JOURNAL OF BACTERIOLOGY, Feb. 2002, p. 992–1002 Vol. 184, No. 4
0021-9193/02/$04.00+0 DOI: 10.1128/JB.184.4.992–1002.2002

Transposable Element IS*Hp608* of *Helicobacter pylori*: Nonrandom Geographic Distribution, Functional Organization, and Insertion Specificity

Dangeruta Kersulyte,[1] Billie Velapatiño,[1,2] Giedrius Dailide,[1] Asish K. Mukhopadhyay,[1] Yoshiyuki Ito,[1] Lizbeth Cahuayme,[2] Alan J. Parkinson,[3] Robert H. Gilman,[2] and Douglas E. Berg[1]*

Departments of Molecular Microbiology and Genetics, Washington University School of Medicine, St. Louis, Missouri 63110[1]; Department of Pathology, Universidad Peruana Cayetano Heredia, Lima, Peru[2]; and Arctic Investigations Program, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Anchorage, Alaska[3]

Received 3 October 2001/Accepted 19 November 2001

A new member of the IS*605* transposable element family, designated IS*Hp608*, was found by subtractive hybridization in *Helicobacter pylori*. Like the three other insertion sequences (ISs) known in this gastric pathogen, it contains two open reading frames (*orfA* and *orfB*), each related to putative transposase genes of simpler (one-gene) elements in other prokaryotes; *orfB* is also related to the *Salmonella* virulence gene *gipA*. PCR and hybridization tests showed that IS*Hp608* is nonrandomly distributed geographically: it was found in 21% of 194 European and African strains, 14% of 175 Bengali strains, 43% of 131 strains from native Peruvians and Alaska natives, but just 1% of 223 East Asian strains. IS*Hp608* also seemed more abundant in Peruvian gastric cancer strains than gastritis strains (9 of 14 versus 15 of 45, respectively; $P = 0.04$). Two IS*Hp608* types differing by ~11% in DNA sequence were identified: one was widely distributed geographically, and the other was found only in Peruvian and Alaskan strains. Isolates of a given type differed by ≤2% in DNA sequence, but several recombinant elements were also found. IS*Hp608* marked with a resistance gene was found to (i) transpose in *Escherichia coli*; (ii) generate simple insertions during transposition, not cointegrates; (iii) insert downstream of the motif 5′-TTAC without duplicating target sequences; and (iv) require *orfA* but not *orfB* for its transposition. IS*Hp608* represents a widespread family of novel chimeric mobile DNA elements whose further analysis should provide new insights into transposition mechanisms and into microbial population genetic structure and genome evolution.

The hundreds of known insertion sequences (ISs) of prokaryotes are diverse in overall structure and in detailed mechanism, specificity, and regulation of transposition (8, 21, 36). Typically, an IS found in one bacterial strain will be absent from many other strains of that species yet be closely related to ISs in other microbes. This pattern bears witness to a rich heritage of interspecies DNA transfer, which has resulted in the spread of individual ISs to taxonomic groups remote from those in which they arose. Their abundance also reflects the action of transposase proteins, which mediate IS movement without need for extensive DNA sequence homology; to the ability of these elements to proliferate in host genomes by transposition itself; and to contributions that these elements or genes associated with them (e.g., in composite drug resistance transposons) sometimes make to bacterial fitness.

Many ISs and other tranposable elements specify just one transposase protein that acts as a multimer on matched (inverted repeat) sequences at each element end. Others, exemplified by Tn*7*, specify two different proteins that form a more complex transposase. With Tn*7*, each transposase protein has a distinct role in transposition, and additional Tn*7*-encoded proteins help select insertion sites and affect the efficiency of transposition, in part through interactions with host proteins (9, 31). A third type of element, exemplified by IS*605* of the gastric pathogen *Helicobacter pylori*, contains just two transposition-related genes, each of which has protein level homology to the single putative transposase genes of other simpler one-gene ISs and thus may be of different phylogenetic origin (18, 19). The termini of these latter two types of elements tend to have less inverted repeat character, which implies that each end may be acted on differently by proteins of the transposition complex.

Each of the two other ISs found to date in *H. pylori* (IS*606* and IS*607*) is related to IS*605* by protein level homologies in *orfB* (some 25 to 35% amino acid sequence identity) (18, 19), as is the new IS*Hp608* element described here (Fig. 1A). This *orfB* (putative transposase gene) is also related to *gipA*, a gene found in a *Salmonella* prophage that contributes to virulence during murine infection (32).

Two branches of the IS*605* family can be distinguished based on protein level homologies in *orfA*, one represented by IS*607*, whose transposition was found to be *orfA* dependent and *orfB* independent (19), and the other represented by IS*605* and IS*606* (18) and also the present IS*Hp608*. The *orfA*s of IS*605*, IS*606*, and IS*Hp608* are protein level homologs of the putative transposase gene of IS*200*, an element that is abundant in natural isolates of *Salmonella* and *Escherichia coli*. Only a few cases of IS*200* transposition have ever been detected, however, and the mechanisms of its movement are not understood (8, 21, 36).

Sequences of IS*605*, IS*606*, and IS*607* were each found in

* Corresponding author. Mailing address: Department of Molecular Microbiology, Campus Box 8230, Washington University School of Medicine, St. Louis, MO 63110. Phone: (314) 362-2772. Fax: (314) 362-1232. E-mail: berg@borcim.wustl.edu.

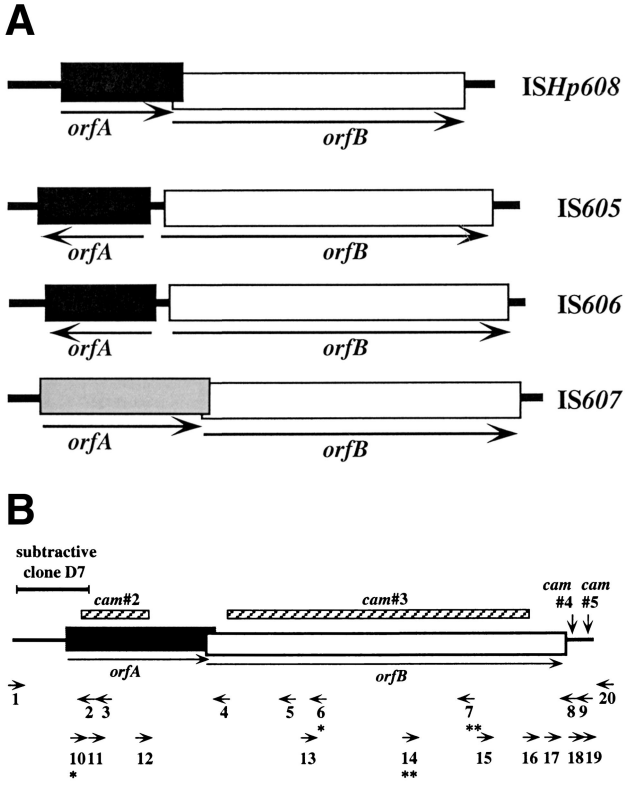

FIG. 1. IS element maps. (A) Structures of ISs of *H. pylori*. IS*Hp608* is 1,833 bp long, making it slightly smaller than its three other known *H. pylori* relatives (size range, 1,888 to 2,028 bp [18, 19]). Boxes represent ORFs: those with the same shading pattern are related at the protein level (range of 28 to 36% amino acid sequence identities). *orfA* sizes are 156, 148, 139, and 218 codons and *orfB* sizes are 384, 427, 446, and 420 codons in IS*Hp608*, IS*605*, IS*606*, and IS*607*, respectively. *orfA* and *orfB* overlap by 10 codons in IS*Hp608*, which is reminiscent of the arrangement in IS*607* (nine-codon overlap). Distances from left and right ends to the 5′ end of *orfA* and 3′ end of *orfB* in a representative IS*Hp608* are 171 bp and 75 bp, respectively (accession no. AF357224). GenBank accession nos. for IS*605*, IS*606*, and IS*607* are AC000108, U95957, and AF189015, respectively. (B) Positions of oligonucleotide primers, the subtractive clone that led to the discovery of IS*Hp608*, and mutations generated in IS*Hp608*. The locations and orientations of PCR and sequencing primers are denoted by arrowheads; these primers are described in detail in Table 1. Primers 1 and 20 are specific to sequences in jhp0924 (an ORF in strain J99 [4]) located to the left and right of IS*Hp608*, respectively, in strain PeCan2A. * and ** identify primers specific for IS*Hp608* types 2 and 3, respectively (defined in Results). Subtractive clone D7 is a 242-bp *Alu*I fragment from strain PeCan18B that begins approximately 30 to 34 bp from the left end of IS*Hp608* copy 2. *cam*#2, *cam*#3, *cam*#4, and *cam*#5 represent mutant derivatives of IS*Hp608* from strain PeCan2A that contain the *cam* (resistance gene) cassette inserted into (*cam*#4 and *cam*#5) or replacing (*cam*#2 and *cam*#3) particular IS*Hp608* sequences. In *cam*#4 (A^+B^+), *cam* was inserted just downstream of *orfB* and 68 bp from the rightmost end of IS*Hp608* (generated with primers 8 and 18); in *cam*#2 (A^-B^+), 154 bp of *orfA* sequence was replaced by the *cam* gene (generated with primers 2 and 12); in *cam*#3 (A^+B^-), 948 bp of *orfB* sequence was replaced by *cam* (generated with primers 4 and 16); in *cam*#5 (A^+B^+), *cam* was inserted just 22 bp from the IS*Hp608* right end, within a 22-bp direct repeat (putative transposase binding site) (generated with primers 9 and 19), as detailed in Results.

only a subset of *H. pylori* strains, always in chimeric (two-gene) elements, as depicted in Fig. 1A. In particular, none of the several hundred *H. pylori* strains tested to date was found to contain only *orfA* (or only *orfB*) without the cognate *orfB* (or *orfA*), positioned as in Fig. 1A (18, 19). Formal models to explain such association include (i) OrfA and OrfB proteins serving together as the functional transposase, at least in *H. pylori*; (ii) transposition mediated by just one of these proteins but regulated (in terms of efficiency or specificity) by the other; (iii) each protein mediating transposition in a different set of bacterial species (implying relatively recent acquisition by *H. pylori*); or (iv) one gene needed for transposition and the other contributing to fitness.

As background to the present studies, *H. pylori* chronically infects more than half of all people worldwide and is implicated in peptic ulcer disease and gastric cancer. Infections often begin in infancy, resulting from preferential transmission within the family or perhaps the local community, and tend to last for decades once established; new adult infections are rare (10, 12, 27, 38). *H. pylori* is also one of the most genetically diverse of bacterial species: independent isolates typically differ from one another by some 3 to 5% in DNA sequence in essential genes and can also differ markedly in gene content and chromosomal gene arrangement (1, 3, 4). This diversity is enhanced by recombination. In general, little if any linkage is found between alleles at different polymorphic sites in collections of strains from the same geographic region, a pattern referred to as free recombination (33).

Superimposed on the extensive recombination between strains from the same region are indications that *H. pylori* gene pools differ geographically. This is based primarily on distributions of DNA sequence motifs in the virulence-associated *vacA* and *cagA* genes and of insertion and deletion motifs downstream of *cagA* (17, 20, 26, 37, 39). Strains from Europe, India, and East Asia were generally easily distinguished from one another using these tests, whereas those from a largely native population in Lima, Peru, seemed closely related to those of Spain. Much of the genetic diversity in local *H. pylori* populations and geographic differences in gene pools can be attributed to *H. pylori*'s patterns of transmission (preferentially to children and within families), the extraordinary chronicity of infection, and the relative isolation of the strain(s) carried by any given person, noted above. In consequence, strains diverge from one another by random genetic drift and by selection for adaptation to the new gastric environments each person may present (15). Despite occasional mixed infection, there is only limited direct competition between *H. pylori* strains of different lineages and no effective selection for just a few genotypes that might be best suited for most people worldwide.

Here we describe the discovery and characterization of IS*Hp608*, found first in an *H. pylori* strain from Peru, features of its structure and transposition behavior, and its remarkably nonrandom distribution in human populations.

MATERIALS AND METHODS

General methods. Standard procedures were used for *H. pylori* growth in a microaerobic atmosphere on brain heart infusion agar (Difco) containing 10% horse blood (2, 18, 19), *E. coli* growth in L broth or on L agar, genomic DNA isolation, DNA electrophoresis, and transformation of competent *E. coli* cells (29). Antibiotics, when needed, were used at the following concentrations (in

TABLE 1. Primers used in this study

Type and primer no.	Primer name	Sequence	Location[a]
Specific for flanking *jhp0924* sequence[b]			
1	jhp0924R	5′-TGCGCTAGTGCAAAAGTTAC	Left of IS*Hp608* in PeCan2A
20	jhp0924F	5′-TTATTAAAGTCATTAGGCAAATT	Right of IS*Hp608* in PeCan2A
Leftward-facing[c] IS*Hp608*-specific			
2	608-3a	5′-GATATGGTATTTGCAAGAATAGAC	233 bp from left end of IS*Hp608*
3	608-1	5′-ACTAATACCTTACGCCTGTA	268 bp from left end of IS*Hp608*
4	608-F1	5′-CCATAACGCCTTAATAGTGTGTC	683 bp from left end of IS*Hp608*
5	608-F	5′-ATCCTCTCAGTCAAATCTTGCATG	851 bp from left end of IS*Hp608*
6	608a112-R[d]	5′-TTTTGTCAATCTTACTCTTA	958 bp from left end of IS*Hp608*
7	608a113-R[e]	5′-TAGAGAGCGGTTGCAAGCACG	1,444 bp from left end of IS*Hp608*
8	608neutral-Fnew	5′-CAAGCAAACTCAGCCTACTAG	1,772 bp from left end of IS*Hp608*
9	608neutral-F	5′-CTAAAGCTAGGGGATTCTAGC	1,815 bp from left end of IS*Hp608*
Rightward-facing[c] IS*Hp608*-specific			
10	608a112-F[d]	5′-TATATAAAAACAACCACAACA	1,647 bp from right end of IS*Hp608*
11	608-3out	5′-GTCTATTCTTGCAAATACCATATCG	1,624 bp from right end of IS*Hp608*
12	608-R	5′-CAAGCTTTGGAGTGATGAAGTTC	1,446 bp from right end of IS*Hp608*
13	608-4out	5′-GTAGCCAACTACAAGTCTTTCAC	921 bp from right end of IS*Hp608*
14	608a113-F[e]	5′-CGAGTTCAATGGCAAACCCAAG	620 bp from right end of IS*Hp608*
15	608-6out	5′-TAGCGAGTTTGTGAAGATCCT	343 bp from right end of IS*Hp608*
16	608-7out	5′-AAGAAAGAGAATACCACTGCA	200 bp from right end of IS*Hp608*
17	608-8out	5′-CTATCAACATTCATAGGGTTG	140 bp from right end of IS*Hp608*
18	608neutral-Rnew	5′-TGCTTGATCCCAATTTTCCTC	67 bp from right end of IS*Hp608*
19	608neutral-R	5′-TTAGCTATGGGGAGTATGTCA	22 bp from right end of IS*Hp608*
Universal primers for adapter PCR			
C1		5′-GTACATATTGTCGTTAGAACGCG	Adapter-specific primer (TaKaRa)
C2		5′-TAATACGACTCACTATAGGGAGA	Adapter-specific primer (TaKaRa)

[a] Where specified in base pairs, location refers to position of 5′ end of primer in sequence of IS*Hp608* from *H. pylori* strain PeCan2A (GenBank accession no. AF357223).
[b] *jhp094* refers to an ORF of unknown function in the genome of *H. pylori* strain J99 (4). The IS*Hp608* element found in strain PeCan2A is inserted in this ORF.
[c] Left and right ends of IS*Hp608* are diagrammed in Fig. 1.
[d] IS*Hp608* type 2 specific.
[e] IS*Hp608* type 3 specific

micrograms per milliliter): ampicillin (Amp), 100; chloramphenicol (Cam), 25; and streptomycin (Str), 150. Plasmid DNAs were isolated from *E. coli* cultures using the Qiagen prep spin miniprep kit (Qiagen, Chatsworth, Calif.).

High-molecular-weight genomic DNA was isolated from *H. pylori* and *E. coli* by a hexadecyltrimethylammonium method (5). Restriction digestion and ligation were carried out as recommended by the manufacturers (generally New England Biolabs, Beverly, Mass.). DNA fragments used for cloning and hybridization were purified from 1% agarose gels using the Geneclean II kit (Bio 101, Vista City, Calif.). A subtractive DNA library was made using the PCR-Select bacterial genome subtraction kit (Clontech) (2). Genomic DNA from a strain from a gastric cancer patient was used as the driver, and a pool of DNAs from three strains from gastritis-only patients was used as the tester. Southern blotting and hybridization were performed using Hybond N+ membranes (Amersham) according to the manufacturer's recommendations. DNA was labeled using the ECL kit (Amersham).

Specific PCR was carried out in 25-μl volumes, containing 5 to 10 ng of DNA, 1 U of *Taq* polymerase (Biolase; Midwest Scientific, St. Louis, Mo.), or Expand High Fidelity *Taq-Pwo* polymerase mixture (Boehringer-Mannheim, Indianapolis, Ind.), 5 pmol of each primer (Table 1), and 0.25 mM of each deoxynucleoside triphosphate (dNTP), in a standard buffer for 30 cycles with the following cycling parameters: denaturation for 30 s at 94°C; annealing for 30 s at a temperature appropriate for the primer (generally 55°C); and DNA synthesis for 1 min or as appropriate (1 min per kb) at 72°C. PCR primers used in these experiments are listed in Table 1, and their positions are diagrammed in Fig. 1B. When needed, PCR products were cloned into *Eco*RV-cleaved pBluescript plasmid DNA containing an additional terminal T (22).

DNA sequencing was carried out using a Big Dye Terminator DNA sequencing kit (Perkin-Elmer) and ABI automated sequencers. Direct sequencing on chromosomal DNA was done with 5 μl of chromosomal DNA (1 to 2 μg), 1 μl of primer (10 pmol/μl), and 6 μl of Big Dye under the following conditions (oil free): 96°C for 5 min and then 90 cycles of denaturation for 15 s at 96°C; annealing for 10 s at a temperature appropriate to the primer; and extension for 4 min at 60°C (Perkin Elmer 2400). DNA sequence editing and analysis were performed with programs in the Genetics Computer Group (Madison, Wis.) package, programs and data in *H. pylori* genome sequence databases (4, 35), and Blast and Pfam (version 5.3) homology search programs (http://www.ncbi.nlm.nih.gov/blast/blast.cgi and http://pfam.wustl.edu/hmmsearch.shtml).

Bacterial strains and plasmids. The *H. pylori* strains used for subtractive library construction were PeCan18B, from a Peruvian gastric cancer patient (tester), and a pool of SJM180A, SJM184A, and SJM189A, from Peruvian gastritis patients (driver). Peruvian gastric cancer strains (designated with the prefix PeCan) were cultured from biopsy samples provided by Alejandro Bussalleu and Juan Combe of the Servicio Universitario de Apoyo, Universidad Peruana Cayetano Heredia (UPCH) in Lima. These and all other biopsy samples were from persons with gastroduodenal complaints, were part of a series taken to help in diagnosis of disease, and were obtained with written informed consent under protocols approved by the local (UPCH) human studies committee. Most other *H. pylori* strains used here were from the Berg laboratory collection and have been described in detail elsewhere (7, 11, 20, 25).

The following notes on the ethnicities of patients and clinical disease associations may be useful in evaluating the present data. The 45 Peruvian gastritis strains were from persons of primarily Amerindian ancestry living in the shanty town of San Juan de Miraflores in Lima, Peru (strains designated by the prefix PeSJM). Each of the 35 strains for which records were available were from patients with gastritis only, not peptic ulcer disease. Twenty-two of the 34 Spanish strains for which records were available were from patients with peptic ulcer disease, and the other 12 were from patients with gastritis only. Sixty-three of the 72 Lithuanian strains were from gastritis patients, and the others were from peptic ulcer patients (three gastric and six duodenal). The 23 African strains were from native black African residents of Soweto (17 strains) and The Gambia (six strains) and were in each case from persons with gastritis only. The 118 Indian strains were from middle or lower middle class residents of Calcutta and were from patients with duodenal ulcer disease (96 strains) or with gastritis only (22 strains). The Bangladeshi strains were from residents of the city of Dhaka and were kindly provided by Motiur Rahman and colleagues at International Centre for Diarrhoeal Diseases Research in Bangladesh. Twelve of the 24 Japanese strains for which records were available were from gastric cancer patients, and the other 12 were from gastric ulcer patients. Most Alaska native strains were from Inupiaq- and Yupik-speaking Eskimo peoples with gastric complaints living in Anchorage and in native villages scattered along the Alaska coast, ranging from the Kenai Peninsula and Anchorage to Barrow on the North Slope.

The *E. coli* K-12 strains used in this study were DH5α, the routine host for recombinant DNA plasmid cloning and DNA preparation; DB1683, which is Str^s and contains pOX38 (an IS-free derivative of the F fertility factor) and was used

as the donor in bacterial conjugation; and MC4100 which is F^- and Str^r and was used as the recipient in conjugation (6). The plasmid pBluescript SK+/− (Stratagene, La Jolla, Calif.) (hereafter called pBS) was used as a cloning vector. A chloramphenicol resistance cassette (*cam*), developed initially by D. E. Taylor as a selectable marker for *H. pylori*, was derived from plasmid pBSC103 by restriction with *Sma*I plus *Eco*RV (as in reference 18) and inserted into IS*Hp608* to generate mutations and allow selection for IS*Hp608* transposition. IS*Hp608* transposition donor plasmids were constructed in the pBS plasmid vector using DNAs that had been PCR amplified with Expand High Fidelity *Taq-Pwo* DNA polymerase mixture (Boehringer-Mannheim) in place of standard *Taq* DNA polymerase to minimize the risk of mutation during PCR.

Plasmid construct 1 [pBS:IS*Hp608*wt(A^+B^+)] contained an entire IS*Hp608* element from strain PeruCan2A, a segment obtained by PCR with primers 1 and 20, which are specific to sequences flanking the site of IS*Hp608* insertion in this strain (Fig. 1B). The PCR product was then cloned into *Eco*RV-cleaved pBS. Plasmid 2 [pBS:IS*Hp608orfA*:*cam*(A^-B^+)] was constructed by linearizing plasmid 1 above by PCR with outward-facing primers 12 and 2 (Fig. 1B) and then ligating the product with *cam*, replacing 154 bp of *orfA* with this selectable marker. Plasmid 3 [pBS:IS*Hp608orfB*:*cam*(A^+B^-)] was generated by linearizing plasmid 1 by PCR with outward-facing primers 16 and 4 (Fig. 1B) and then ligating the product with *cam*, replacing 948 bp of *orfB* with this selectable marker. Plasmid 4 [pBS:IS*Hp608*neutral2(A^+B^+)] was generated by linearizing plasmid 1 by PCR with outward-facing primers 8 and 18 (Fig. 1B) and ligating the product with *cam*, inserting this selectable marker into a noncoding sequence 68 bp from the rightmost IS*Hp608* end. Plasmid 5 [pBS:IS*Hp608*neutral1(A^+B^+)] was generated by linearizing plasmid 1 by PCR with outward-facing primers 9 and 19 (Fig. 1B) and ligating the product with *cam*, inserting this selectable marker in a noncoding sequence 22 bp from the rightmost IS*Hp608* end.

IS*Hp608* transposition in *E. coli*. Strain DB1683 was transformed with plasmids containing marked IS*Hp608* elements (plasmids 2, 3, 4, and 5, above). Transposition of IS*Hp608* elements to an F factor was then scored in a mating-out assay with selection for the IS*Hp608* Cam^r marker. In brief, the cells from single transformant colonies were grown overnight with aeration, diluted 100-fold into 4 ml of LB plus 0.5% glucose in a petri dish, and incubated for another 2 h at 37°C without shaking. They were then mixed with 4 ml of an exponentially growing culture of the MC4100 recipient strain at equal cell density in a total volume of 8 ml and incubated for another 3 to 4 h at 37°C without shaking. One milliliter of mating mix was concentrated by centrifugation, and the pellet was spread on L agar containing Str and Cam. This selected for transposition of the IS*Hp608cam* element to pOX38 and then transfer of pOX38::IS*Hp608cam* to MC4100 (6, 18, 19). The values reported here are expressed as Cam^r Str^r exconjugants per initial donor cell in matings that were carried out under the same conditions in each trial. The F factor plasmid was not marked with a separate drug resistance determinant, and thus the absolute efficiency of F factor transfer was not determined. The selected Cam^r Str^r exconjugants were scored for the presence or absence of the plasmid vector (Amp^s or Amp^r phenotype, respectively).

Definition of IS*Hp608* insertion sites. Sites of IS*Hp608cam* insertion in different *H. pylori* strains were identified by direct chromosomal sequencing, using outward facing primers as detailed above and/or Adapter PCR (PCR in vitro single site amplification and cloning kit; TaKaRa; PanVera, Madison, Wis.). In Adapter PCR, total genomic DNA was digested with restriction enzyme *Sau*3A and ligated with the *Sau*3A adapter cassette supplied with the kit. The first nested PCR was carried out with cassette primer C1 and IS*Hp608* terminus-specific primers (primer 3 for the left end and primer 16 for the right end; Table 1 and Fig. 1B). A second nested PCR was then carried out using cassette primer C2 and second IS*Hp608* terminus-specific primers (2 for the left end and 17 for the right end; Fig. 1B). Gel-purified PCR fragments were sequenced directly.

Sites of IS*Hp608* insertion in *E. coli* were identified by direct chromosomal DNA sequencing with outward-facing primers and genomic DNA from IS*Hp608*-containing exconjugants which had been generated by transposition to the pOX38 F factor and transfer of pOX38::IS*Hp608* by conjugation into the MC4100 F^- recipient strain.

Phylogenetic analysis. Sequence data for phylogenetic analyses were aligned and edited using Vector NTI v0.5.2.1 (http://www.informaxinc.com/). Maximum-likelihood analyses were carried out using the Phylip 3.573c package of J. Felsenstein (http://evolution.genetics.washington.edu/phylip.html). Tree plot analysis and printing were carried out using TreeviewPPC (http://taxonomy.zoology.gla.ac.uk/rod/rod.html).

Nomenclature. The IS*Hp608* element described here was originally accessioned in GenBank as IS*608*. This element was renamed with the initials *Hp* to reflect the species in which the element was discovered, as recommended (21), upon finding that Eiichi Ohtsubo and colleagues had independently used IS*608* to name an IS from an *E. coli* pathogen (accession no. AP002563). Their IS*608* element is distantly related to IS*50* (Tn*5*) and is not related to IS*Hp608* from *H. pylori* that we describe here.

Nucleotide sequence accession no. numbers. The sequence of IS*Hp608* copy 1 from strain PeCan18B corresponds to nucleotides (nts) 1 through 1833 of the 1,833-nt sequence in GenBank accession no. AF357224; that of copy 2 from strain PeCan18B (lacking about 30 nts from the left) corresponds to nts 1 through 1787 of 1,787 nts in accession no. AF411947. The other IS*Hp608* elements were accessioned in GenBank as follows: from strain PeCan2A, nts 41 through 1872 of the 1,897 nts in accession no. AF357223; from strain India49, nts 173 through 2004 of the 2,153 nts in accession no. AF411939; from Lithuanian strain Lith5, nts 5 through 1838 of the 1,838 nts in accession no. AF411940; from strain SpainB53, nts 7 through 1839 of the 1,839 nts in accession no. AF411941; from strain PeruSJM92, nts 14 through 1836 of the 1,838 nts in accession no. AF411942; from strain AlaskaP76, nts 7 through 1832 of the 1,832 nts in accession no. AF411943; from strain AlaskaP20, nts 5 through 1827 of the 1,828 nts in accession no. AF411944; from strain India77, nts 168 through 2000 of the 2,088 nts in accession no. AF411945; and from strain PeCan4A, nts 94 through 1898 of the 1,903 nts in accession no. AF411946.

RESULTS

IS*Hp608* discovery. Subtractive hybridization was carried out to search for sequences in an *H. pylori* strain from a native Peruvian with gastric cancer (PeCan18B) that were absent from each of three strains from native Peruvians with gastritis only (SJM180A, SJM184A, and SJM189A, used as a pool). One end of one subtractive clone (D7, 242 bp long) exhibited protein level homology to amino termini of a family of putative transposases: identities of 13 of 16 amino acids to transposases from *Yersinia enterocolitica* and *Enterococcus faecium* (AJ238015 and AF125554) and 12 of 16 amino acids to transposases of IS*200* of *E. coli* and IS*1004* of *Vibrio cholerae* (2002282D and AE004380). However, no significant DNA sequence matches to known *H. pylori* ISs or other database entries were detected in BlastN homology searches. These data suggested that a previously unknown IS had been found, which we provisionally designated IS*Hp608*.

IS*Hp608* sequence and diversity. Sequencing of IS*Hp608* was initiated by primer walking, using genomic DNA of PeCan18B and primers based on the sequence of subtractive clone D7. Five primers were used in succession (primers 3, 11, 12, 13, and 15; Fig. 1B). Each of the three internal primers gave clear, unambiguous sequence traces of at least 500 nts, whereas two others (outward-facing primers 3 and 15, near left and right ends) yielded mixed traces after about 100 and 320 nts, respectively. These results suggested that this strain contained two or more copies of IS*Hp608* that were identical for nearly 1.7 kb and that either (i) the ends of these IS*Hp608* elements were traversed where sequences became mixed or (ii) the copies of this element differed in sequence at one or both ends.

This IS*Hp608* sequence contained open reading frames (ORFs) of 156 and 384 codons that were protein level homologs of *orfA* and *orfB*, respectively, in IS*605* (Fig. 1A). However, the two ORFs in IS*Hp608* were in the same orientation and overlapped by 10 codons, whereas their homologs in IS*605* and IS*606* were divergent. Their parallel orientation in IS*Hp608* is reminiscent of that in IS*607*, an element with homology to IS*Hp608* only in *orfB*, not *orfA*.

IS*Hp608* copy numbers in PeCan18B and four other IS*Hp608*-positive strains (identified in preliminary dot blot tests) were estimated in Southern blots of genomic DNAs that had been singly digested by *Hin*dIII and *Ssp*I and probed

	IS*Hp608*		
	Left	Right	
aaatacttac	CAAAAACTAA	..AGTATGTCAA	mixed	PeCan18B-1
acaatcttac			aaaatattat	HP0456
		..AGTATGTCAA	aatcttcatg	PeCan18B-2
aattccttac			aatcctcaca	jhp0942
aattctttac			aatcttcatg	HP0996
caaaaattac	CAAAA.CTAA	..AGTATGTCAA	caaatttgcc	PeCan2A
caaaagttac			caaatttgcc	jhp0924
ctatacttac	CCAAAACTAA	..AGTATGTCAA	mixed	PeCan16A
ctatacttac			acactaaaat	*orfA* IS*605*
gtgtttttac	CCAAAACTAA	..AGTATGTCAA	ccaataaagg	India49
gcgtctttac			ccaataaaga	jhp0974
ttttatttac	CCAAAACTAA	..AGTATGTCAA	taggcttgcg	India77
tttaagttac			tagacttacg	intergenic,J99
tttaagttac			taggcttgcg	intergenic,26695

FIG. 2. IS*Hp608* terminal sequences in several *H. pylori* strains. Left and right termini of IS*Hp608*s determined by sequence comparison are shown in uppercase, and flanking DNA and empty sites in reference strains 26695 and J99 (HP and jhp gene designations, respectively [4, 35]) are shown in lowercase letters. The left end of IS*Hp608* copy 1 in strain PeCan18B and flanking sequence were determined by sequencing adapter PCR products. This left end was located in a gene with low homology to HP0456 of strain 26695 and was absent from strain J99. The sequence of the right end was obtained directly from genomic DNA, but flanking sequence was not determined because this strain contained two IS*Hp608* copies. The right end of IS*Hp608* copy 2 from strain PeCan18B and flanking sequence were determined by adapter PCR and sequencing. Attempts to obtain the leftmost ~30 bp of IS*Hp608* copy 2 by PCR with jhp0942-specific primers and by sequencing directly on genomic DNA were not successful. The left and right ends of IS*Hp608* in strain PeCan2A and flanking sequences were determined by adapter PCR and by direct chromosomal sequencing using a primer matching the inferred flanking sequence, respectively. The left end of IS*Hp608* in strain PeCan16A and the flanking sequence were determined by adapter PCR; this left end was located in an IS*605 orfA* sequence. The right end was identified by genomic DNA sequencing, but the flanking sequence was not determined because this strain contained two IS*Hp608* copies and because no product was obtained after PCR using primers specific for IS*Hp608* and for the anticipated flanking IS*605* sequence. Both ends and flanking sequences of the single copy of IS*Hp608* in strain India49 were determined directly from genomic DNA. The element was found inserted into the gene corresponding to jhp0974 of reference strain J99 (4). Both ends and flanking sequences of the single copy of IS*Hp608* in strain India77 were determined directly from genomic DNA and were found inserted into an intergenic (noncoding) region (between genes jhp1113 and jhp1114 in J99 and HP1188 and HP1189 in 26695).

with an internal 460-bp PCR fragment from IS*Hp608* (made with primers 5 and 12; Fig. 1B). Two strains (PeCan4A and PeCan10B) yielded just one hybridizing band in each digest, inicating just one IS*Hp608* copy per strain, whereas PeCan18 and the two other strains (PeCan2A and PeCan16A) each yielded two hybridizing bands, indicating two IS*Hp608* copies per strain.

The sequences of IS*Hp608* insertion sites were defined more closely by adapter PCR or direct sequencing of genomic DNAs from several representative IS*Hp608*-carrying strains and identification of corresponding "empty site" sequences (lacking IS*Hp608*) in databases of reference strains J99 and 26695. Sequences of each IS*Hp608* end and adjacent DNAs were obtained from three strains (PeCan2A, India49, and India77), and additional useful, although less complete, sequences were obtained in three other cases (strains PeCan16A and copies 1 and 2 from PeCan18B) (detailed in Fig. 2 legend). These *H. pylori* sequence data indicated that IS*Hp608* inserts preferentially with its left end adjacent to 5′-TTAC in target sequences, that it has little or no specificity for sequences next to its right end, and that it does not duplicate target DNA. This specificity is reminiscent of that seen with IS*605* and IS*606*, which insert preferentially downstream of 5′-TTTAA and 5′-TTTAT, respectively (18). The organization of sequences within the termini of IS*Hp608* will be presented in the context of sequence conservation and diversity among IS*Hp608* elements below.

One of two elements from PeCan2A was sequenced next, in part to prepare for functional tests (below). The entire element was PCR amplified with primers specific for flanking sequences (in jhp0924, primers 1 and 20; Fig. 1B) under high-fidelity conditions (Expand High Fidelity polymerase; Boehringer-Mannheim), cloned in a pBS plasmid vector, and then sequenced. This element was 1.83 kb long with only 92% identity (DNA level) to that from strain PeCan18B. However, it also contained two ORFs that were 97% (*orfA*) and 91% (*orfB*) identical in amino acid sequence to those of the first isolate (PeCan18B) and that also overlapped by 10 codons.

Eight additional elements were then characterized, primarily by genomic DNA sequencing without PCR or cloning, to better understand phylogenetic relationships among IS*Hp608* isolates. In overview, each element was closely matched to the others in size and ORF content, and 7 of 11 elements sequenced belonged to two major types (type I from strains PeCan2, India49, SpainB53, and Lith5, and type 2 from PeSJM92, AlaskaP20, and AlaskaP76). These two types differed by about 11% in DNA sequence, in contrast to ~1 to 2% divergence among isolates of a given type, as diagrammed in Fig. 3A. At the protein level, identities and similarities between types 1 and 2 were 95 and 98%, respectively, in *orfA*, and 88 and 92%, respectively, in *orfB*. The other four elements (from strains PeCan4A, India77, and the two elements from PeCan18B) were inferred to be recombinant (Fig. 3B). In particular, the element in strain PeCan4A consisted primarily of type 2 sequence but with replacement of an internal 600-bp segment with type 1 sequence. The two copies of IS*Hp608* from strain PeCan18B seemed to contain seven or eight interspersed patches of type 1 and type 2 sequences, suggesting a history of numerous homologous recombination events. The two copies seemed to be identical (as noted above), except for about 160 bp near their left ends. The element from strain India77 consisted primarily of type 1 sequence but also contained an internal 400-bp segment that was only 72 and 75% matched to corresponding sequences in types 1 and 2 elements, respectively. The ≥25% DNA divergence from other IS*Hp608* sequences suggested that this 400-bp patch was derived from a third type of element.

A well-conserved DDE(K) motif, reminiscent of those found to be important in active sites in several families of transposases (21), was found in *orfB* of each IS*Hp608* element sequenced (identified with an asterisk in Fig. 4), in the *orfB*s of the three other *H. pylori* ISs, and indeed in each of the 26 significant *orfB* homologs from other organisms that were found in the public database (Fig. 4). Most of these DDE motifs are unusual, however, in the relatively short distance between D_2 and E (just 18 amino acids, rather than 33 to more than 100 in most other systems analyzed to date) (14, 21).

Sequences near the ends of ISs generally constitute sites of

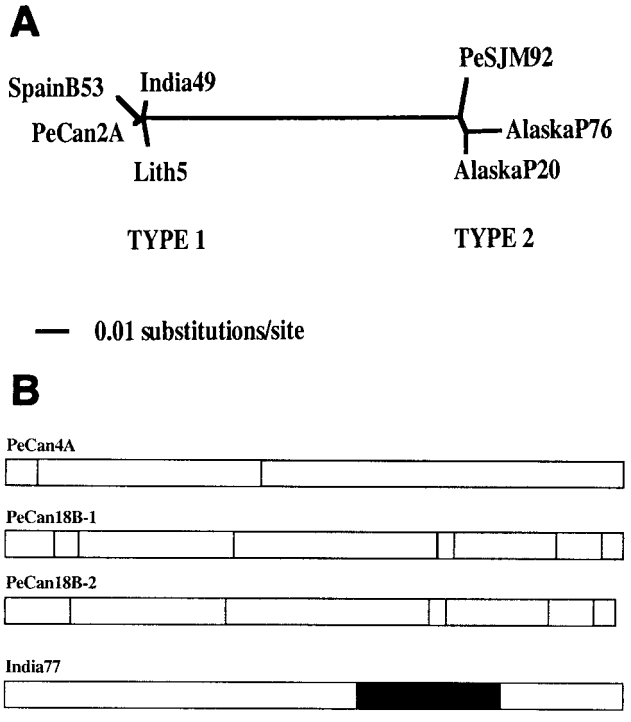

FIG. 3. DNA sequence diversity of IS*Hp608* elements. (A) Maximum-likelihood phylogenetic tree generated using DNA sequences of type 1 and type 2 elements. DNA sequence divergence between types was about 11%; divergence within a given type was about 1 to 2%. The four IS*Hp608*s that seemed to have resulted from recombination between elements of different types were not included because this tree analysis assumes divergence by mutation alone, without recombination between divergent lineages. (B) Recombinant IS*Hp608* elements identified by DNA sequencing. Areas of type 1, type 2, and type 3 sequence are depicted as white, grey, and black boxes, respectively. Strain PeCan4A contains just one copy of IS*Hp608*, which was inferred to be recombinant, based on the following DNA sequence matches: bp 1 to 114, 93% identity with type 2 but 80% with type 1; bp 123 to 715, 99% identity with type 1 but 93% with type 2; bp 732 to 1832, 98% identity with type 2 but 87% with type 1. Strain PeCan18B contained two copies of IS*Hp608*, designated PeCan18B-1 and PeCan18B-2 (for which we do not have the leftmost approximately 34 bp of sequence, as detailed in the Fig. 2 legend). The only difference detected between the two copies was in the leftmost 160 bp, in which copy 1 matched type 1, and the available 130 bp of copy 2 matched type 2 (97% identity with type 2 but 78% with type 1). In other regions the two copies seemed to be identical, although a mosaic of type 1 and type 2 sequences. DNA sequence matches for PeCan18B-2 were as follows: bp 1 to 160, 97% type 1 and 84% type 2; bp 160 to 220, 97% type 2 and 89% type 1; bp 220 to 680, 99% type 1 and 94% type 2; bp 680 to 1270, 97% type 2 and 89% type 1; bp 1270 to 1310, 100% type 1 and 88% type 2; bp 1310 to 1630, 97% type 2 and 87% type 1; bp 1630 to 1750, 97% type 1 and 82% type 2; bp 1750 to 1833, 98% type 2 and 87% type 1. Most of the sequencing of these two copies was done by primer walking on chromosomal DNA, but with the right end of copy 1 also determined by the adapter PCR method (which showed that this copy was inserted into gene jhp0942). The nearly complete copy 2 left-end sequence was obtained from subtractive clone D7 (lacks 30 to 34 bp to the very end); the left end of copy 1 was from sequencing an adapter PCR product. The genomic DNA sequence profiles were carefully searched for mixed traces that would indicate sequence differences between the copies, and none was found except at the left end, as indicated in the figure. Strain India77 contained just one copy of IS*Hp608*, which was inferred to be recombinant based on the following DNA sequence matches: bp 1 to 1021 and also bp 1445 to 1833, 97% identity with type 1 in each segment; bp 1029 to 1419, 75% identity with type 1 and 72% with type 2. This divergent 400 bp was interpreted to indicate the existence of a third type (type 3) of IS*Hp608* element.

transposition protein action. These sites may contain or be included in simple inverted repeats of about 10 to 25 bp in some elements but be more complex in structure (e.g., direct and interrupted inverted repeats) in other elements, depending on the binding specificity of cognate transposition proteins (8, 21, 28). Type 1 and type 2 IS*Hp608* elements each contained (i) a well-conserved direct repeat of 22 bp in the left and right ends; (ii) an octanucleotide (5′AACGCCTT) or penta- or hexanucleotide (5′AACGC or 5′AACGCC) direct repeats in the left ends of type 1 and type 2 elements, respectively; and (iii) conserved decamer (5′-GCTTTAGCTA) or dodecamer (5′-TAGCTTTAGCTA) direct repeats in the right ends of type 1 and type 2 elements, respectively (boxes and horizontal lines in Fig. 5A). These various direct repeat motifs were embedded in a longer and much-interrupted (~37 of 54 bp) inverted repeat (Fig. 5B). Given the apparently strict specificity for insertion of the IS*Hp608* left end downstream of 5′-TTAC target sequences, it is noteworthy that this left end varied in length and sequence: 5′-CC$(A)_4$CT or 5′-C$(A)_5$CT in type 1 elements and 5′-C(A/T)CC(C/T) in type 2 elements, whereas the right end, which seemed to join to target sequences less specifically, was constant in length in the elements analyzed (Fig. 2 and 5A).

IS*Hp608* geographic distribution. The frequency of IS*Hp608* carriage in various *H. pylori* populations was estimated by dot blot hybridization with 80 strains and PCR with 723 strains. Hybridization was done with a probe containing parts of *orfA* and *orfB* (460 bp, generated with primers 5 and 12; Fig. 1B), and PCR was carried out with primers that generated a ~296-bp product from both type I and type 2 IS*Hp608*s (primers 4 and 12; Fig. 1B). Hybridization and PCR results were complementary and entirely consistent; the same 22 strains were scored as positive for IS*Hp608* in both tests, and no strains were scored as positive in one test but not the other. In particular, hybridization test results with DNAs from 24 Japanese strains supported the conclusion from PCR (below) that IS*Hp608* is rare or absent in East Asian strains.

IS*Hp608* was found in some 15 to 20% of strains from Europeans, Africans, Indians, and Bangladeshis, at somewhat higher frequency (41%) in strains from native peoples of the Americas (Peru and Alaska), but in only 1% of East Asian (Japanese, Korean, and Chinese) strains (Table 2). IS*Hp608* also seemed to be more abundant in our relatively few strains from Peruvian patients with gastric cancer than from those with gastritis only (~64% versus ~33%, respectively; $P = 0.04$, Fischer's exact probability). No other correlation between IS*Hp608* carriage and overt disease was observed among cases for which clinical data were available.

All strains found to carry IS*Hp608* were screened by PCR with primers specific for type 2 elements (primers 6 and 10; Fig. 1B). These tests indicated that 8 of 24 native Peruvian and 14 of 32 Alaska native IS*Hp608* elements were type 2 (or contained type 2 sequences if recombinant), whereas none of the 53 elements from European, Indian subcontinent, or African strains seemed to be (Table 2).

The existence of a probable third type of IS*Hp608* element with ≥25% DNA divergence from all others was suggested by the sequence of a 400-bp segment in the recombinant element from strain India77 (noted above). No other elements with this type 3 sequence were found by PCR using type 3-specific

```
                 *                                                                  *                  *    *
ISHp608   CDKSVGFDFGLRTFLTGSDNTK1IESPLFFSQYLPLIKRLSKNLSKKVK-GSNNFKKAKKKLAQLHQKIKHLRTDFFYKLALKLSRKYQSIFIEDLNMKAMQK
IS607     SHNKLGIDIGIKSFVSLSNGLN-IYAPKPLDKLTRKLVRISRQLSKKIH2KSNNYLKHSKKLTHLHEKIANIRLDFLHKLTSSLIRHSNSFCLESLKVKNMFK
IS606     PKNGVGLDLNILDIACSCGVNN-YSKLYSLKKYSKKFKRLQRKQSRRVL3LGGNFYKTQKKLNQAFDKSSHQKTDRYHKITSELSKQFELVVVEDLQVKNMTK
IS605     IKKAVGLDMGLRTLIVTSDKIE-YPHIRFYQKLEKKLTKAQRRLSKKVK-GSNNRKKQAKKVARLHLACSNTRDDYLHKISNEITNQYDLIGVETLNVKGLMR
B1432     SASMVGLDAGVAKLATLSDGTV-FEPVNSFQKNQKKLARLQRQLSRKVK-FSNNWQKQKRKIQRLHSCIANIRRDYLHKVTTAVSKNHAMIVIEDLKVSNMSK
GipA      --SMVGIDVGLKDLFVTDTGFR-SGNPRHTAKYAARLALLQRRLSKKAK-GSKNRAKAHLKVARLHAKIADCRLDALHKATRKLINDNQVVCVESLKVRNMIR

Cons      ----VGLDIGLKTLITTSDG---Y--PR-F-KY-KKL-RLQRRLSKKVK-GSNNF-K--KKLARLH-KIANIR-DFLHKITS-LSKQYQLI-VEDLKVKNM-K

Inserts
1  (in ISHp608)     HKKLTDFKQYPTDMKELLGIEIDEELDTKRLIPT
2  (in IS606)       PKTKGDKTR
3  (in IS607)       K-SKQNKTK
```

FIG. 4. Multiple alignments of amino acid sequences in region of putative DDE(K) motifs in OrfB and homologs. Asterisks identify residues of the DDE(K) motif that, by extrapolation from other systems, could be critical features of an active site (namely, coordination of divalent cations and nucleophilic attack leading to DNA cleavage early in transposition [14, 21]). The consensus sequence (Cons) was developed by multiple alignment in ClustalW in the NTI vector program. Underlining identifies residues also conserved in the consensus based on 26 sequences of OrfB homologs found in GenBank as of September 2001. GenBank accession nos. used in developing this alignment are as follows: AC000108, IS*605*; U95957, IS*606*; AF189015, IS*607*; AF357223, IS*Hp608* type 1; AF411944, IS*Hp608* type 2; C64895, *E. coli* B1432; AAF98319, GipA. The other OrfB homolog accession nos. used here are AC000108, E64644, U95957, A33489, AAC97568, CAC03683, G75376, CAB41498, BAA15060, C64895, CAA60219, F70884, E70811, AAB12365, AAF05601, AF357223, JC4292, AAK40078, AAA26505, AAF98319, T36649, AAA83564, S74909, F75401, C72305, and A82794.

primers (7 and 14), either among the 123 strains that had been scored as containing IS*Hp608* by PCR with general primers (4 and 12) or among any of 212 strains tested that had been scored as lacking IS*Hp608* by PCR with general primers (103 Indian, 48 Bangladeshi, 36 Lithuanian, and 25 Alaska native) (Table 2). Thus, the putative type 3 elements may be rare in *H. pylori*, at least in the regions that we have studied to date.

IS*Hp608* transposition in *E. coli* and mutations affecting transposition. The IS*Hp608* element from strain PeCan2A that had been cloned and sequenced (above) was used to test for transposition in *E. coli* and to identify IS genes and sites needed for this process. The cloned IS*Hp608* element was marked genetically using a *cam* (resistance) gene; plasmids containing these IS*Hp608*:*cam* elements were introduced by transformation into *E. coli* strain DB1683, which contains pOX38, a conjugative (F factor) plasmid. Cultures from single transformant colonies were mated with the Str^r *E. coli* recipient strain MC4100, and exconjugants were selected on medium containing Cam and Str. To help guard against false leads caused by mutation or transposition jackpots during clonal growth, each assay was carried out with four or five separate transformants and with at least two single colonies from each transformant clone.

IS*Hp608* derivatives containing the *cam* cassette in place of most of *orfB* (construct 3) or in a noncoding sequence 68 bp from the right end (construct 4) yielded Cam^r Str^r exconjugants at average frequencies of 5×10^{-8} to 10×10^{-8} per starting donor cell, and nearly all these exconjugants (179 of 180 and 169 of 177, respectively) were Amp^s. This result indicated that IS*Hp608* could transpose in *E. coli* and that most transposition involved simple separation of the marked IS*Hp608* element from vector sequences and insertion into pOX38, not cointegrate formation, as is common with the two other members of this family studied to date (IS*605* and IS*607*) (18, 19).

In contrast, IS*Hp608* elements containing *cam* in place of *orfA* (construct 2) or in a noncoding sequence near the right end (within the 22-bp direct repeat; construct 5) (Fig. 1B and 5A) yielded Cam^r Str^r exconjugants only at frequencies of about 10^{-9}. These yields were at background level and 50- to 100-fold lower than with the other transposition-proficient constructs. Each of 10 exconjugants obtained from these matings that we tested was Amp^r, as expected of replicon fusions generated by illegitimate recombination. Thus, *orfA* is needed for IS*Hp608* transposition in *E. coli*, whereas *orfB* is not. The data also indicated that a sequence near the IS*Hp608* right end is needed for transposition, probably as a site on which transposition proteins act.

The ends and insertion specificity of IS*Hp608* after transposition in *E. coli* were defined by analysis of 13 representative transposition products: 5 from the element marked with *cam* 68 bp from the right end (construct 4; *orfA* and *orfB* each intact) and 8 from the element in which most of *orfB* had been replaced with *cam* (construct 3). Sequencing was carried out directly on *E. coli* genomic DNA, using outward-facing IS*Hp608*-specific primers 2 and 18. Figure 6 shows that in all 13 cases, each end of IS*Hp608* matched that expected from analysis of the ancestral *H. pylori* strain PeCan2A. The left end was immediately downstream of a 5′-TTAC sequence, and the right end was joined to many different sequences, although CG or GC base pairs were preferred in the first two positions (CG or GC in 12 of 13 first positions and also in 12 of 13 second positions).

DISCUSSION

Prokaryotic transposable elements can affect microbial evolution, serve as sensitive indicators of population genetic structure and evolutionary change, act as "selfish" DNAs that increase in copy number by transposition, and exhibit diversity in detailed mechanisms of transposition. The IS*Hp608* element described here represents a widespread but as yet relatively unstudied family whose members contain two genes (*orfA* and *orfB*) that are each homologs of the single putative transposase genes of simpler (one-gene) ISs. We identified two main IS*Hp608* sequence types and found that IS*Hp608* was nonrandomly distributed in *H. pylori* populations, that it formed sim-

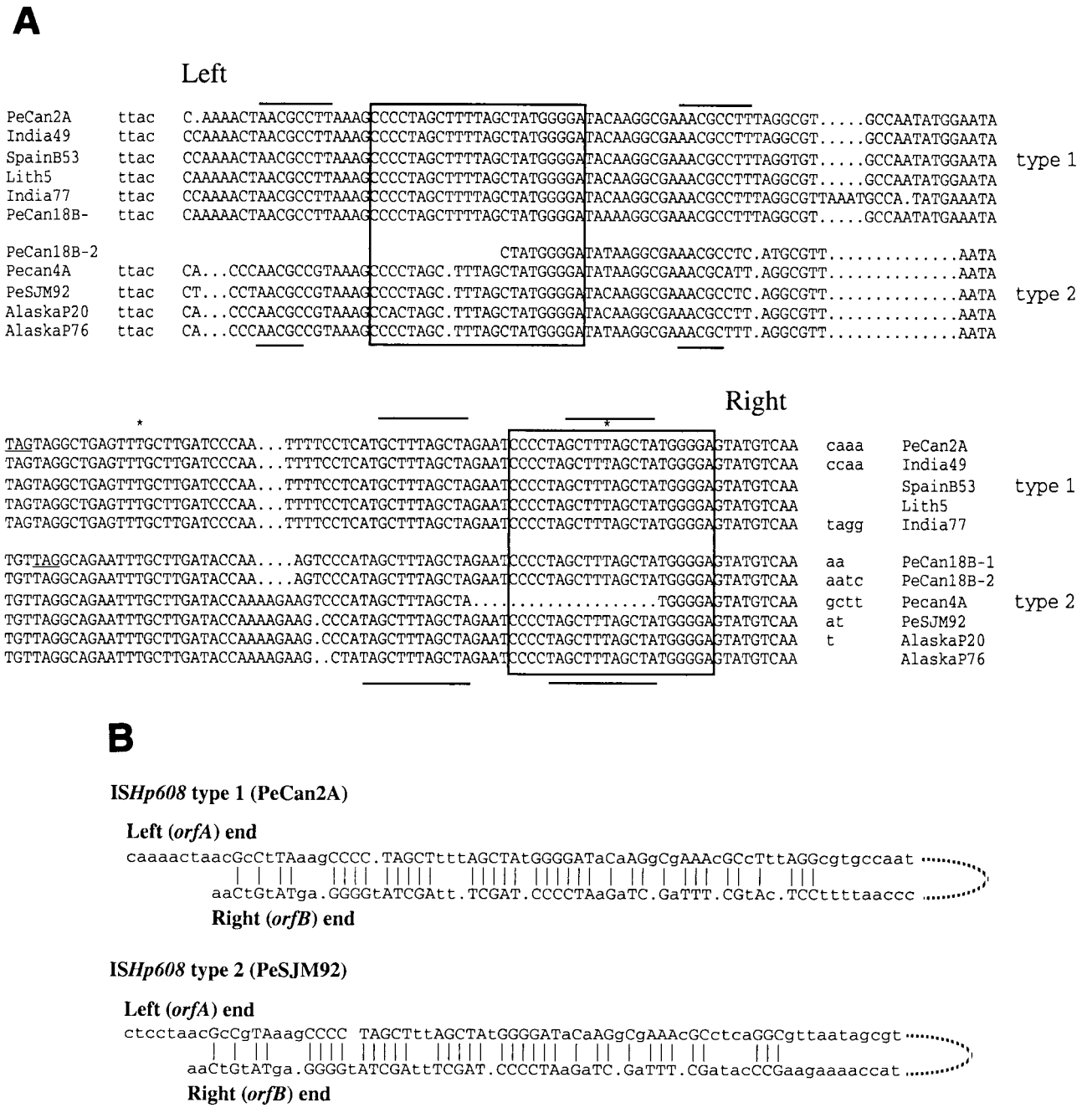

FIG. 5. Repeat sequences in the termini of IS*Hp608* elements. (A) Size and spacing of direct repeats of type 1 and type 2 element termini. Boxed sequences represent a well-conserved 22 (or 23)-bp direct repeat present in both IS*Hp608* ends. Horizontal lines identify shorter direct repeats: octanucleotide (5′AACGCCTT) and related penta- or hexanucleotides (5′AACGC and 5′AACGCC) in the left ends of type 1 and type 2 elements, respectively; and decamer (5′-GCTTTAGCTA) or dodecamer (5′-TAGCTTTAGCTA) in the right ends of type 1 and type 2 elements, respectively. *TAG* is the stop codon at the 3′ end of *orfB*. ∗ identifies the site of insertion of the *cam* cassette in IS*Hp608* right end (insertions 4 and 5; see Fig. 1B). Lowercase letters represent flanking sequences. (B) Subterminal imperfect inverted repeats of IS*Hp608*. Left and right ends of IS*Hp608* (same strand) are aligned to highlight inverted repeat sequence (capital letters).

ple insertions with limited target site specificity, and that it required *orfA* but not *orfB* for transposition in *E. coli*.

Population genetics and evolution. The observed *orfA-orfB* arrangements in the four known ISs of *H. pylori* (Fig. 1A) and the need for *orfA* but not *orfB* for transposition in *E. coli* (19; this study) contributes to interest in the evolutionary history of this mobile DNA family. The two unrelated *orfA*s (e.g., in IS*Hp608* and IS*607*) and related *orfA*s in two orientations (e.g., in IS*Hp608* and IS*605*) suggest that these elements arose several times, perhaps by transposition of simpler elements or specialized or illegitimate recombination. Explanations for the conservation of a particular *orfA-orfB* association in each IS species include (i) both genes being important for transposition in *H. pylori*; (ii) a need for just one gene for transposition,

TABLE 2. Distribution of IS*Hp608* in different *H. pylori* populations

Region	No. of strains	No. (%) of IS*Hp608*-positive strains			
		Total	Type 1	Type 2	Other[a]
Europe					
Spain	51	9 (18)	9 (100)	0 (0)	0 (0)
England	48	8 (17)	8 (100)	0 (0)	0 (0)
Lithuania	72	20 (28)	20 (100)	0 (0)	0 (0)
Africa	23	4 (17)	4 (100)	0 (0)	0 (0)
East Asia					
Japan	84	0 (0)			
Korea	63	0 (0)			
China	76	2[b] (2.6)	2 (100)	0 (0)	0 (0)
South Asia					
India	118	15 (13)	14 (93)	0 (0)	1 (7)
Bangladesh	57	9 (16)	9 (100)	0 (0)	0 (0)
The Americas (natives)					
Peru					
Gastritis	45	15 (33)	9 (60)	6 (40)	0 (0)
Cancer	14	9 (64)	7 (78)	2 (22)	0 (0)
Alaska	72	32 (44)	18 (56)	14 (44)	0 (0)

[a] The one other IS*Hp608* element in Indian strains was identified by sequencing as a type 1-type 3 recombinant (Fig. 3B).
[b] These two strains were from Hong Kong.

with the other contributing to fitness; and (iii) carriage of both genes together reflecting evolutionary forces in earlier microbial hosts, but not important in *H. pylori*.

IS*Hp608* was found in some 15 to 60% of *H. pylori* strains from many parts of the world, but in only 1% of strains from East Asia. These results are reminiscent of findings that East Asian and other strains are generally distinguishable based on alleles of the *vacA* and *cagA* virulence genes and sets of insertion and deletion motifs downstream of *cagA* (17, 20, 26, 37, 39). Collectively these findings indicated that the gene pools of East Asian and Western *H. pylori* strains are distinct and that *H. pylori* is not truly a "panmictic" species, as had been postulated (33). Our IS*Hp608* data suggest that IS*Hp608* was transferred into *H. pylori* after the East Asian and Western *H. pylori* populations were separated. IS*Hp608* would then have spread by DNA transfer between strains in a population and by recombination and transposition. This spread would have been accelerated if IS*Hp608* contributed to bacterial fitness. The paucity of IS*Hp608* in East Asian *H. pylori* strains might then reflect a relative lack of *H. pylori* gene flow (due to the thousands of years in which there was rather little human migration) between East Asia and many other parts of the world.

Our sequence and PCR analyses indicated that one-third or more of the IS*Hp608*s in Alaskan and Peruvian strains were distinct (type 2), differing by some 11% in DNA sequence from most other IS*Hp608* elements (Table 2 and Fig. 3A). The abundance of type 2 elements in Alaskan strains is difficult to interpret because there is little other information at present on Alaska native *H. pylori* genotypes. The abundance of type 2 elements in Peru was surprising, however, because the Peruvian strains analyzed here seemed to be closely related to Spanish strains in tests of several diagnostic markers (20). Nevertheless, no type 2 elements were found among nine IS*Hp608* elements from Spain or among the 28 other IS*Hp608*s from elsewhere in Europe (United Kingdom and Lithuania).

Many explanations seem tenable, given our present limited understanding of *H. pylori* genome evolution, and two will be considered here. In one explanation, the ancestors of modern native Americans might have brought *H. pylori* carrying type 2 IS*Hp608* with them from central and northern Asia thousands of years ago. If those putative original *H. pylori* strains differed genetically from the Spanish-type strains that now predominate, it would imply that type 2 IS*Hp608*s had been transferred into the new Spanish-type strains before they had completely displaced the original (Asian-type) Peruvian strains. This explanation would also imply, however, that the original putative Asian-Peruvian *H. pylori* strains had often carried type 2 IS*Hp608* elements, even though IS*Hp608* is very rare now in East Asia (the two East Asian exceptions, being from Hong Kong, might actually be Western in origin). Alternatively, *H. pylori* might have been brought to the Americas by early European conquerors (20). In this case, type 2 IS*Hp608* elements might have been uncommon in the original European *H. pylori* populations but then became abundant in Peru by chance (founder effect) and/or due to special conditions accompanying the spread of *H. pylori* in the New World. For example, rapid person-to-person transmission in the large, *H. pylori*-free, highly susceptible native population might have created con-

```
                          Left                                  Right
2-1   gctggacacacacattac CAAAACTAACGCCTTAAA..GCTAGGGGAGTATGTCAA cgtaacgccaggacacca
5-2   aacaatcacacagattac CAAAACTAACGCCTTAAA..GCTAGGGGAGTATGTCAA ccgtaaacagcctgaatg
6-1   gacgtaacatatgtttac CAAAACTAACGCCTTAAA..GCTAGGGGAGTATGTCAA tgtattcgcgacgtccca
2-2   aggaggctgcgcagttac CAAAACTAACGCCTTAAA..GCTAGGGGAGTATGTCAA cgggtatatataacatgg
1-1   tgccgtggtttcttttac CAAAACTAACGCCTTAAA..GCTAGGGGAGTATGTCAA cgtatcccacgaccagat
5-1   gtgctgggtgatatttac CAAAACTAACGCCTTAAA..GCTAGGGGAGTATGTCAA ctgaatcataaatacacg
6-2   tgccgtaggagtaattac CAAAACTAACGCCTTAAA..GCTAGGGGAGTATGTCAA cgtcaaccacactgcctg
1-2   caggcgaatgctgtttac CAAAACTAACGCCTTAAA..GCTAGGGGAGTATGTCAA ggggtttttactggttg
18-1  aggaggctgcgcagttac CAAAACTAACGCCTTAAA..GCTAGGGGAGTATGTCAA cgggtatatataagatgg
18-2  ccgccatccgtcctttac CAAAACTAACGCCTTAAA..GCTAGGGGAGTATGTCAA ggtggtttctgagcagat
23-1  ccgccatccgtcctttac CAAAACTAACGCCTTAAA..GCTAGGGGAGTATGTCAA ggtggtttctgagcagat
23-2  gcctgacggtcctttac CAAAACTAACGCCTTAAA..GCTAGGGGAGTATGTCAA ccgcaaacatgccgaagc
24-1  acagcgtaaggtgtttac CAAAACTAACGCCTTAAA..GCTAGGGGAGTATGTCAA ccgtaatgagggaaaaca
```

FIG. 6. IS*Hp608* insertions in F factor in *E. coli.* Termini and sites of IS*Hp608cam* insertions in F factor DNA (GenBank accession no. NC002483). Left and right termini of IS*Hp608*s are shown in uppercase and flanking DNAs are shown in lowercase letters. The products of transposition from donor plasmid 3 (A^+B^-) (eight sites of insertion) were as follows: 2–1, position 60704/5; 5–2, position 62725/6; 6–1, position 64675/6; 2–2, position 65422/3; 1–1, position 80949/50; 5–1, position 85934/5; 6–2, position 86796/7; and 1–2, position 99/100. The products of transposition from donor plasmid 4 (A^+B^+) (five sites of insertion) were: 18–1, position 65422/3; 18–2 and 23–1 each at (although independent), position 95172/3; 23–2, position 59288/9; and 24–1, position 70860/1.

ditions that also fostered interstrain gene transfer and/or transposition of type 2 IS*Hp608* elements.

The *orfB-gipA* (*Salmonella* virulence gene) homology (32) and contributions of the unrelated Tn*5* tranposase to *E. coli* fitness (16) serve as reminders that the frequency of an IS*Hp608* type would be further increased if it contributed to the efficiency of *H. pylori* transmission to new hosts in any way. Direct tests for effects of IS*Hp608* elements on *H. pylori* fitness in culture and in in vivo models and studies of strains from other populations may help resolve these issues.

Transposition mechanism. The transposition of IS*Hp608* in *E. coli* was *orfA* dependent but *orfB* independent, as was also the case with IS*607*, a family member related to IS*Hp608* only in *orfB* (protein level) (19). Essentially all IS*Hp608* transposition products were simple insertions, not cointegrates (which contain both plasmid vector and IS*Hp608* DNAs), whereas most IS*605* and IS*607* transposition products were cointegrates (18, 19). Where tested (IS*Hp608* and IS*607*), these outcomes were not affected by *orfB* inactivation. Simple insertions and cointegrates often result from double- versus single-strand DNA breaks by transposase, resulting in conservative (break-join) and replicative transposition, respectively (6, 28, 30, 36). Studies with Tn*7* and IS*903* have shown that elements can be switched from conservative to replicative modes by subtle mutations affecting transposase activity or accessibility of cognate DNA sites (23, 34). Equivalent differences might account for the different spectra of transposition products generated with IS*Hp608* versus IS*605* and IS*607*.

Inspection of the sequences flanking IS*Hp608* in *H. pylori* and after transposition in *E. coli* indicated that (i) IS*Hp608* inserts with its left (*orfA*) end immediately downstream of 5′-TTAC; (ii) its right end joins target DNAs nonspecifically; and (iii) it neither duplicates nor deletes sequences at its site of insertion. Similar patterns were evident with IS*605* and IS*606*, except that they inserted next to 5′-TTTAA and 5′-TTTAT, respectively (18). In contrast, IS*607*, whose OrfA is unrelated, inserted preferentially between adjacent G nucleotides in target DNA (19). These patterns suggest that OrfA proteins help select DNA targets, a property exhibited by many but not all transposase proteins (13).

The 22-bp direct repeats embedded within the imperfect inverted repeats near each IS*Hp608* end (Fig. 5) may constitute sites of transposase action. In accord with this view, transposition was blocked by placing a *cam* cassette within one repeat unit near the right end but not by placing it interior to this repeat (68 bp from the end).

One model to explain transposition in *E. coli* of *orfB* mutant derivatives of IS*Hp608* and related elements holds that their movement is truly independent of any OrfB-like protein. Alternatively, an OrfB-like protein might be needed but be satisfied in *E. coli* by the chromosomally encoded homolog (B1432; 25% identity and 42% similarity to IS*Hp608* OrfB, accession no. C64895). In accord with this view are (i) the strong *orfA-orfB* associations discussed above (18, 19; this study), (ii) the frequent annotation of OrfB homologs as transposases, and (iii) conserved DDE(K) motifs (Fig. 4), reminiscent of those in active sites of transposases of several families.

Different distances can separate D_1, D_2, and E residues in the linear transposase sequence. In the Tn*5* transposase, whose D_2-E distance is unusually large (138 amino acids), the intervening peptide contributes importantly to transposase-DNA binding (14). If the DDE(K) motifs of OrfB homologs also form part of transposase active sites, they would, in contrast, be among the most tightly spaced of any known to date (18 amino acids between D_2 and E in *orfB* versus 33 to more than 100 amino acids in others [21]). Based on the Tn*5* precedent (14), this short spacing might affect IS*Hp608* OrfB protein interaction with DNA substrates.

Continued population-level analyses of IS*Hp608* and related elements and the proteins they encode may help elucidate the origins and modes of evolution of prokaryotic transposable elements, forces that shape their distributions in bacterial populations, and factors important in the evolution of their *H. pylori* hosts as well. Similarly, detailed biochemical analyses, focused on the two families of OrfA proteins, each needed for transposition although lacking obvious DDE(K) motifs, and also on the universally present OrfB proteins with their tightly spaced DDE(K) motifs (if OrfB ever participates in IS movement), should provide new insights into the diversity of mechanisms of transposition and DNA rearrangement.

ACKNOWLEDGMENTS

We thank many colleagues for providing biopsy samples and *H. pylori* cultures used in the present experiments, in particular, A. Bussalleu, J. Combe, T. Alarcon, J. C. Atherton, H. Chalkauskas, A. Chowdhury, A. Janulaitis, L. Kiuduliene, H. K. Lee, M. Lopez Brea, G. B. Nair, M. Rahman, B. Wong, I. Segal, and J. E. Thomas. We thank J. Slauch for helpful discussions about *gipA*.

This research was supported in part by U.S. Public Health Service grants AI38166, AI49161, DK53727, and P30 DK52574.

REFERENCES

1. **Achtman, M., T. Azuma, D. E. Berg, Y. Ito, G. Morelli, Z. J. Pan, S. Suerbaum, S. A. Thompson, A. van der Ende, and L. J. van Doorn.** 1999. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. Mol. Microbiol. **32:**459–470.
2. **Akopyants, N. S., A. Fradkov, L. Diatchenko, P. D. Siebert, S. Lukyanov, E. D. Sverdlov, and D. E. Berg.** 1998. PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. Proc. Natl. Acad. Sci. USA **95:**13108–13113.
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. S. 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:**176–180.
5. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A., Smith, and K. A. Struhl.** 1994. Current protocols in molecular biology, supplement 27, page 2.4.1. Greene Publishing and Wiley Interscience, New York, N.Y.
6. **Berg, D. E.** 1983. Structural requirement for IS*50*-mediated gene transposition. Proc. Natl. Acad. Sci. USA **80:**792–796.
7. **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 individual Peruvian patients. Clin. Infect. Dis. **25:**996–1002.
8. **Berg, D. E., and M. M. Howe (ed.).** 1989. Mobile DNA. American Society for Microbiology, Washington, D.C.
9. **Biery, M. C., M. Lopata, and N. L. Craig.** 2000. A minimal system for Tn*7* transposition: the transposon-encoded proteins TnsA and TnsB can execute DNA breakage and joining reactions that generate circularized Tn*7* species. J. Mol. Biol. **297:**25–37.
10. **Blanchard, T. G., and S. J. Czinn.** 2001. *Helicobacter pylori* acquisition and transmission: where does it all begin? Gastroenterology **121:**483–485.
11. **Chalkauskas, H., D. Kersulyte, I. Cepuliene, V. Urbonas, D. Ruzeviciene, A. Barakauskiene, A. Raudonikiene, and D. E. Berg.** 1998. Genotypes of *Helicobacter pylori* in Lithuanian families. Helicobacter **3:**296–302.
12. **Cover, T. L., D. E. Berg, M. J. Blaser, and H. L. T. Mobley.** 2001. *H. pylori* pathogenesis, p. 509–558. *In* E. A. Groisman (ed.), Principles of bacterial pathogenesis. Academic Press, New York, N.Y.

13. **Craig, N. L.** 1997. Target site selection in transposition. Annu. Rev. Biochem. **66:**437–474.
14. **Davies, D. R., I. Y. Goryshin, W. S. Reznikoff, and I. Rayment.** 2000. Three-dimensional structure of the Tn*5* synaptic complex transposition intermediate. Science **289:**77–85.
15. **Dubois, A., A. Welch, D. E. Berg, and M. J. Blaser.** 2000. *Helicobacter pylori*, p. 263–280. *In* J. P. Nataro, M. J. Blaser, and S. Cunningham-Rundles (ed.), Persistent bacterial infections. ASM Press, Washington, D.C.
16. **Hartl, D. L., D. E. Dykhuizen, R. D. Miller, L. Green, and J. de Framond.** 1983. Transposable element IS*50* improves growth rate of *E. coli* cells without transposition. Cell **35:**503–510.
17. **Ito, Y., T. Azuma, S. Ito, H. Suto, H. Miyaji, Y. Yamazaki, Y. Kohli, and M. Kuriyama.** 1998. Full-length sequence analysis of the *vacA* gene from cytotoxic and noncytotoxic *Helicobacter pylori*. J. Infect. Dis. **178:**1391–1398.
18. **Kersulyte, D., N. S. Akopyants, S. W. Clifton, B. A. Roe, and D. E. Berg.** 1998. Novel sequence organization and insertion specificity of IS*605* and IS*606*: chimeric transposable elements of *Helicobacter pylori*. Gene **223:**175–186.
19. **Kersulyte, D., A. K. Mukhopadhyay, M. Shirai, T. Nakazawa, and D. E. Berg.** 2000. Functional organization and insertion specificity of IS*607*, a chimeric element of *Helicobacter pylori*. J. Bacteriol. **182:**5300–5308.
20. **Kersulyte, D., A. K. Mukhopadhyay, B. Velapatino, W. W. Su, Z. J. Pan, C. Garcia, V. Hernandez, Y. Valdez, R. S. Mistry, R. H. Gilman, Y. Yuan, H. Gao, T. Alarcon, M. Lopez-Brea, G. Balakrish Nair, A. Chowdhury, S. Datta, M. Shirai, T. Nakazawa, R. Ally, I. Segal, B. C. Wong, S. K. Lam, F. O. Olfat, T. Boren, L. Engstrand, O. Torres, R. Schneider, J. E. Thomas, S. Czinn, and D. E. Berg.** 2000. Differences in genotypes of *Helicobacter pylori* from different human populations. J. Bacteriol. **182:**3210–3218.
21. **Mahillon, J., and M. Chandler.** 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. **62:**725–774.
22. **Marchuk, D., M. Drumm, A. Saulino, and F. S. Collins.** 1991. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucleic Acids Res. **19:**1154.
23. **May, E. W., and N. L. Craig.** 1996. Switching from cut-and-paste to replicative Tn*7* transposition. Science **272:**401–404.
24. **Mitchell, H. K.** 2001. Epidemiology of infection, p. 7–18. *In* H. L. Mobley, G. L. Mendz, and S. L. Hazell (ed.), *Helicobacter pylori*: physiology and genetics. American Society for Microbiology, Washington, D.C.
25. **Mukhopadhyay, A. K., D. Kersulyte, J. Y. Jeong, S. Datta, Y. Ito, A. Chowdhury, S. Chowdhury, A. Santra, S. K. Bhattacharya, T. Azuma, G. B. Nair, and D. E. Berg.** 2000. Distinctiveness of genotypes of *Helicobacter pylori* in Calcutta, India. J. Bacteriol. **182:**3219–3227.
26. **Pan, Z. J., R. W. van der Hulst, M. Feller, S. D. Xiao, G. N. Tytgat, J. Dankert, and A. van der Ende.** 1997. Equally high prevalences of infection with *cagA*-positive *Helicobacter pylori* in Chinese patients with peptic ulcer disease and those with chronic gastritis-associated dyspepsia. J. Clin. Microbiol. **35:**1344–1347.
27. **Parsonnet, J.** 1999. *Helicobacter* and gastric adenocarcinoma, p. 372–408. *In* J. Parsonnet (ed.), Microbes and malignancy: infection as a cause of human cancers. Oxford University Press, New York, N.Y.
28. **Reznikoff, W. S., A. Bhasin, D. R. Davies, I. Y. Goryshin, L. A. Mahnke, T. Naumann, I. Rayment, M. Steiniger-White, and S. S. Twining.** 1999. Tn*5*: a molecular window on transposition. Biochem. Biophys. Res. Commun. **266:**729–734.
29. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
30. **Shapiro, J. A.** 1979. Molecular model for the transposition and replication of bacteriophage *Mu* and other transposable elements. Proc. Natl. Acad. Sci. USA **76:**1933–1937.
31. **Sharpe, P. L., and N. L. Craig.** 1998. Host proteins can stimulate Tn*7* transposition: a novel role for the ribosomal protein *L29* and the acyl carrier protein. EMBO J. **17:**5822–5831.
32. **Stanley, T. L., C. D. Ellermeier, and J. M. Slauch.** 2000. Tissue-specific gene expression identifies a gene in the lysogenic phage Gifsy-1 that affects *Salmonella enterica* serovar Typhimurium survival in Peyer's patches. J. Bacteriol. **182:**4406–4413.
33. **Suerbaum, S.** 2000. Genetic variability within *Helicobacter pylori*. Int. J. Med. Microbiol. **290:**175–181.
34. **Tavakoli, N. P., and K. M. Derbyshire.** 2001. Tipping the balance between replicative and simple transposition. EMBO J. **20:**2923–2930.
35. **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.
36. **Turlan, C., and M. Chandler.** 2000. Playing second fiddle: second-strand processing and liberation of transposable elements from donor DNA. Trends Microbiol. **8:**268–274.
37. **van der Ende, A., Z. J. Pan, A. Bart, R. W. van der Hulst, M. Feller, S. D. Xiao, G. N. Tytgat, and J. Dankert.** 1998. *cagA*-positive *Helicobacter pylori* populations in China and The Netherlands are distinct. Infect. Immun. **66:**1822–1826.
38. **Westblom, T. U., S. J. Czinn, and J. G. Nedrud (ed.).** 1999. Current topics in microbiology and immunology, vol. 241. Springer-Verlag GmbH & Co., KG, Berlin, Germany. Gastroduodenal disease and *Helicobacter pylori*: pathophysiology, diagnosis and treatment.
39. **Yamaoka, Y., M. S. Osato, A. R. Sepulveda, O. Gutierrez, N. Figura, J. G. Kim, T. Kodama, K. Kashima, and D. Y. Graham.** 2000. Molecular epidemiology of *Helicobacter pylori*: separation of *H. pylori* from East Asian and non-Asian countries. Epidemiol. Infect. **124:**91–96.