:**2493–2499.
41. **Marshall, B. J., H. Royce, D. I. Annear, C. S. Goodwin, J. W. Pearman, J. R. Warren, and J. A. Armstrong.** 1984. Original isolation of *Campylobacter pyloridis* from human gastric mucosa. Microbios Lett. **25:**83–88.
42. **Marshall, D. G., W. G. Dundon, S. M. Beesley, and C. J. Smyth.** 1998. *Helicobacter pylori*—a conundrum of genetic diversity. Microbiology **144:** 2925–2939.
43. **Matysiak-Budnik, T., and F. Megraud.** 1997. Epidemiology of *Helicobacter pylori* infection with special reference to professional risk. J. Physiol. Pharmacol. **48**(Suppl. 4)**:**3–17.
44. **Odenbreit, S., M. Till, D. Hofreuter, G. Faller, and R. Haas.** 1999. Genetic and functional characterization of the *alpAB* gene locus essential for the adhesion of *Helicobacter pylori* to human gastric tissue. Mol. Microbiol. **31:**1537–1548.
45. **O'Toole, P. W., J. W. Austin, and T. J. Trust.** 1994. Identification and molecular characterization of a major ring-forming surface protein from the gastric pathogen *Helicobacter mustelae*. Mol. Microbiol. **11:**349–361.
46. **O'Toole, P. W., L. Janzon, P. Doig, J. Huang, M. Kostrzynska, and T. J. Trust.** 1995. The putative neuraminyllactose-binding hemagglutinin HpaA of *Helicobacter pylori* CCUG 17874 is a lipoprotein. J. Bacteriol. **177:**6049–6057.
47. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence analysis. Proc. Natl. Acad. Sci. USA **85:**2444–2448.
48. **Peck, B., M. Ortkamp, K. D. Diehl, E. Hundt, and B. Knapp.** 1999. Conservation, localization and expression of HopZ, a protein involved in adhesion of *Helicobacter pylori*. Nucleic Acids Res. **27:**3325–3333.
49. **Pohlner, J., R. Halter, K. Beyreuther, and T. F. Meyer.** 1987. Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. Nature **325:**458–462.
50. **Rutz, J. M., J. Liu, J. A. Lyons, J. Goranson, S. K. Armstrong, M. A. McIntosh, J. B. Feix, and P. E. Klebba.** 1992. Formation of a gated channel by a ligand-specific transport protein in the bacterial outer membrane. Science **258:**471–475.
51. **Schirmer, T., T. A. Keller, Y. F. Wang, and J. P. Rosenbusch.** 1995. Structural basis for sugar translocation through maltoporin channels at 3.1 A resolution. Science **267:**512–514.
52. **Schmitt, W., and R. Haas.** 1994. Genetic analysis of the *Helicobacter pylori* vacuolating cytotoxin: structural similarities with the IgA protease type of exported protein. Mol. Microbiol. **12:**307–319.
53. **Schuler, G. D., S. F. Altschul, and D. J. Lipman.** 1991. A workbench for multiple alignment construction and analysis. Proteins Struct. Funct. Genet. **9:**180–190.
54. **Sherburne, R., and D. E. Taylor.** 1995. *Helicobacter pylori* expresses a complex surface carbohydrate, Lewis X. Infect. Immun. **63:**4564–4568.
55. **Stern, A., M. Brown, P. Nickel, and T. F. Meyer.** 1986. Opacity genes in *Neisseria gonorrhoeae*: control of phase and antigenic variation. Cell **47:**61–71.
56. **Struyve, M., M. Moons, and J. Tommassen.** 1991. Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. J. Mol. Biol. **218:**141–148.
57. **Szmelcman, S., M. Schwartz, T. J. Silhavy, and W. Boos.** 1976. Maltose transport in *Escherichia coli* K12. A comparison of transport kinetics in wild-type and lambda-resistant mutants as measured by fluorescence quenching. Eur. J. Biochem. **65:**13–19.
58. **Takami, S., T. Hayashi, Y. Tonokatsu, T. Shimoyama, and T. Tamura.** 1993. Chromosomal heterogeneity of *Helicobacter pylori* isolates by pulsed-field gel electrophoresis. Zentbl. Bakteriol. **280:**120–127.
59. **Taylor, D. E., M. Eaton, N. Chang, and S. M. Salama.** 1992. Construction of a *Helicobacter pylori* genome map and demonstration of diversity at the genome level. J. Bacteriol. **174:**6800–6806.
60. **Tee, W., J. Lambert, R. Smallwood, M. Schembri, B. C. Ross, and B. Dwyer.** 1992. Ribotyping of *Helicobacter pylori* from clinical specimens. J. Clin. Microbiol. **30:**1562–1567.
61. **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. Fitzegerald, 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.
62. **Whittam, T. S.** 1995. Genetic population structure and pathogenicity in enteric bacteria. Symp. Soc. Gen. Microbiol. **52:**217–245.
63. **Worst, D. J., B. R. Otto, and J. de Graaff.** 1995. Iron-repressible outer membrane proteins of *Helicobacter pylori* involved in heme uptake. Infect. Immun. **63:**4161–4165.

Editor: A. D. O'Brien