Transposable element ISHp608 of Helicobacter pylori: Nonrandom geographic distribution, functional organization, and insertion specificity

Dangeruta Kersulyte
*Washington University School of Medicine in St. Louis*

Billie Velapatiño
*Washington University School of Medicine in St. Louis*

Giedrius Dailide
*Washington University School of Medicine in St. Louis*

Asish K. Mukhopadhyay
*Washington University School of Medicine in St. Louis*

Yoshiyuki Ito
*Washington University School of Medicine in St. Louis*

See next page for additional authors

Follow this and additional works at: [https://digitalcommons.wustl.edu/open_access_pubs](https://digitalcommons.wustl.edu/open_access_pubs)

**Recommended Citation**


[https://digitalcommons.wustl.edu/open_access_pubs/2809](https://digitalcommons.wustl.edu/open_access_pubs/2809)
Authors
Dangeruta Kersulyte, Billie Velapatiño, Giedrius Dailide, Asish K. Mukhopadhyay, Yoshiyuki Ito, Lizbeth Cahuayme, Alan J. Parkinson, Robert H. Gilman, and Douglas E. Berg
Transposable Element ISHp608 of Helicobacter pylori: Nonrandom Geographic Distribution, Functional Organization, and Insertion Specificity

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

Transposable Element ISHp608 of Helicobacter pylori: Nonrandom Geographic Distribution, Functional Organization, and Insertion Specificity

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

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

Received 3 October 2001/Accepted 19 November 2001

A new member of the IS605 transposable element family, designated ISHp608, 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 ISHp608 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. ISHp608 also seemed more abundant in Peruvian gastric cancer strains than gastritis strains (9 of 14 versus 15 of 45, respectively; \( P = 0.04 \)). Two ISHp608 types differing by \( \sim 11\% \) 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 \( \leq 2\% \) in DNA sequence, but several recombinant elements were also found. ISHp608 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. ISHp608 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 history 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 transposable elements specify just one transposase protein that acts as a multimer on matched (inverted repeat) sequences at each element end. Others, exemplified by Tn7, specify two different proteins that form a more complex transposase. With Tn7, each transposase protein has a distinct role in transposition, and additional Tn7-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 IS605 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 (IS606 and IS607) is related to IS605 by protein level homologies in orfB (some 25 to 35% amino acid sequence identity) (18, 19), as is the new ISHp608 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 IS605 family can be distinguished based on protein level homologies in orfA, one represented by IS607, whose transposition was found to be orfA dependent and orfB independent (19), and the other represented by IS605 and IS606 (18) and also the present ISHp608. The orfAs of IS605, IS606, and ISHp608 are protein level homologs of the putative transposase gene of IS200, an element that is abundant in natural isolates of Salmonella and Escherichia coli. Only a few cases of IS200 transposition have ever been detected, however, and the mechanisms of its movement are not understood (8, 21, 36).

Sequences of IS605, IS606, and IS607 were each found in...
known H. pylori IS is 1,833 bp long, making it slightly smaller than its three other Hp608 cam sequence was replaced by cam end, within a 22-bp direct repeat (putative transposase binding site) Hp608 and 420 codons in IS orfB sizes are 156, 148, 139, and 218 codons and protein level (range of 28 to 36% amino acid sequence identities). represent ORFs: those with the same shading pattern are related at the VOL. 184, 2002 IS cam 606, IS GenBank accession nos. for IS are 171 bp and 75 bp, respectively (accession no. AF357224). Hp608 arrangement in IS Hp608 #2 (A cam primers, the subtractive clone that led to the discovery of IS U95957, and AF189015, respectively. (B) Positions of oligonucleotide Hp608 fi primers speci, respectively, in strain PeCan2A. ct o ers are described in detail in Table 1. Primers 1 and 20 are speci, respectively, in PCR and sequencing primers are denoted by arrowheads; these prim- mutations generated in IS Hp608 PeCan18B that begins approximately 30 to 34 bp from the left end of (Results). Subtractive clone D7 is a 242-bp Alu (resistance gene) cassette inserted into (allele) or replac- #5) or replac- #4, and IS are AC000108, U95957, and AF189015, respectively. (B) Positions of oligonucleotide primers, the subtractive clone that led to the discovery of ISHp608, and mutations generated in ISHp608. The locations and orientations of PCR and sequencing primers are denoted by arrowheads; these prim- es 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 ISHp608, respectively, in strain PeCan2A. * and ** identify primers specific for ISHp608 types 2 and 3, respectively (defined in Results). Subtractive clone D7 is a 242-bp Alu fragment from strain PeCan18B that begins approximately 30 to 34 bp from the left end of IShp608 copy 2, cam#2, cam#3, cam#4, and cam#5 represent mut- tant derivatives of ISHp608 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 IShp608 sequences. In cam#4 (A’ B’), cam was inserted just downstream of orfB and 68 bp from the rightmost end of ISHp608 (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 ISHp608 right end, within a 22-bp direct repeat (putative transposase binding site) (generated with primers 9 and 19), as detailed in Results.

FIG. 1. IS element maps. (A) Structures of ISs of H. pylori. ISHp608 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 ISHp608, IS605, IS606, and IS607, respectively, orfA and orfB overlap by 10 codons in ISHp608, which is reminiscent of the arrangement in IS607 (nine-codon overlap). Distances from left and right ends to the 5’ end of orfA and 3’ end of orfB in a representative ISHp608 are 171 bp and 75 bp, respectively (accession no. AF357224). GenBank accession nos. for IS605, IS606, and IS607 are AC000108, U95957, and AF189015, respectively. (B) Positions of oligonucleotide primers, the subtractive clone that led to the discovery of ISHp608, and mutations generated in ISHp608. The locations and orientations of PCR and sequencing primers are denoted by arrowheads; these prim- es 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 ISHp608, respectively, in strain PeCan2A. * and ** identify primers specific for ISHp608 types 2 and 3, respectively (defined in Results). Subtractive clone D7 is a 242-bp Alu fragment from strain PeCan18B that begins approximately 30 to 34 bp from the left end of IShp608 copy 2, cam#2, cam#3, cam#4, and cam#5 represent mut- tant derivatives of ISHp608 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 IShp608 sequences. In cam#4 (A’ B’), cam was inserted just downstream of orfB and 68 bp from the rightmost end of ISHp608 (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 ISHp608 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 ISHp608, 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
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 subclone of the bacterial 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-Manheim, Indianapolis, Ind.), 5 μl 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 EcoRV-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 the E. coli K-12 strains used in this study were DH5α, the routine host for H. pylori strain PeCan2A (GenBank accession no. AF537223). H. pylori strains for which records were available were from gastric cancer patients, and the other 12 were from patients with gastritis only. Sixty-three of the 45 Peruvian gastritis 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). 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 PeCan) were cultured from biopsy samples provided by Alejandro Busalque and Juan Combe of the Servicio Universitario de Apoyo, Universidad Peruana Cayetano Heredia (UPCH) in Lima. These and all other biopsy samples were from patients 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).

**TABLE 1. Primers used in this study**

<table>
<thead>
<tr>
<th>Type and primer no.</th>
<th>Primer name</th>
<th>Sequence</th>
<th>Locationa</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>jhp0924R</td>
<td>5'-TGCCTGATGCGAAAATGCACGTCA 22 bp from right end of ISHp608</td>
<td>Left of ISHp608 in PeCan2A</td>
</tr>
<tr>
<td>20</td>
<td>jhp0924F</td>
<td>5'-TTATTAAAGTCTATTGCAGAACATT 140 bp from right end of ISHp608</td>
<td>Right of ISHp608 in PeCan2A</td>
</tr>
<tr>
<td>2</td>
<td>608-3a</td>
<td>5'-GATAGTTGCTTTGAGAATATGAC 233 bp from left end of ISHp608</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>608-I</td>
<td>5'-ACTAATCCCTGACCCGTGA 208 bp from left end of ISHp608</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>608-EF1</td>
<td>5'-CCCTACGCTTGCTTGAGTGTGTC 683 bp from left end of ISHp608</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>608-F</td>
<td>5'-ATCTCTTCAGTAAAATCTTGACGTGATTC 851 bp from right end of ISHp608</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>608a112-R+</td>
<td>5'-TTTTCCTCCTAATCTCTTTA 958 bp from left end of ISHp608</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>608a113-R+</td>
<td>5'-TAGGACGGTGTTGAACAGCGC 1,444 bp from left end of ISHp608</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>608neutral-F</td>
<td>5'-CAAGAAAATCGACGCTCTAG 1,772 bp from left end of ISHp608</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>608neutral-R</td>
<td>5'-CTTACCTGATTGGGACAGTCTA 1,815 bp from right end of ISHp608</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>608a112-F+</td>
<td>5'-TATGATGAAAATACACCAACATAA 1,647 bp from right end of ISHp608</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>608-out</td>
<td>5'-GTCTATTTCTTGCAATACACATCCG 1,624 bp from right end of ISHp608</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>608-R</td>
<td>5'-CAACGTTCGGTGTGAACTGTC 1,446 bp from right end of ISHp608</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>608-out</td>
<td>5'-GTAGCCTACAGTACGCTTACGC 921 bp from right end of ISHp608</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>608a113-F+</td>
<td>5'-CGGTTAGGTTGAGTTCCACCAG 620 bp from right end of ISHp608</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>608-out</td>
<td>5'-TAGGGATTTGGTGAAAGCTCTTC 343 bp from right end of ISHp608</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>608-out</td>
<td>5'-AAGAAGGAGAATACCCAGTGA 200 bp from right end of ISHp608</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>608-F</td>
<td>5'-CTTACACATTCAATGGGGGTTG 140 bp from right end of ISHp608</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>608-R</td>
<td>5'-CTCCTCTCAGTCTAAATCTTGCATG 851 bp from left end of ISHp608</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>608neutral-R</td>
<td>5'-TTTACCTGATTGGGACAGTCTA 22 bp from right end of ISHp608</td>
<td></td>
</tr>
</tbody>
</table>

Adaptor-specific primer (TaKaRa)

Adaptor-specific primer (TaKaRa)

**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 Busalque and Juan Combe of the Servicio Universitario de Apoyo, Universidad Peruana Cayetano Heredia (UPCH) in Lima. These and all other biopsy samples were from patients 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).
as the donor in bacterial conjugation; and MC4100 which is F− and Strr and was used as the recipient in conjugation (6). The plasmid pBluescript SK− (Stragen, La Jolla, Calif.) (hereafter called vector A) and a chromophenicol resistance cassette (cam), developed initially by D. E. Taylor as a selectable factor and transfer of pOX38::ISHp608 containing exconjugants which had been generated by transposition to the pOX38 F factor and cam marker in a noncoding sequence 22 bp from the rightmost ISHp608 element from strain PeruSJM92, nts 14 through 1836 of the 1,838 nt sequence in GenBank accession no. AF411943; from strain PeCan18B and primers based on the sequence of 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 IS200 of E. coli and IS1004 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 ISHp608.

RESULTS

ISHp608 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 IS200 of E. coli and IS1004 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 ISHp608.

ISHp608 sequence and diversity. Sequencing of ISHp608 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 ISHp608 that were identical for nearly 1.7 kb and that either (i) the ends of these ISHp608 elements were traversed where sequences became mixed or (ii) the copies of this element differed in sequence at one or both ends.

This ISHp608 sequence contained open reading frames (ORFs) of 156 and 384 codons that were protein level homologs of orfA and orfB, respectively, in IS605 (Fig. 1A). However, the two ORFs in ISHp608 were in the same orientation and overlapped by 10 codons, whereas their homologs in IS605 and IS606 were divergent. Their parallel orientation in ISHp608 is reminiscent of that in IS607, an element with homology to ISHp608 only in orfB, not orfA.

ISHp608 copy numbers in PeCan18B and four other ISHp608-positive strains (identified in preliminary dot blot tests) were estimated in Southern blots of genomic DNAs that had been singly digested by HindIII and SphI and probed

as the donor in bacterial conjugation; and MC4100 which is F− and Strr and was used as the recipient in conjugation (6). The plasmid pBluescript SK− (Stragen, La Jolla, Calif.) (hereafter called vector A) and a chromophenicol resistance cassette (cam), developed initially by D. E. Taylor as a selectable factor and transfer of pOX38::ISHp608 containing exconjugants which had been generated by transposition to the pOX38 F factor and cam marker in a noncoding sequence 22 bp from the rightmost ISHp608 element from strain PeruSJM92, nts 14 through 1836 of the 1,838 nt sequence in GenBank accession no. AF411943; from strain PeCan18B and primers based on the sequence of 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 IS200 of E. coli and IS1004 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 ISHp608.

RESULTS

ISHp608 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 IS200 of E. coli and IS1004 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 ISHp608.

ISHp608 sequence and diversity. Sequencing of ISHp608 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 ISHp608 that were identical for nearly 1.7 kb and that either (i) the ends of these ISHp608 elements were traversed where sequences became mixed or (ii) the copies of this element differed in sequence at one or both ends.

This ISHp608 sequence contained open reading frames (ORFs) of 156 and 384 codons that were protein level homologs of orfA and orfB, respectively, in IS605 (Fig. 1A). However, the two ORFs in ISHp608 were in the same orientation and overlapped by 10 codons, whereas their homologs in IS605 and IS606 were divergent. Their parallel orientation in ISHp608 is reminiscent of that in IS607, an element with homology to ISHp608 only in orfB, not orfA.

ISHp608 copy numbers in PeCan18B and four other ISHp608-positive strains (identified in preliminary dot blot tests) were estimated in Southern blots of genomic DNAs that had been singly digested by HindIII and SphI and probed
FIG. 2. ISHp608 terminal sequences in several H. pylori strains. Left and right termini of ISHp608 determined by sequence comparison are shown in uppercase, and flanking DNA and empty sites in reference strains 26065 and J99 (HP and jhp gene designations, respectively [4, 35]) are shown in lowercase letters. The left end of ISHp608 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 26065 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 ISHp608 copies. The right end of ISHp608 copy 2 from strain PeCan18B and flanking sequence were determined by adapter PCR and sequencing. Attempts to obtain the leftmost ~30 bp of ISHp608 copy 2 by PCR with jhp0942-specific primers and by sequencing directly on genomic DNA were not successful. The left and right ends of ISHp608 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 ISHp608 in strain PeCan16A and the flanking sequence were determined by adapter PCR and sequencing. The left end was identified by genomic DNA sequencing, but the flanking sequence was not determined because this strain contained two ISHp608 copies and because no product was obtained after PCR using primers specific for ISHp608 and for the anticipated flanking IS605 sequence. Both ends and flanking sequences of the single copy of ISHp608 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 ISHp608 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 26065).

with an internal 460-bp PCR fragment from ISHp608 (made with primers 5 and 12; Fig. 1B). Two strains (PeCan4A and PeCan10B) yielded just one hybridizing band in each digest, indicating just one ISHp608 copy per strain, whereas PeCan18 and the two other strains (PeCan2A and PeCan16A) each yielded two hybridizing bands, indicating two ISHp608 copies per strain.

The sequences of ISHp608 insertion sites were defined more closely by adapter PCR or direct sequencing of genomic DNAs from several representative ISHp608-carrying strains and identification of corresponding “empty site” sequences (lacking ISHp608) in databases of reference strains J99 and 26065. Sequences of each ISHp608 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 ISHp608 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 IS605 and IS606, which insert preferentially downstream of 5′-TTTA and 5′-TTTTAT, respectively (18). The organization of sequences within the termini of ISHp608 will be presented in the context of sequence conservation and diversity among ISHp608 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 ISHp608 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 1 from strains PeCan2, India49, Spain55, and Lih5, 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 ISHp608 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 ISHp608 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 ISHp608 element sequenced (identified with an asterisk in Fig. 4), in the orfBs 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 D2 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
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 ISHp608 elements each contained (i) a well-conserved direct repeat of 22 bp in the left and right ends; (ii) an octanucleotide (5’AACGCTT) or penta- or hexanucleotide (5’AAGCG or 5’AACGCC) direct repeats in the left ends of type 1 and type 2 elements, respectively; and (iii) conserved decaemer (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 ISHp608 left end downstream of 5’-TTAC target sequences, it is noteworthy that this left end varied in length and sequence: 5’-CC(A)CT or 5’-C(A)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).

**ISHp608 geographic distribution.** The frequency of ISHp608 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 1 and type 2 ISHp608s (primers 4 and 12; Fig. 1B). Hybridization and PCR results were complementary and entirely consistent; the same 22 strains were scored as positive for ISHp608 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 ISHp608 is rare or absent in East Asian strains.

ISHp608 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). ISHp608 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 ISHp608 carriage and overt disease was observed among cases for which clinical data were available.

All strains found to carry ISHp608 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 ISHp608 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 ISHp608 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

---

**FIG. 3.** DNA sequence diversity of ISHp608 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 ISHp608s 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 ISHp608 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 ISHp608, 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 and 87% with type 1. Strain PeCan18B contained two copies of ISHp608, 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 ISHp608, 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 ISHp608 element.
primers (7 and 14), either among the 123 strains that had been scored as containing ISHp608 by PCR with general primers (4 and 12) or among any of 212 strains tested that had been scored as lacking ISHp608 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.

ISHp608 transposition in E. coli and mutations affecting transposition. The ISHp608 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 ISHp608 element was marked genetically using a cam (resistance) gene; plasmids containing these ISHp608: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<sup>r</sup> 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 single colonies from each transformant clone. ISHp608 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<sup>r</sup> Str<sup>r</sup> exconjugants at average frequencies of $5 \times 10^{-8}$ to $10 \times 10^{-8}$ per starting donor cell, and nearly all these exconjugants (179 of 180 and 169 of 177, respectively) were Amp<sup>r</sup>. This result indicated that ISHp608 could transpose in E. coli and that most transposition involved simple separation of the marked ISHp608 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 (IS605 and IS607) (18, 19).

In contrast, ISHp608 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<sup>r</sup> Str<sup>r</sup> 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<sup>r</sup>, as expected of replicon fusions generated by illegitimate recombination. Thus, orfA is needed for ISHp608 transposition in E. coli, whereas orfB is not. The data also indicated that a sequence near the ISHp608 right end is needed for transposition, probably as a site on which transposition proteins act.

The ends and insertion specificity of ISHp608 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 ISHp608-specific primers 2 and 18. Figure 6 shows that in all 13 cases, each end of ISHp608 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 ISHp608 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 ISHp608 sequence types and found that ISHp608 was nonrandomly distributed in H. pylori populations, that it formed sim-
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 orfAs (e.g., in _ISHp608_ and _IS607_) and related orfAs in two orientations (e.g., in _ISHp608_ and _IS605_) 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.

**FIG. 5.** Repeat sequences in the termini of _ISHp608_ 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 _ISHp608_ ends. Horizontal lines identify shorter direct repeats: octanucleotide (5'-AACGCCTT-3') and related penta- or hexanucleotides (5'-AACGC and 5'-AACGCC) in the left ends of type 1 and type 2 elements, respectively, and dodecamer (5'-GCTTTAGCTA-3') or 12-mer (5'-TAGCTTTAGCTA-3') 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 _ISHp608_ right end (insertions 4 and 5; see Fig. 1B). Lowercase letters represent flanking sequences. (B) Subterminal imperfect inverted repeats of _ISHp608_. Left and right ends of _ISHp608_ (same strand) are aligned to highlight inverted repeat sequence (capital letters).
alleles of the Asian and other strains are generally distinguishable based on East Asia. These results are reminiscent of from many parts of the world, but in only 1% of strains from 2, position 99/100. The products of 2, position 65422/3; 1 – 1, position 80949/50; 5 – 1, position 70860/1. The one other IS608 element in Indian strains was identified by sequencing as a type 1-type 3 recombinant (Fig. 3B).

ISHp608 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 IS608 data suggest that ISH608 was transferred into H. pylori after the East Asian and Western H. pylori populations were separated. ISH608 would then have spread by DNA transfer between strains in a population and by recombination and transposition. This spread would have been accelerated if ISH608 contributed to bacterial fitness. The paucity of ISH608 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 ISH608s in Alaskan and Peruvian strains were distinct (type 2), differing by some 11% in DNA sequence from most other ISH608 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 IS608 elements from Spain or among the 28 other ISH608s 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 ISH608 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 ISH608s 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 ISH608 elements, even though ISH608 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 ISH608 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-
ditions that also fostered interstrain gene transfer and/or transposition of type 2 ISHp608 elements.

The orfB-gipA (Salmonella virulence gene) homology (32) and contributions of the unrelated Tn5 transposase to E. coli fitness (16) serve as reminders that the frequency of an ISHp608 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 ISHp608 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 ISHp608 in E. coli was orfA dependent but orfB independent, as was the case with IS607, a family member related to ISHp608 only in orfB (protein level) (19). Essentially all ISHp608 transposition products were simple insertions, not cointegrates (which contain both plasmid vector and ISHp608 DNAs), whereas most IS605 and IS607 transposition products were cointegrates (18, 19). Where tested (ISHp608 and IS607), 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-joint) and replicative transposition, respectively (6, 28, 30, 36). Studies with Tn7 and IS903 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 ISHp608 versus IS605 and IS607.

Inspection of the sequences flanking ISHp608 in H. pylori and after transposition in E. coli indicated that (i) ISHp608 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 IS605 and IS606, except that they inserted next to 5′-TTTAA and 5′-TTTAT, respectively (18). In contrast, IS607, whose OrfA is unrelated, inserted preferentially between adjacent G nucleotides in target DNA (19). These patterns suggest that OrfA proteins help insert 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 ISHp608 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 ISHp608 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 ISHp608 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 D1, D2, and E residues in the linear transposase sequence. In the Tn5 transposase, whose D2-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 D2 and E in orfB versus 33 to more than 100 amino acids in others [21]). Based on the Tn5 precedent (14), this short spacing might affect ISHp608 OrfB protein interaction with DNA substrates.

Continued population-level analyses of ISHp608 and related elements and the proteins they encode may help elucidate the origins and modes of evolution of prokaryotic transposable elements, factors 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. Busseleu, J. Combe, T. Alarcon, J. C. Atherton, H. Chaloukas, A. Chowdhury, A. Janulaitis, L. Kiuvaliene, H. K. Lee, M. Lopez Brea, G. B. Nair, M. Rahman, B. Wong, I. Segal, and E. J. Thomas. We thank J. Slauch for helpful discussions about gipA.

This research was supported in part by U.S. Public Health Service grants AI8166, AI49161, DK53727, and P30 DK52574.

**REFERENCES**


