2004

Evolutionary dynamics of insertion sequences in Helicobacter pylori

Awdhesh Kalia
Washington University School of Medicine in St. Louis

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

Giedrius Dailide
Washington University School of Medicine in St. Louis

Yoshiyki Ito
Washington University School of Medicine in St. Louis

Takeshi Azuma
Fukui Medical School

See next page for additional authors

Recommended Citation

Kalia, Awdhesh; Mukhopadhyay, Asish K.; Dailide, Giedrius; Ito, Yoshiyki; Azuma, Takeshi; Wong, Benjamin C. Y.; and Berg, Douglas E., "Evolutionary dynamics of insertion sequences in Helicobacter pylori." Journal of Bacteriology. 186, 22. 7508-7520. (2004).
https://digitalcommons.wustl.edu/open_access_pubs/2714

Journal of Bacteriology

Evolutionary Dynamics of Insertion Sequences in *Helicobacter pylori*

Awdhesh Kalia, Asish K. Mukhopadhyay, Giedrius Dailide, Yoshiyki Ito, Takeshi Azuma, Benjamin C. Y. Wong and Douglas E. Berg

J. Bacteriol. **2004, 186(22):7508. DOI: 10.1128/JB.186.22.7508-7520.2004.**

JOURNAL OF BACTERIOLOGY, Nov. 2004, p. 7508–7520 Vol. 186, No. 22
0021-9193/04/$08.00+0 DOI: 10.1128/JB.186.22.7508–7520.2004

Evolutionary Dynamics of Insertion Sequences in *Helicobacter pylori*

Awdhesh Kalia,[1] Asish K. Mukhopadhyay,[1]† Giedrius Dailide,[1] Yoshiyki Ito,[1,2] Takeshi Azuma,[2] Benjamin C. Y. Wong,[3] and Douglas E. Berg[1]*

Department of Molecular Microbiology, Washington University School of Medicine, Saint Louis, Missouri[1]; Second Department of Internal Medicine, Fukui Medical School, Fukui, Japan[2]; and Division of Gastroenterology, Department of Medicine, University of Hong Kong, Queen Mary Hospital, Hong Kong[3]

Received 21 April 2004/Accepted 5 August 2004

Prokaryotic insertion sequence (IS) elements behave like parasites in terms of their ability to invade and proliferate in microbial gene pools and like symbionts when they coevolve with their bacterial hosts. Here we investigated the evolutionary history of IS*605* and IS*607* of *Helicobacter pylori*, a genetically diverse gastric pathogen. These elements contain unrelated transposase genes (*orfA*) and also a homolog of the *Salmonella* virulence gene *gipA* (*orfB*). A total of 488 East Asian, Indian, Peruvian, and Spanish isolates were screened, and 18 and 14% of them harbored IS*605* and IS*607*, respectively. IS*605* nucleotide sequence analysis ($n = 42$) revealed geographic subdivisions similar to those of *H. pylori*; the geographic subdivision was blurred, however, due in part to homologous recombination, as indicated by split decomposition and homoplasy tests (homoplasy ratio, 0.56). In contrast, the IS*607* populations ($n = 44$) showed strong geographic subdivisions with less homologous recombination (homoplasy ratio, 0.2). Diversifying selection (ratio of nonsynonymous change to synonymous change, $\gg$1) was evident in ~15% of the IS*605 orfA* codons analyzed but not in the IS*607 orfA* codons. Diversifying selection was also evident in ~2% of the IS*605 orfB* and ~10% of the IS*607 orfB* codons analyzed. We suggest that the evolution of these elements reflects selection for optimal transposition activity in the case of IS*605 orfA* and for interactions between the OrfB proteins and other cellular constituents that potentially contribute to bacterial fitness. Taken together, similarities in IS elements and *H. pylori* population genetic structures and evidence of adaptive evolution in IS elements suggest that there is coevolution between these elements and their bacterial hosts.

The insertion sequences (IS) of bacteria are discrete DNA segments that are distinguished by their ability to move within genomes without a need for extensive DNA homology. These elements move within and between species by DNA transfer and transposition, and they proliferate within bacterial genomes by transposition per se (for general reviews see references 3, 15, and 16). Most elements seem to be innocuous or parasitic (19). However, coadaptation between elements and their hosts during long-term association may also make carriage beneficial in certain cases (38, 43).

Chance and natural selection may each contribute to the maintenance of IS elements in bacterial populations. Once a given element has entered a gene pool, its abundance should reflect its rates of interstrain transfer and transposition, coupled with any benefit or cost resulting from its carriage in terms of bacterial survival, growth, or adaptability in variable and often hostile environments. Elements that confer resistance to antibiotics or metals provide dramatic examples of contributions to host fitness. Certain mutations caused by IS element movement may also be adaptive (11, 53, 56), and the product of the transposase gene of at least one element (IS*50*) contributes to host fitness, independent of movement (25). The known deleterious effects of transposable elements are related to their movement, and they include inactivation of genes by insertion, adjacent deletion, and other chromosomal rearrangements. The position of a mobile element in the spectrum from parasitic to beneficial in its interactions with its host may depend, at least in part, on the selection imposed by the host's background genotype and environment.

Much of our understanding of mechanisms that regulate IS element maintenance and evolution in bacterial populations has come from studies of the elements that are common in natural isolates of *Escherichia coli* and *Salmonella* (5, 26, 41, 45, 48, 53). These enteric organisms have strongly clonal population structures and often carry multiple unrelated elements in their chromosomes and/or plasmids. Interstrain transfer is relatively frequent, and transfer between species has also been observed. Consequently, IS phylogenies are often incongruent with the phylogenies of the chromosomal genes of the bacterial hosts; nearly identical elements are often found in divergent bacterial lineages, and conversely, divergent elements are sometimes found in closely related lineages (5, 24, 41, 54). In *E. coli* nucleotide sequences within each class of IS elements are highly conserved (>99% similar to a consensus sequence), and comparisons with sequences from chromosomal genes suggest that these elements move often among host lineages (41). Nucleotide sequence analysis of IS*1* and IS*3* homologs from diverse enteric species also indicated that there is significant intragenic recombination, which might contribute to IS element evolution (41).

To gain insight into the role of natural selection in the maintenance and evolution of IS elements in bacterial populations, we investigated the evolutionary histories of IS*605* and IS*607* of *Helicobacter pylori*, a gram-negative bacterium that is

* Corresponding author. Mailing address: Box 8230, Department of Molecular Microbiology, 4940 Parkview Place, Washington University School of Medicine, Saint Louis, MO 63110. Phone: (314) 362-2772. Fax: (314) 362-1232. E-mail: berg@borcim.wustl.edu.

† Present address: National Institute for Cholera and Enteric Diseases, Kolkata, India.

TABLE 1. Primers used in the present study

Gene	Primer	Nucleotide sequence	Location
IS*605 orfA*	ORF18F	5′-CGCCTTGATCGTTTCAGGATTAGC	112 bp from left end of IS*605*
	ORF18R	5′-CAACCAACCGAAGCAAGCATAATC	482 bp from left end of IS*605*
IS*605 orfB*	ORF19F	5′-GGCTGTTCTAGGGTCGTGTATAAC	658 bp from left end of IS*605*
	ORF19R	5′-CAAGCTAGATGCAATCTAGCTACC	1319 bp from left end of IS*605*
IS*607 orfA*	FlkF	5′-GGCTACAAACAGAAACTAAAAT	237 bp from left end of IS*607*
	F3	5′-ATGCGTGATTGAGATAGCTG	740 bp from left end of IS*607*
IS*607 orfB*	R6	5′-GCGCTAGATTGTATGGCTCT	1386 bp from right end of IS*607*
	F2	5′-GCATAGATATTTAAGCCATTAGA	1351 bp from left end of IS*607*
	F10	5′-CTCTTAATTTAGGATTTTTGTTG	1984 bp from left end of IS*607*

implicated in peptic ulcer disease and gastric cancer (6, 14). *H. pylori* is of particular interest for such studies because it is extremely diverse genetically and has robust geographic subdivisions that are far more robust than those of any other bacterial pathogen (1, 22, 23, 36, 46, 52). Its great genetic diversity can be ascribed to multiple factors, including (i) mutation, (ii) recombination, (iii) highly localized (preferentially intrafamilial) transmission and the resulting lack of species-wide selective sweeps, and (iv) selection for different bacterial traits in different human hosts (20, 21, 47, 61). Barriers to gene flow resulting from geographic isolation of ancestral *H. pylori* populations may also have caused considerable genetic drift and thus differentiation of various *H. pylori* subpopulations (22). Each of these factors may also have influenced the evolutionary dynamics of IS elements in the *H. pylori* gene pool.

IS*605* and IS*607* belong to a novel family of transposable elements that are chimeric; each element contains two genes, *orfA* and *orfB*, which apparently have different phylogenetic origins (9, 10, 34, 35). The *orfA* genes of IS*605* and IS*607* encode transposases belonging to two different families. IS*607 orfA* is predicted to encode a serine recombinase, most homologs of which catalyze only site-specific recombination (59), although one case in which this enzyme mediates both site-specific recombination and transposition has also been reported (39). In contrast, the IS*605 orfA* product is not considered to be a serine recombinase based on amino acid homologies. Consistent with transposase differences, the elements differ in their insertion specificities. IS*605* inserts preferentially downstream of a pentanucleotide motif, 5′-TTTAA, whereas IS*607* inserts preferentially between G residues in a GG target sequence. The second gene of these elements, *orfB*, which is related to the genes encoding several putative bacterial transposases (10), is not required for transposition in the *E. coli* model, but it is related to the gene encoding the *Salmonella* prophage-encoded GipA protein, which promotes *Salmonella*'s growth and survival in Peyer's patches of the intestine (60).

Nucleotide sequence analysis of IS elements in bacterial populations can provide insights into the evolutionary histories of the elements and sometimes also into the histories of their microbial hosts (5, 41). In this study, we elucidated the evolutionary histories of IS*605* and IS*607* in the context of *H. pylori*'s extensive genetic diversity and geographic population subdivision.

MATERIALS AND METHODS

Bacterial isolates and geographical origins. *H. pylori* isolates were cultured by using standard conditions and were maintained as frozen stocks at −70°C in sterile brain heart infusion broth containing 15% glycerol. A total of 488 *H. pylori* isolates from different parts of the world were included in this study. The individual population profiles are described below.

Sixty-eight Spanish isolates obtained from Teresa Alarcon and Manuel Lopez Brea have been described previously (37), as have 76 isolates cultured from biopsies provided by Abhijit Choudhury from ethnic Bengali patients in Calcutta, India (47). Fifty-one strains isolated from Amerindian (native Peruvian) patients from Lima, Peru, were provided by Robert H. Gilman. For the East Asian strains, 140 isolates were obtained from ethnic Japanese patients in Fukui, Honshu, Japan, 103 isolates were obtained from ethnic Japanese or Okinawan patients in Okinawa, and 50 isolates were obtained from ethnic Chinese patients in Hong Kong. All isolates were obtained from patients with gastric complaints that had undergone diagnostic endoscopy with informed consent.

PCR amplification, dot blot hybridization, and nucleotide sequencing. Chromosomal DNA was prepared from confluent cultures on brain heart infusion agar plates either by the hexadecyltrimethylammonium bromide extraction method as described previously (39) or with a QIAamp DNA minikit (QIAGEN Inc., Chatsworth, Calif.). Primers used for generating probes for hybridization, PCR, and nucleotide sequencing are listed in Table 1. Dot blot hybridization was performed by using Hybond-N1 nylon membranes and an ECL kit (Amersham Pharmacia Biotech, Piscataway, N.J.) according to the manufacturer's instructions. Probes for the IS*605 orfA* and *orfB* genes were generated by PCR by using *H. pylori* strain 26695 genomic DNA as the template with primer pairs ORF18F-ORF18R and ORF19F-ORF19R, respectively. A probe for IS*607* was generated by using *H. pylori* strain CPY41 genomic DNA as the template with primers FlkF and F10. Specific PCR were carried out in 20-μl mixtures containing 10 ng of DNA, 1 U of *Taq* polymerase (Promega, Madison, Wis.), 10 pmol of each primer, each deoxynucleoside triphosphate at a concentration of 0.25 mM, and 2 to 3 mM $MgCl_2$ in standard PCR buffer for 30 cycles, generally under the following conditions: 94°C for 40 s, 55°C for 40 s, and 72°C for a time chosen based on the size of the expected fragment (1 min/kb).

To assess the extent of geographic subdivision in IS element populations, PCR products obtained from IS*605* ($n = 42$) and IS*607* ($n = 44$) were sequenced. Isolates were selected primarily to sample the geographic extremes of East Asia (Japan, Hong Kong) and Europe (Spain). Peruvian and Indian *H. pylori* isolates in our collection exhibited substantial European influence (36, 46). Therefore, for the purposes of this study, Indian, Spanish, and Peruvian strains were placed in the Indo-European group, and strains from Japan (including Honshu and Okinawa) and Hong Kong were placed in the East Asian group. Sampling within each group was randomized to test the hypothesis of geographic isolation between the two groups.

The IS*605 orfA* and *orfB* genes were amplified in a single PCR with primers ORF18F and ORF19R. To ensure twofold redundancy, the PCR product was sequenced by using primers ORF18F and ORF19R and also primers ORF18R and ORF19F (internal primers) (Table 1). Similarly, the IS*607* FlkF-F10 PCR product was sequenced by using primers FlkF and F10 and also primers F3, R6, and F2 (internal primers) (Table 1). PCR products were sequenced directly after purification with a QIAquick gel extraction kit (QIAGEN Inc.) by using a Big-Dye terminator cycle sequencing kit (Perkin-Elmer Applied Biosystems, Foster City, Calif.) and an ABI 377 automated sequencer. Of the 86 IS elements for which nucleotide sequences were determined, 38 were resequenced to check for possible sequencing or PCR errors. The IS elements selected for verification included those which had identical sequences or those which had stop codons in the coding region. The resequencing data were 100% concordant with the first data set.

Computations. Preliminary sequence data management, including editing and alignment, was done by using the VectorNTI suite of programs (Informax Inc., Carlsbad, Calif.).

(i) Polymorphism and divergence. Tests to obtain estimates of the average nucleotide diversity within populations (average nucleotide divergence per site [π]), the average nucleotide diversity at synonymous sites (π_S) and nonsynonymous sites (π_{NS}) corrected for multiple hits, the average nucleotide divergence between populations (D_{xy}), and F_{ST} and permutation tests for geographic isolation and genetic differentiation were done with DNASP, version 4 (58). A permutation (randomization) approach was used to detect geographic subdivision and genetic differentiation (30, 31). The significance of the observed values for K_{S*}, K_{ST}, Z_S, Z_{S*}, and S_{NN} test statistics was obtained by 10,000 random permutations as implemented in DNASP, version 4.

(ii) Phylogenetic analysis. Phylogenetic reconstruction of IS*605* and IS*607* nucleotide sequences was performed by using the maximum-likelihood (ML) approach implemented in PAUP*4b10 (63). All available sequences were used. An optimal model of DNA substitution that best described the pattern of nucleotide variation and sequence evolution of IS elements was determined by using MODELTEST, version 3.06 (55). For this purpose, an initial neighbor-joining tree was computed with the uncorrected p-distance measure and used as the input to estimate the likelihood parameters for 56 nested models of evolution. Models that were nested (i.e., where the simpler model [null hypothesis, H_0] was a special case of the more general model [alternate hypothesis, H_1]) were compared statistically by using the hierarchical likelihood ratio test (LRT) (32). The LRT includes generating a likelihood ratio statistic (LRS), which is computed as follows: $LRS = 2 \times [(-InLH_1) - (-InLH_0)]$, where InL is the log-likelihood score of a given model. The product 2LRS follows a χ^2 distribution with the degrees of freedom equal to the difference in the number of parameters between the two models. An ML phylogeny was reconstructed under the best-fit model by using a combination of heuristic searches and branch swapping (with the tree bisection reconnection algorithm) to further optimize the likelihood score and substitution parameters, including among- and within-site rate variation. The significance of observed groupings in the phylogeny was assessed by a bootstrap analysis performed with 1,000 replicates under the distance optimality criterion, while the ML-optimized model and parameters were incorporated by using a combination of heuristic searches and branch swapping. Phylogenetic trees were visualized with TreeView, version 1.6.6 (49), and were edited further in Microsoft PowerPoint.

Homologous recombination in IS*605* and IS*607* was assessed by split decomposition (SD) analysis (2) and the homoplasy test (44). SD analysis uses a relaxed condition for estimating splitting patterns among taxa, and both the signal (suggesting the tree structure) and the noise (suggesting patterns inconsistent with the tree structure, such as those arising from recombination and/or convergent evolution) can be presented simultaneously. Unlike standard tree-building methods, SD analysis can usually produce a network that helps delineate anomalies in the history of gene trees (3, 50, 62). In particular, recombination generates sequences linked by multiple ancestors, which with this method are depicted as an interconnected network of phylogenetic relationships (50). For the analysis reported below, nucleotide data were transformed into a matrix of pairwise distances incorporating ML-optimized substitution parameters. In some cases, SD analysis was also done by using the uncorrected p-distance measure. The analysis was done with SplitsTree, version 3.2 (33). The homoplasy test was done by using the program HOMOPLASY (kindly provided by Mark Achtman). Estimates of *Se* (effective site number) and the homoplasy ratio (*h*) were obtained by assuming all three levels of expression (low, medium, and high) in three independent runs. Results were reported by averaging over all runs.

(iii) Analysis of selection pressures. The selection pressure on a protein-encoding gene can be measured by comparing the nonsynonymous substitution rate (d_N) (amino acid altering) to the synonymous substitution rate (d_S) (silent, no amino acid change) to obtain the d_N/d_S ratio (ω). An ω of 1 indicates neutral evolution (relaxed selective constraint; nonsynonymous changes have no associated fitness advantage and are fixed at the same rate as synonymous changes); an ω of <1 indicates purifying selection (strong functional constraint; nonsynonymous changes are deleterious for protein function and are fixed at a lower rate than synonymous changes); and an ω of >1 indicates positive selection (adaptive evolution; nonsynonymous changes are favored because they confer a fitness advantage and are fixed at a higher rate than synonymous changes).

Here we used an ML-based approach for measuring selective pressures acting on *H. pylori*'s IS elements. ML analysis with site-specific codon substitution models that allow ω to vary among sites but not among lineages (68) detects positive selection at individual sites only if the average d_N over all lineages is higher than the average d_S. Thus, the test is still relatively conservative in terms of detecting positive selection. ω for codons of IS*607* and IS*605 orfA* and *orfB* were computed by using ML-optimized phylogenies of the *orf* genes (available from us upon request). Sequences with stop codons were excluded from this analysis.

The following six site-specific models of codon substitution were used in this study. (i) M0 assumes that all codons (sites) are subject to the same selection pressure, so a single ω value is estimated. (ii) M1 divides codons into two categories, representing the codons that are invariant (p_0), with ω_0 fixed at 0, and codons that are neutral (p_1), where ω_1 is 1. (iii) M2 accounts for positive selection by adding a third category of codons (p_2) with ω_2, which is estimated from the data and can be >1. (iv) M3 also invokes three codon site classes and provides a more sensitive test for positive selection by estimating all ω values from the data, and all ω values may be >1. (v) M7 uses a discrete β distribution (with 10 rate categories), whose shape varies depending on the parameters *p* and *q*, to model ω among codons; with M7, no class of codons can have an ω of >1. (vi) M8 also uses a β distribution, but an extra class of codons is incorporated, in which ω can be >1.

TABLE 2. Geographic prevalence of IS elements

Geographic origin	No. of strains	IS element prevalence		
		IS*605*	IS*607*	Both IS*605* and IS*607*
HongKong	50	7 (14)[a]	9 (18)	4 (8)
Japan (Honshu)	140	7 (5)	7 (5)	0 (0)
Japan (Okinawa)	103	8 (7.5)	18 (17.5)	1 (0.97)
India	76	19 (25)	10 (13)	1 (1.3)
Peru	51	18 (35)	14 (27)	1 (1.96)
Spain	68	28 (41)	12 (18)	2 (2.94)

[a] The numbers in parentheses are percentages.

Positive selection was inferred when codons with an ω of >1 were identified and the likelihood of the codon substitution model in question was significantly higher than the likelihood of a nested model that did not take positive selection into account. When the more general models indicated the presence of sites with ω of >1, the comparison constituted an LRT (see above) of positive selection (68). Finally, by using Bayesian methods we calculated the probability that a specific codon belonged to the neutral, negative, or positively selected class. This analysis was done with the CODEML program in the PAML suite, version 3.13d (67).

Nucleotide sequence accession numbers. Sequence data determined in this study have been deposited in the GenBank/EMBL/DDBJ databases under accession numbers AY687641 to AY687726, AY687742 to AY687785, and AY687787 to AY687828.

RESULTS

Population structure of *H. pylori* IS elements: geographic distribution of IS*605* and IS*607* in *H. pylori* populations. The distribution of IS*605* and IS*607* in *H. pylori* isolates from several parts of the world was analyzed by dot blot hybridization, followed by confirmatory PCR. IS*605* was found in 18% of the isolates worldwide, but the frequency varied geographically, ranging from only 5% in Japanese (Honshu) isolates to 41% in Spanish isolates (Table 2). These outcomes agree with previous reports of IS*605* carriage in 28 to 50% of mainly Western *H. pylori* collections (9, 29, 34), but they indicate that this element is significantly less common in East Asian isolates ($P <$ 0.05 for all pairwise comparisons except Hong Kong versus India).

IS*607* was found in 14% of the isolates tested, but the frequency also varied geographically, ranging from 5% in Japanese (Honshu) *H. pylori* isolates to 27% in Peruvian isolates (Table 2). As observed with IS*605*, significantly fewer Honshu isolates than isolates from elsewhere carried IS*607* ($P < 0.05$). However, the IS*607* prevalence in Okinawan and Hong Kong isolates did not differ significantly from that in Spanish, Indian, or Peruvian isolates ($P > 0.1$). IS*607* was less abundant than IS*605*, especially among Spanish isolates (for IS*607* versus IS*605*, $P < 0.01$); this general trend was also detected among Peruvian and Indian isolates, although the differences were not statistically significant.

Of the 157 isolates that were positive for IS*605* and IS*607*, only 9 (6%) carried both elements (Table 2). The co-occurrence of these elements in most geographic regions was ex-

TABLE 3. DNA divergence in *H. pylori* IS elements and chromosomal genes

Gene	No. of sequences	Sequence length (bp)	Mean nucleotide diversity [$\pi_{(\mathrm{JC\ corrected})}$][b]	Mean nucleotide diversity at synonymous sites [$\pi_{\mathrm{S(JC\ corrected)}}$]	Mean nucleotide diversity at nonsynonymous sites [$\pi_{\mathrm{NS(JC\ corrected)}}$]
IS*605 orfA*	42	378	0.051	0.153	0.029
IS*605 orfB*	42	549	0.035	0.128	0.013
IS*607 orfA*	44	537	0.021	0.078	0.007
IS*607 orfB*	44	563	0.033	0.113	0.013
atpA[a]	310	627	0.032	0.136	0.002
efp[a]	303	410	0.035	0.175	0.001
ppa[a]	317	398	0.025	0.107	0.005
yphC[a]	332	510	0.068	0.164	0.014
ureI[a]	335	585	0.031	0.123	0.006

[a] Sequences for the housekeeping genes listed were downloaded from http://helicobacter.mlst.net and have been described previously (22).
[b] Jukes-Cantor correction was applied for multiple hits.

tremely low and ranged from 0% in Japan (Honshu) to 8% in Hong Kong.

Nucleotide sequence diversity. We determined the nucleotide sequences for 378 bp of *orfA* and 549 bp of *orfB* from IS*605* in 42 strains and for 537 bp of *orfA* and 563 bp of *orfB* from IS*607* in 44 strains; in each case the strain was chosen as described in Materials and Methods. The average nucleotide diversity (i.e., nucleotide divergence per site) in these elements ranged from 0.021 (IS*607 orfA*) to 0.051 (IS*605 orfA*). The average nucleotide diversity at synonymous sites, where the accumulated changes reflect the ages of alleles (4), ranged from 0.078 (IS*607 orfA*) to 0.153 (IS*605 orfA*). The IS π and π_S values were fairly similar to those of five other *H. pylori* normal chromosomal genes (housekeeping genes) (Table 3). The average nucleotide diversity values for nonsynonymous sites ranged from 0.007 in IS*607 orfA* to 0.029 in IS*605 orfA*. Stop codons due to nucleotide substitutions were identified in 1 of 44 IS*607 orfA* genes, in 5 of 42 IS*605 orfA* genes, and in one IS*607 orfB* gene. In all other cases the *orfA* and *orfB* sequences were not interrupted.

Phylogenetic relationships inferred from IS*605* sequences. A 927-bp sequence obtained from concatenating the IS*605 orfA* and *orfB* sequences was used for phylogenetic reconstruction. Figure 1 shows an unrooted ML tree of IS*605* nucleotide sequences obtained from 42 *H. pylori* isolates. The East Asian isolates formed one weakly supported group (group I; bootstrap support, 41%) and sequences from Hong Kong, Honshu, and Okinawa were intermingled with sequences from Spain and Peru. A second weakly supported group (group II; bootstrap support, 50%) consisted mostly of Spanish isolates that were intermingled with Peruvian and Indian isolates; one Honshu isolate (CPY3281) also clustered with group II. Three distinct groups were seen among 14 Indian *H. pylori* isolates. IS*605* sequences from four Indian *H. pylori* isolates (I-41, I-69, I-111, and I-48) clustered with group II Spanish and Peruvian *H. pylori* isolates, and IS*605* sequences of the other 10 isolates formed two additional clusters, group IndA (four isolates; bootstrap support, 99%) and group IndB (six isolates; bootstrap support, 51%). Taken together, the IS*605* phylogeny exhibited blurred population subdivisions, in which three of the four major groups were only weakly supported.

The nucleotide sequence diversity within subpopulations ranged from 1% in group IndA to 3.4% in group II (Table 4). In all pairwise comparisons of IS*605* subpopulations, the mean divergence between groups (D_{xy}) was higher than the mean divergence within groups, which supported the phylogenetic distinctness (51). Statistical tests that measured the extent and significance of genetic differentiation (30, 31) suggested that IS*605* sequences from different populations had evolved with significant geographic isolation ($P < 0.01$ for K_{S*}, K_{ST}, Z_S, Z_{S*}, and S_{NN} test statistics) and had undergone very great genetic differentiation ($F_{ST} = 0.28l$) (Table 4) (27). Individual pairwise comparisons also revealed significant geographic isolation and genetic differentiation among IS*605* subpopulations.

Subpopulations that diverge from a common ancestor and subsequently evolve in isolation are expected to accumulate fixed polymorphisms (sites at which all sequences in one population differ from all sequences in a second population),

TABLE 4. Sequence divergence within and between phylogenetically distinct IS*605* taxa

Comparison	Genes	Mean nucleotide diversity within group[a]	Mean nucleotide diversity between groups[a]	No. of fixed polymorphisms[b]	No. of shared polymorphisms	F_{ST}	PM test *P* value[c]
Group I vs Group II	*orfA* + *orfB*	$\pi_I = 0.033$; $\pi_{II} = 0.034$	0.05	0	52	0.303	<0.001
Group I vs IndA	*orfA* + *orfB*	$\pi_I = 0.033$; $\pi_{IndA} = 0.01$	0.047	9 (6S + 3N)	5	0.341	<0.001
Group I vs IndB	*orfA* + *orfB*	$\pi_I = 0.033$; $\pi_{IndB} = 0.025$	0.037	0	29	0.225	<0.001
Group II vs IndA	*orfA* + *orfB*	$\pi_{II} = 0.034$; $\pi_{IndA} = 0.01$	0.041	1N	5	0.163	<0.05
Group II vs IndB	*orfA* + *orfB*	$\pi_{II} = 0.034$; $\pi_{IndB} = 0.025$	0.048	0	25	0.364	<0.001
IndA vs IndB	*orfA* + *orfB*	$\pi_{IndA} = 0.01$; $\pi_{IndB} = 0.025$	0.035	14 (10S + 4N)	9	0.252	<0.01
All groups	*orfA* + *orfB*	NA[d]	NA	NA	NA	0.281	<0.01

[a] Values were calculated with the Jukes-Cantor correction (correction for multiple hits).
[b] S, synonymous polymorphisms; N, nonsynonymous polymorphisms.
[c] A permutation test was done with 10,000 replications; *P* values were calculated for the K_{s*}, K_{ST}, Z_s, Z_{s*} (30), and S_{nn} (31) test statistics. All test statistics demonstrated identical significance levels.
[d] NA, not applicable.

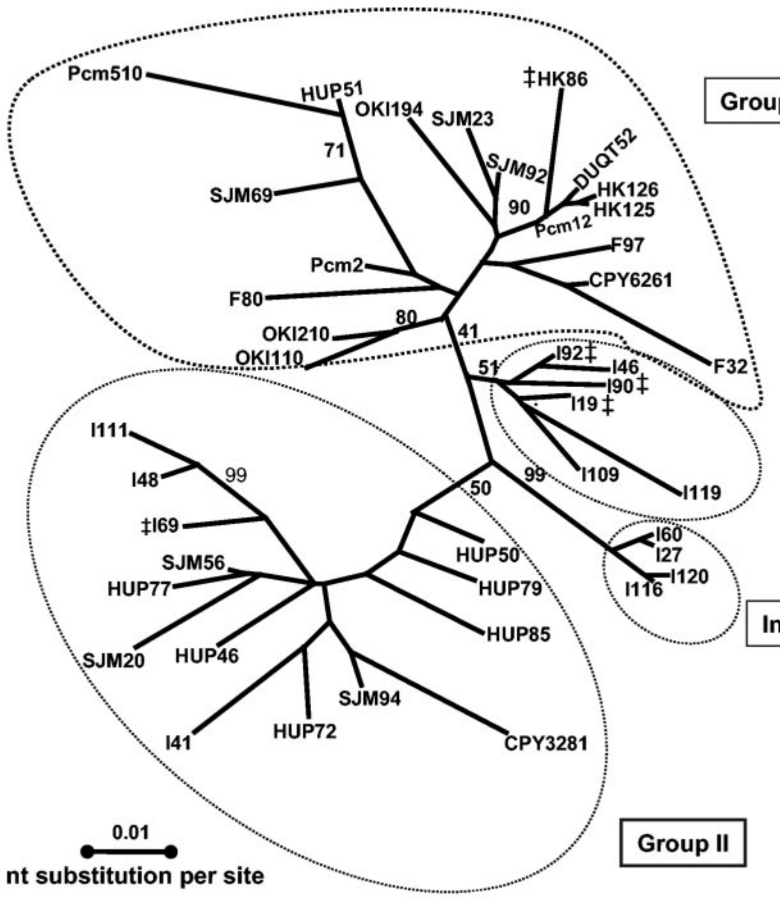

FIG. 1. Unrooted, radial gene tree generated by the ML method by using a 927-bp concatenated sequence from IS*605 orfA* (378 bp) and *orfB* (549 bp). All available sequences were used in the phylogenetic reconstruction. *orfA* sequences with stop codons are indicated by double daggers. The most appropriate model for IS*605* sequence evolution (TIM + I + Γ) was used for phylogenetic reconstruction with a discrete gamma distribution (Γ) shape parameter (α = 0.803) and an assumed proportion of invariate sites (I = 0.69). The TIM model (a constrained submodel of the general time reversible model) specifies a rate substitution matrix in which transitions (A↔G and C↔T) have only one rate category and the four possible transversions (A↔C, G↔T, A↔T, and C↔G) have four distinct rate categories. Bootstrap values of ≥50 are indicated at the nodes. *H. pylori* strain designations indicate the geographic origins, as follows: HUP, Spain; SJM, Peru; I, India; OKI, Okinawa, Japan; F or CPY, Honshu, Japan; and HK, DUQT, or Pcm, Hong Kong. Major sequence similarity clusters are circled. nt, nucleotide.

shared polymorphisms, and unique polymorphisms (i.e., sites polymorphic in one population but monomorphic in a second population) in their DNA sequences (28). The distributions of these polymorphisms are useful in inferring the times of population divergence; recently split populations show only shared polymorphisms, and populations that split a long time ago show an inverse relationship between fixed and shared polymorphisms or only fixed polymorphisms (64). Of six possible pairwise comparisons of IS*605* subpopulations, only three provided evidence for the presence of fixed polymorphisms, and all three of these involved the well-supported IndA group (Table 4). For all other comparisons, only polymorphisms shared by subpopulations were observed, suggesting that either the subpopulations diverged very recently or that recombination between subpopulations had obscured population differentiation (see below).

Phylogenetic relationships inferred from IS*607* sequences. An 1,100-bp sequence obtained from concatenating IS*607 orfA* and *orfB* was used for phylogenetic reconstruction. Figure 2A shows an unrooted ML tree for IS*607* nucleotide sequences obtained from the 44 *H. pylori* isolates. Two distinct sequence similarity clusters were evident that broadly distinguished group I ($n = 20$), consisting of East Asian *H. pylori* isolates, and group II ($n = 22$), consisting of Spanish, Peruvian, and Indian *H. pylori* isolates. Only 3 of the 23 East Asian *H. pylori* isolates, F-37 from Japan and pcm2 and DUQT-52 from Hong Kong, were not assigned to group I. The phylogeny also provided evidence for population subdivision within each group. Specifically, IS*607* elements from Japanese (Honshu) isolates and from Okinawan-Hong Kong *H. pylori* isolates formed two well-supported groups, designated group Ia (bootstrap support, 99%) and group Ib (bootstrap support, 85%), although two Okinanwan IS*607* sequences clustered with group Ia. IS*607* sequences from three Indian isolates formed a strongly supported group, designated group IIa (bootstrap support, 99), which was distinct from all other group II IS*607* sequences. Taken together, the IS*607* phylogeny exhibited well-supported geographic clustering that was markedly stronger than that in the IS*605* phylogeny.

The nucleotide diversity within IS*607* subpopulations ranged from 0.4% in group Ia to 1.8% in group II (Table 5). As seen with IS*605* populations, D_{xy}, the mean divergence between

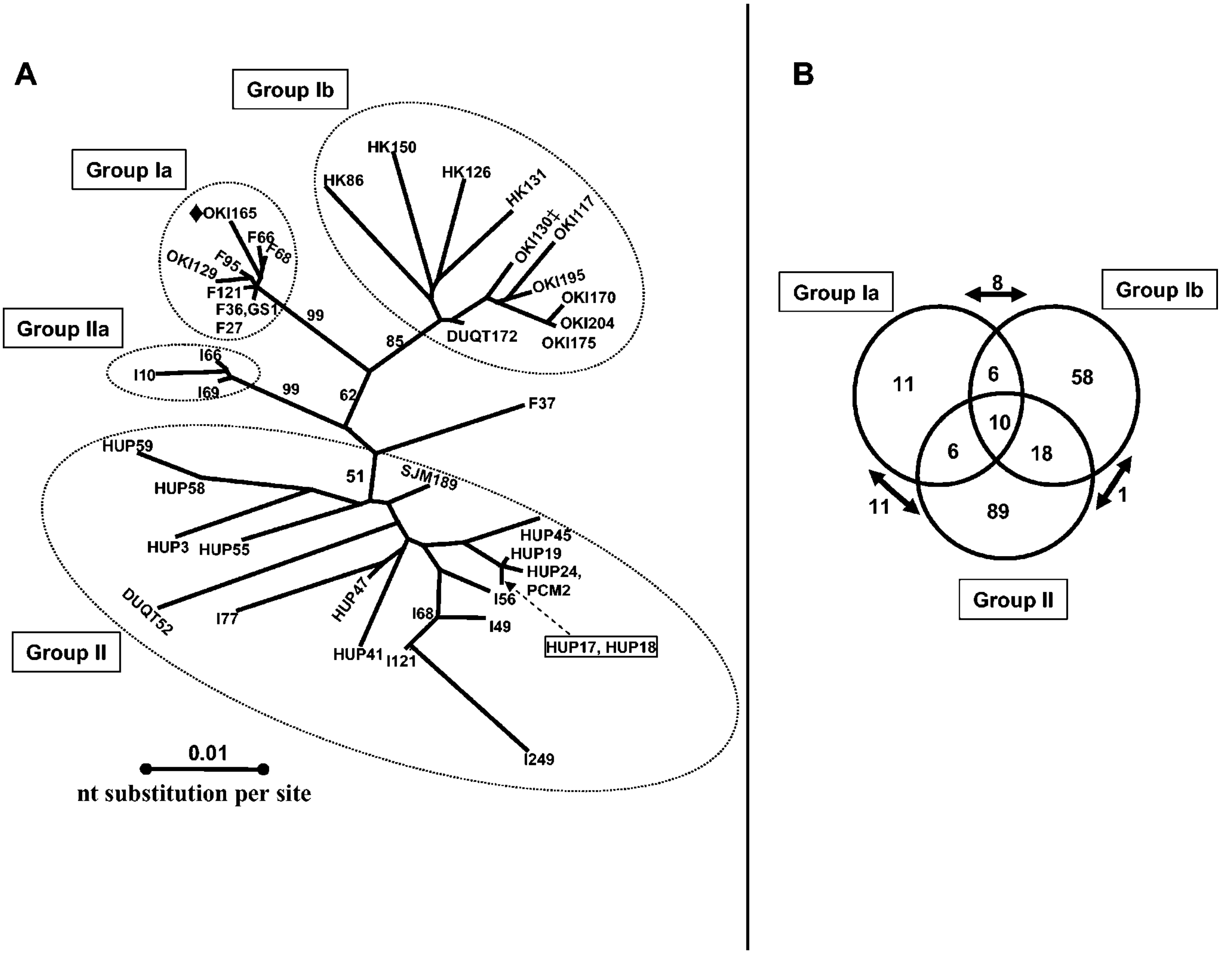

FIG. 2. (A) Unrooted, radial gene tree generated by the ML method by using an 1,100-bp concatenated sequence from IS*607 orfA* (537 bp) and *orfB* (563 bp). Sequences with stop codons in *orfA* and *orfB* are indicated by double daggers and solid diamonds, respectively. The TVM + I + Γ model was used in phylogenetic reconstruction with $\alpha = 0.304$ and $I = 0.46$. The TVM rate substitution matrix specifies that transitions have two distinct rate categories and transversions have only two specified rate categories (A↔C and G↔T have one rate and A↔T and C↔G have one rate). Bootstrap values of ≥50 are indicated at the nodes; distinct IS*607* sequence similarity clusters are circled. *H. pylori* strain GS1 is from Honshu, Japan; for an explanation of all other strain designations see the legend to Fig. 1. nt, nucleotide. (B) Summary of polymorphisms in IS*607* subpopulations: Venn diagram summarizing the unique, shared, and fixed polymorphisms observed in IS*607*group Ia ($n = 9$), group Ib ($n = 11$), and group II ($n = 19$). Fixed differences between populations are indicated below the arrows outside the conjoined circles.

groups, was higher than π, the mean divergence within groups, suggesting the phylogenetic distinctness of IS*607* subpopulations. The probabilities for the K_{S*}, K_{ST*}, Z_S, Z_{S*}, and S_{NN} test statistics obtained by the permutation test with 10,000 replicates were highly significant ($P < 0.001$) (Table 5). This suggested that (i) the data deviated significantly from the null hypothesis of no genetic differentiation or geographic isolation and (ii) the observed IS*607* popula-

TABLE 5. Sequence divergence within and between phylogenetically distinct IS*607* taxa[a]

Comparison	Genes	Mean nucleotide diversity within group	Mean nucleotide diversity between groups	No. of fixed polymorphisms	No. of shared polymorphisms	F_{ST}	PM test *P* value
Group Ia vs Group II	*orfA* + *orfB*	$\pi_{Ia} = 0.004$; $\pi_{II} = 0.018$	0.035	11 (8S + 3N)	6	0.691	<0.001
Group Ib vs Group II	*orfA* + *orfB*	$\pi_{Ib} = 0.015$; $\pi_{II} = 0.018$	0.034	1S	18	0.499	<0.001
Group Ia vs Group Ib	*orfA* + *orfB*	$\pi_{Ia} = 0.004$; $\pi_{Ib} = 0.015$	0.026	8 (5S + 3N)	6	0.634	<0.001
Group Ia vs Group IIa	*orfA* + *orfB*	$\pi_{Ia} = 0.004$; $\pi_{IIa} = 0.005$	0.028	24 (19S + 5N)	0	0.839	<0.001
Group Ib vs Group IIa	*orfA* + *orfB*	$\pi_{Ib} = 0.015$; $\pi_{IIa} = 0.005$	0.031	16 (12S + 4N)	1	0.659	<0.001
Group II vs Group IIa	*orfA* + *orfB*	$\pi_{II} = 0.018$; $\pi_{IIa} = 0.005$	0.029	4 (2S + 2N)	3	0.593	<0.001
All groups	*orfA* + *orfB*	NA	NA	NA	NA	0.648	<0.001

[a] See footnotes to Table 4.

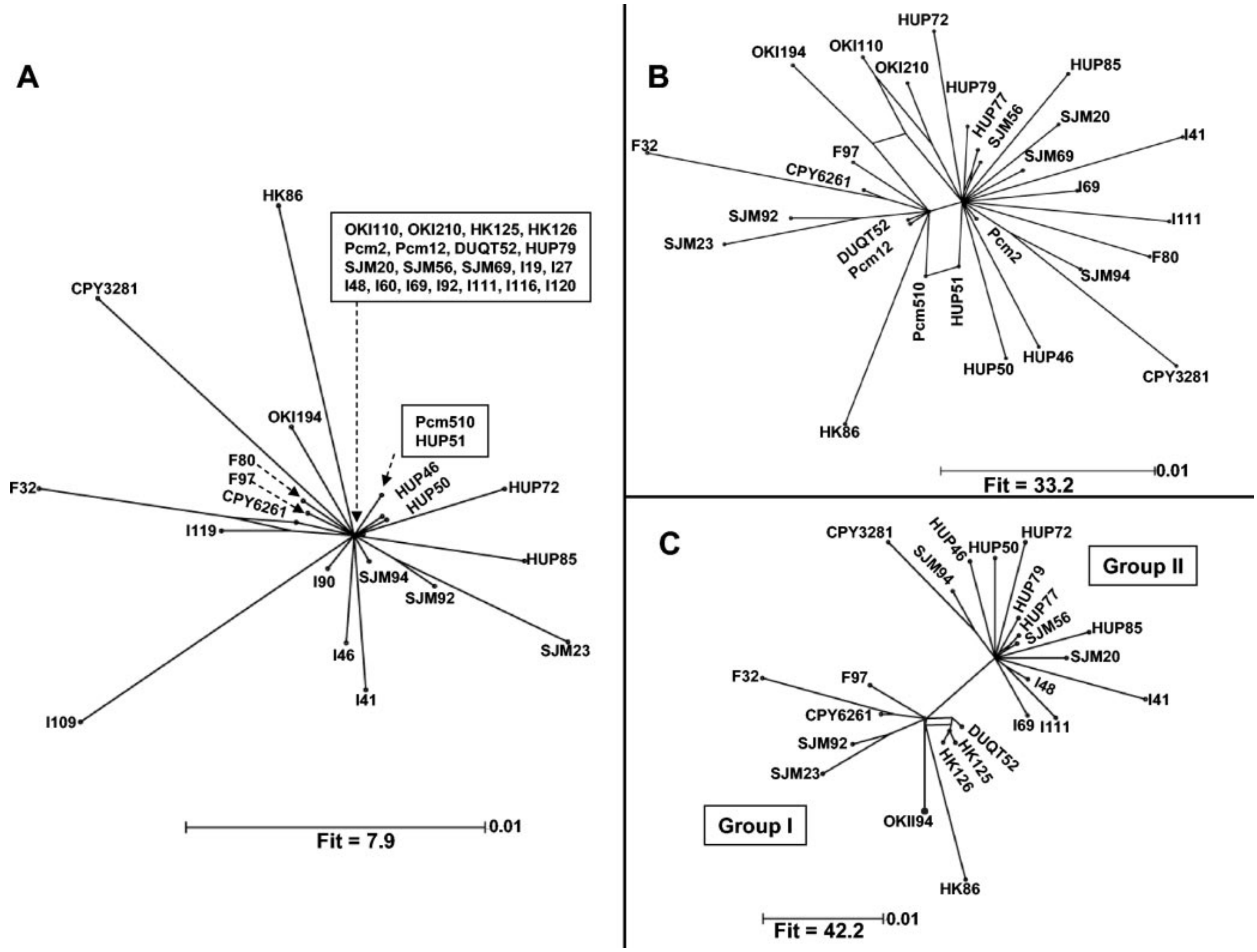

FIG. 3. Split decomposition analysis of IS*605*. (A) Annotated splits graph of the concatenated 927-bp IS*605 orfA* and *orfB* sequence generated with a pairwise TIM + I + Γ distance matrix. A fit parameter of 7.9 indicated the virtual absence of a tree-like structure. (B) Recalculation of the splits graph after removal of IS*605* sequences from clusters IndA and IndB improved the fit parameter and resolved networks linking the East Asian and Indo-European elements. (C) Splits graph of IS*605* showing population subdivision within IS*605*, which was masked by a history of recombinational exchanges between the two lineages. All branch lengths are drawn to scale.

tion structure was highly improbable by chance alone under a neutral model. The F_{ST} value was 0.65, which suggested very great genetic differentiation (27) between IS*607* subpopulations.

Pairwise comparisons of IS*607* groups Ia, Ib, II, and IIa revealed a negative correlation between fixed and shared polymorphisms (Table 5), which indicated that these subpopulations had diverged from a common ancestor. Since group IIa had only three sequences, a detailed analysis of polymorphisms is presented here for only groups Ia, Ib, and II. Group Ia IS*607* shared six polymorphisms with group Ib and six polymorphisms with group II, and groups Ib and II shared 18 polymorphisms (Fig. 2B). Furthermore, group Ia IS*607* showed only 11 unique polymorphisms, in contrast to the 58 and 89 unique polymorphisms in groups Ib and group II, respectively. These patterns suggest that there was strong genetic drift from few founders (founder effect) or a population bottleneck in group Ia IS*607* rather than a long time of separation from other IS*607* populations. A long time of separation would have allowed steady accumulation of unique polymorphisms and greater nucleotide diversity in group Ia, similar to the diversity seen in group Ib and group II (Table 5). Only one fixed difference between group Ib and group II with correspondingly large numbers of shared polymorphisms was identified. This suggests either that group Ib experienced a recent increase in population size or that the two groups diverged recently from their common ancestor.

Homologous recombination among *H. pylori* IS elements. The impact of recombination on IS*605* and IS*607* population structure was analyzed by using split decomposition. The splits graph of the complete IS*605* data set showed that the majority of sequences were linked to a single multifurcating node (star phylogeny) (Fig. 3A), which reaffirmed the lack of the sharp population subdivision seen in the IS*605* ML tree. Distant outgroups often show large random fluctuations and also different systematic biases (i.e., deviations between a parameter and parameter estimate due incorrect assumptions in the estimation method), with a tendency to hide information on in-group systematic bias (62). In such situations the splits graph

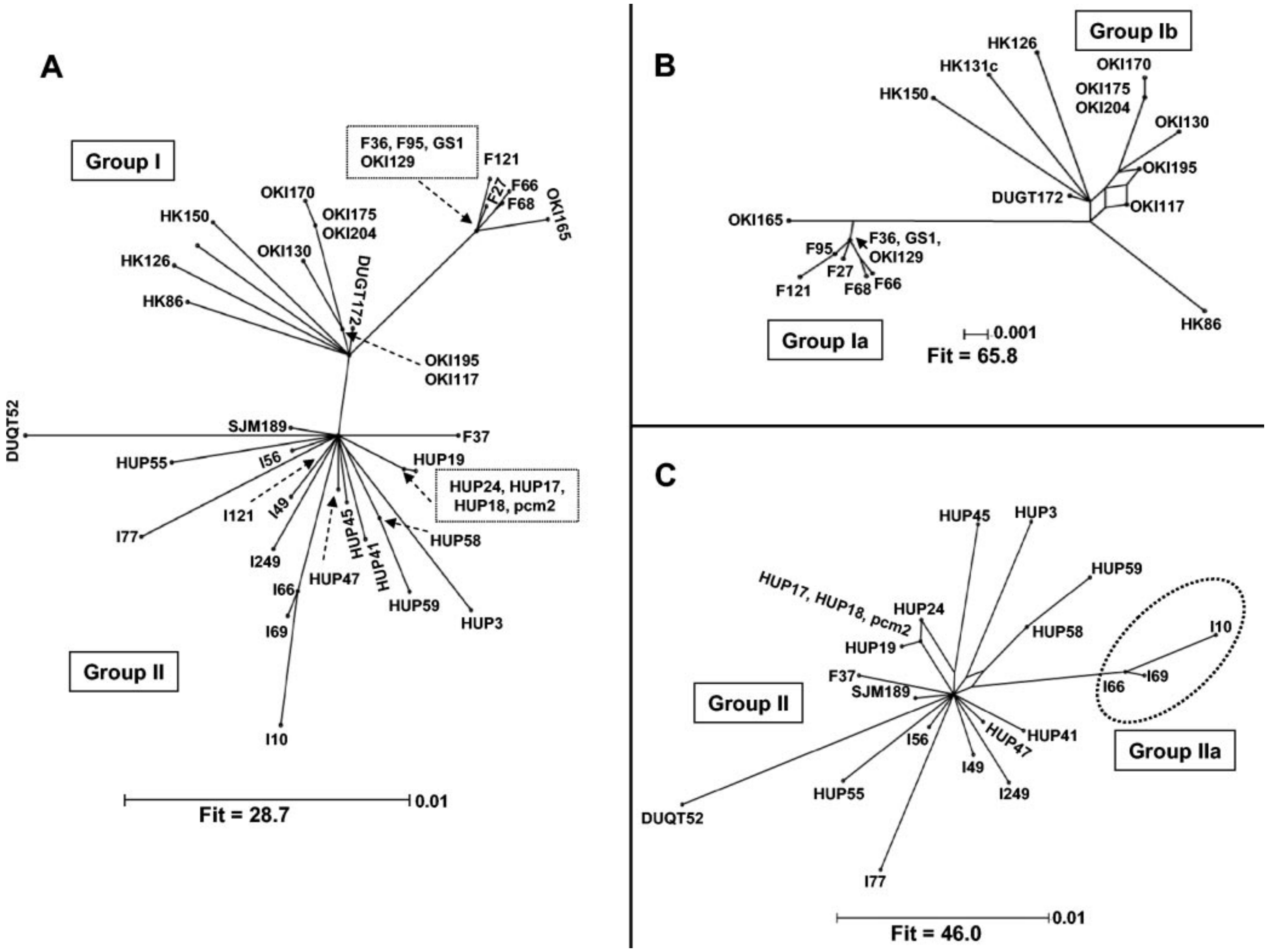

FIG. 4. Split decomposition analysis of IS*607*. (A) Annotated splits graph of the 1,100-bp concatenated sequence from IS*607 orfA* and *orfB* generated with a pairwise TVM + I + Γ distance matrix. A fit parameter of 28.7 indicated that all conflicts in the data set could not be resolved. Removal of distant outgroups in the data set improved the fit parameter (indicating better resolution of conflicts), as shown in the annotated splits graph for group I IS*607* (B) and the annotated splits graph for group II IS*607* (C). All branch lengths are drawn to scale.

loses resolution, as indicated by a low fit parameter (a fit parameter of 100 indicates that all conflicts in the data set are fully resolved). With the IS*605* splits graph a low fit parameter was obtained (fit parameter, 7.9), indicating that very few conflicts (multiple pathways linking sequences) in the data set could be resolved. Recalculation of the splits graph after removal of cluster IndA and IndB IS*605* sequences resulted in an improved fit parameter and revealed networks among IS*605* sequences (Fig. 3B). The multiple pathways linking East Asian and Spanish-Peruvian-Indian IS*605* sequences suggest that recombination had blurred the phylogenetic distinction between lineages. This suggestion was strengthened by removing sequences forming the network linking East Asian and Indo-European sequences in the splits graph. This revealed distinct East Asian and Indo-European IS*605* subpopulations (Fig. 3C).

In the case of IS*607*, the splits graph showed the clear phylogenetic distinctness of the East Asian elements from the elements in other regions but provided no evidence for networked evolution (Fig. 4A). Because the great length of the branch linking the two groups could have obscured networked evolution within groups, we also analyzed separately sequences from the two major lineages corresponding to groups I and II. In both groups few sequences formed interconnected networks, suggesting that there was a history of limited recombination (Fig. 4B and C).

Homoplasy ratio calculations complemented the outcomes from the SD analysis (Table 6). The complete concatenated IS*607* data set yielded an extremely low *h* value (0.2), indicating a lack of gene flow between IS*607* subpopulations. Even within IS*607* subpopulations, only low to moderate levels of recombination were detected. In contrast, the *h* values for IS*605* were in greater accord with values observed for other *H. pylori* chromosomal genes, which typically range from 0.6 to 1 (1, 61). Furthermore, for the complete data set, intragenic recombination seemed to be relatively more frequent within IS*605 orfA* (*h* = 0.53) than within *orfB (h* = 0.32). Consistently low levels of recombination were detected in both IS*607 orfA* (*h* = 0.21) and *orfB* (*h* = 0.201). These outcomes suggest that recombination between divergent lineages has been more common in the evolution of IS*605* than in the evolution of IS*607*.

TABLE 6. Homologous recombination within IS elements

IS element	No. of sequences	No. of variable sites	No. of informative sites	Mean shuffled *h*	*h* ratio[a]
IS*607AB*	44	61	40	89.2	0.217
Group Ia	7	ND[b]	ND	ND	ND
Group Ib	11	26	14	17.7	0.348
Group II	19	59	39	119.2	0.525
IS*605AB*[c]	42	67	53	261.9	0.558
GroupI	10	41	18	19.2	0.654
GroupII	14	54	39	90	0.701

[a] The *h* ratio is the average of *h* ratio assuming high, low, and medium expression.

[b] ND, not determined due to insufficient number of synonymous variable sites.

[c] IS*605* group designations are based on the split decomposition analysis shown in Fig. 3C.

Natural selection and evolution of *H. pylori* IS elements. We assessed the relative contributions of selective forces on the evolution of *H. pylori* IS elements by measuring site-specific d_N/d_S ratios using the ML approach implemented in the CODEML program of PAML, version 3.13d (68).

(i) IS*605*. The IS*605 orfA* gene encodes a 142-amino-acid transposase that is not related to the IS*607* transposase. In the 126 codons analyzed here, the average ω ranged from 0.31 to 0.47, implying that purifying selection dominated the evolution of this transposase (Table 7). Models that allowed for positive selection (M2, M3, and M8) all suggested the presence of positively selected sites. M3 suggested more sites than M2 or M8 because the parameter estimates with this model identified an additional 12.4% of the sites that were under relatively weak positive (diversifying) selection ($\omega_1 = 1.12$). Since the ML approach explicitly assumes a phylogenetic tree, to test whether high levels of recombination in IS*605 orfA* ($h = 0.56$ to 0.7) could have produced a false-positive signal for positive selection, we repeated the analysis assuming a star phylogeny. In a star phylogeny lineages diverge simultaneously from a single root node, thereby removing the effect of phylogenetic history. This analysis again produced significant evidence for positive selection ($P < 0.0001$ for M3 versus M0, M1, and M2 and for M8 versus M7), and in general the same sites fell into the positively selected class that they fell into in the original analysis (data not shown). This indicated that recombination did not adversely affect the ability of the ML method to detect positive selection in the IS*605* phylogeny. Taken together, these data provide evidence for adaptive evolution in IS*605* transposase.

In 183 of the 427 IS*605 orfB* codons analyzed, the average ω for all sites ranged from 0.12 to 0.15 for all models except M1, which, although better than M0, was rejected in favor of all other models that allowed for sites under positive selection (Table 8). M3 fit the data better than M0, M1, or M2, and parameter estimates from M3 suggested that ~90% of the sites were under functional constraints (purifying selection) with an ω_0 of 0.071, ~8% of the sites were moderately conserved with an ω_1 of 0.47 (which also implies purifying selection), and ~1.5% of the sites were under positive selection with an ω_2 of 2.613. Both M3 and M8 identified sites 100 and 183 with a *P* value of >99%.

(ii) IS*607*. In 179 of the 217 *orfA* codons analyzed, the average ω ranged from 0.2 to 0.25, indicating that the transposase was under purifying selection (Table 9). The more complex models (M1, M2, and M3) were significantly better than M0, thereby suggesting that selective pressures (indicated by ω) varied among sites. The best-supported M1 (neutral) model estimated that ~75% of the codons analyzed were subject to strong purifying selection (with ω_0 fixed at 0), while the rest of the codons underwent neutral evolution ($\omega_1 = 1$). These data suggest that the IS*607* transposase evolution was dominated by purifying and neutral selection, in contrast to the positive selection evident in the unrelated IS*605* transposase.

In 189 of the 419 IS*607 orfB* codons analyzed, the average ω ranged from 0.21 to 0.27 among the best-fitting models (Table 10). Parameter estimates and LRTs suggested the presence of sites under positive selection. For example, the discrete model (M3), which fit the data significantly better than M0, M1, or M2, estimated that ~10% of the codons analyzed were under positive selection with an ω_2 of 1.665. Similarly, parameter estimates from M8, which fit the data better than M7, also suggested that ~10% of the *orfB* codons analyzed were under positive selection with an ω_2 of 1.65. The failure of M2 to detect positive selection in biological data sets has been noted previously and is apparently due to the fact that the M1 model, on which M2 is based, is unrealistic and does not optimally

TABLE 7. Selection pressures acting on IS*605 orfA*: ML parameter estimates

Model	lnL[a]	d_N/d_S[b]	Estimates of parameters	LRT (df)[c]	χ^2	*P*	Positively selected sites ($\omega > 1$)[d]
M0 (one ratio)	−1,716.813	0.263	$\omega = 0.263$				None
M1 (neutral)	−1,679.783	0.336	$p_0 = 0.663, \omega_0 = 0; p_1 = 0.336, \omega_1 = 1$				Not allowed
M2 (selection)	−1,663.452	0.475	$p_0 = 0.656; p_1 = 0.321;$ $(p_2 = 0.021), \omega_2 = 7.066$	M0 vs M2 (2) M1 vs M2 (2)	74.06 32.66	<0.000 <0.000	**6E, 19P**, 104R
M3 (discrete)	−1,649.582	0.318	$p_0 = 0.856, \omega_0 = 0.073;$ $p_1 = 0.124, \omega_1 = 1.125;$ $p_2 = 0.019, \omega_2 = 5.842$	M0 vs M3 (4) M1 vs M3 (4) M2 vs M3 (2)	134.46 60.40 27.74	<0.000 <0.000 <0.000	**6E, 10N**, **18I**, **19P**, **48H**, **50I**, **62N**, **63F**, **92S**, **104R**, **105D**, 59T, 3K, 21N
M7 (β)	−1,667.806	0.212	$p = 0.134, q = 0.497$				Not allowed
M8 (β and ω)	−1,654.705	0.302	$p_0 = 0.972; p = 0.210,$ $q = 0.940; p_2 = 0.027,$ $\omega_2 = 4.767$	M7 vs M8 (2)	26.20	<0.000	**6E**, **19P**, 104R

[a] InL, log-likelihood score. Tree lengths and kappa estimates were homogeneous for all models.

[b] d_N/d_S averaged over all lineages and all sites.

[c] Likelihood ratio test (degrees of freedom).

[d] Site probability in boldface type, >95%; site probability underlined, >75% and <95%; site probability not in boldface type and not underlined, >50% and <75%.

TABLE 8. Selection pressures acting on IS*605 orfB*: ML parameter estimates[a]

Model	lnL	d_N/d_S	Estimates of parameters	LRT (df)	χ^2	*P*	Positively selected sites ($\omega > 1$)
M0 (one ratio)	−1,947.005	0.141	$\omega = 0.141$				None
M1 (neutral)	−1,942.164	0.27	$p_0 = 0.729$, $\omega_0 = 0$; $p_1 = 0.270$, $\omega_1 = 1$				Not allowed
M2 (selection)	−1,923.208	0.125	$p_0 = 0.0$, $p_1 = 0.054$; ($p_2 = 0.945$), $\omega_2 = 0.075$	M0 vs M2 (2) M1 vs M2 (2)	47.59 37.91	<0.000 <0.000	None
M3 (discrete)	−1,920.305	0.145	$p_0 = 0.90$, $\omega_0 = 0.071$; $p_1 = 0.084$, $\omega_1 = 0.474$; $p_2 = 0.015$, $\omega_2 = 2.613$	M0 vs M3 (4) M1 vs M3 (4) M2 vs M3 (2)	53.40 43.71 5.80	<0.000 <0.000 <0.05	**100D**, **183T**
M7 (β)	−1,927.993	0.1531	$p = 0.202$, $q = 1.096$				Not allowed
M8 (β and ω)	−1,920.445	0.145	$p_0 = 0.982$; $p = 0.869$, $q = 7.280$; $p_2 = 0.017$, $\omega_2 = 2.511$	M7 vs M8 (2)	15.09	<0.000	**100D**, **183T**

[a] See Table 7 footnotes.

account both for sites for which ω is >0 but <1 and for sites for which ω is >1 (68).

DISCUSSION

Population structures of IS*605* and IS*607*. Analysis of IS*605* and IS*607* sequences indicated that most East Asian elements were phylogenetically distinct from Indo-European elements. Statistical tests for population subdivision suggest that the differences resulted from evolution in geographic isolation. The nucleotide sequences of normal chromosomal (housekeeping) genes of East Asian and Indo-European *H. pylori* strains also form distinct phylogenetic clusters (1, 22, 23, 36, 46, 52). Thus, the IS*605* and IS*607* population structures seemed to be similar to those of their *H. pylori* hosts. This was most striking in the case of IS*607*, in which there was little recombination between subpopulations (Fig. 2 and 4A and Table 6). In the case of IS*605*, geographic isolation between East Asian and Indo-European elements (Fig. 3C and Table 4) was blurred by recombination between subpopulations (Fig. 3B and Table 6). Similarities in the IS and *H. pylori* population structures suggest parallel evolutionary histories and ancient host-element associations. Differences in the detailed population genetic structure of IS*605* and IS*607* are discussed below.

Population differences in IS element alleles might be evolutionarily stable, reflecting long-term IS residence in the *H. pylori* gene pool, or might reflect transient polymorphisms that arose recently as elements spread among lineages and that remain to be eliminated. The levels of nucleotide divergence at synonymous sites, which are generally more uniform than those at nonsynonymous sites (27), can be used to test these alternatives because changes accumulated at synonymous sites reflect the ages of alleles (4, 66). The cumulative (all populations) mean divergence values at synonymous sites (corrected for multiple hits) for the IS*607 orfA* and *orfB* and *IS605 orfA* and *orfB* sequences were 7.8, 11.3, 15.3, and 12.8%, respectively (Table 3). These values resembled the mean divergence values at synonymous sites seen in several of *H. pylori*'s normal chromosomal genes, such as *ppa* (10.7%), *ureI* (12.3%), *atpA* (13.6%), *yphC* (16.4%), and *efp* (17.5%) (22). The levels of nonsynonymous site divergence, which is subject to greater selective constraints, were generally less than the synonymous site divergence levels for both IS and other chromosomal genes (Table 3). These data suggest that IS elements are ancient components of the *H. pylori* gene pool and evolved at approximately the same rate as normal chromosomal genes. Long-term IS residence in the *H. pylori* gene pool would provide ample opportunities for element adaptation to the host or coevolution.

Polymorphism levels within geographic subdivisions depend on population history, including bottlenecks that directly decrease effective population sizes (8, 27). Assuming that the IS elements each invaded the *H. pylori* gene pool prior to separation of populations and that only a fraction of the ancestral *H. pylori* strains also harbored each IS element, the effect of random genetic drift, which reduces the net nucleotide diversity, would have been greater on IS elements than on normal

TABLE 9. Selection pressures acting on IS*607 orfA*: ML parameter estimates[a]

Model	lnL	d_N/d_S	Estimates of parameters	LRT (df)	χ^2	*P*	Positively selected sites ($\omega > 1$)
M0 (one ratio)	−1,393.247	0.207	$\omega = 0.207$				None
M1 (neutral)	−1,385.371	0.246	$p_0 = 0.753$, $\omega_0 = 0$; $p_1 = 0.246$, $\omega_1 = 1$				Not allowed
M2 (selection)	−1,384.631	0.206	$p_0 = 0.673$, $p_1 = 0.119$; ($p_2 = 0.207$), $\omega_2 = 0.41$	M0 vs M2 (2) M1 vs M2 (2)	17.232 1.48	<0.000 NS[b]	None
M3 (discrete)	−1,384.77	0.208	$p_0 = 0.115$, $\omega_0 = 0.0001$; $p_1 = 0.62$, $\omega_1 = 0.01$; $p_2 = 0.26$, $\omega_2 = 0.76$	M0 vs M3 (4) M1 vs M3 (4) M2 vs M3 (2)	16.954 1.202 −0.278	<0.001 NS NS	**None**
M7 (β)	−1,384.701	0.208	$p = 0.086$, $q = 0.329$				Not allowed
M8 (β and ω)	−1,384.483	0.217	$p_0 = 0.98$; $p = 0.144$, $q = 0.638$; $p_2 = 0.011$, $\omega_2 = 3.059$	M7 vs M8 (2)	0.138	NS	**None**

[a] See Table 7 footnotes.
[b] NS, not significant.

TABLE 10. Selection pressures acting on IS607 *orfB*: ML parameter estimates[a]

Model	lnL	d_N/d_S	Estimates of parameters	LRT (df)	χ^2	*P*	Positively selected sites ($\omega > 1$)[a]
M0 (one ratio)	−1,903.608	0.224	$\omega = 0.224$				None
M1 (neutral)	−1,866.1	0.239	$p_0 = 0.760, \omega_0 = 0;$ $p_1 = 0.239, \omega_1 = 1$				Not allowed
M2 (selection)	−1,864.976	0.203	$p_0 = 0.0, p_1 = 0.190;$ $(p_2 = 0.809), \omega_2 = 0.023$	M0 vs M2 (2) M1 vs M2 (2)	77.264 2.248	<0.000 NS[b]	173G
M3 (discrete)	−1,860.354	0.245	$p_0 = 0.74, \omega_0 = 0.01;$ $p_1 = 0.146, \omega_1 = 0.4;$ $p_2 = 0.104, \omega_2 = 1.665$	M0 vs M3 (4) M1 vs M3 (4) M2 vs M3 (2)	86.508 11.492 10.694	<0.000 <0.02 <0.01	**155N**, **168V**, **173G**, **174V**, **180K**, **186Q**, 50T, 51N, 53L, 94R, 148L, 179Y, 29A, 178E
M7 (β)	−1,865.261	0.217	$p = 0.041, q = 0.158$				Not allowed
M8 (β and ω)	−1,862.362	0.245	$p_0 = 0.893; p = 0.158,$ $q = 1.755; p_2 = 0.106,$ $\omega_2 = 1.65$	M7 vs M8 (2)	5.798	<0.05	**155N**, **168V**, **173G**, **174V**, **180K**, **186Q**, 50T, 51N, 53L, 94R, 148L, 179Y, 29A, 178E

[a] See Table 7 footnotes.
[b] NS, not significant.

H. pylori genes, because of smaller IS population sizes. Additionally, constraints on lateral movement of these elements (e.g., due to *H. pylori*'s preferential intrafamilial transmission) or possible selection against IS element carriage in certain lineages would increase drift by keeping the fraction of IS element-containing strains small (8). Analysis of IS*607* subpopulations suggested that there was particularly strong drift among group Ia elements (Fig. 2a); the nucleotide diversity in this group ($\pi = 0.4\%$) was ~5- to 10-fold less than that in normal chromosomal genes from Japanese strains (π ranged from 1.5 to 2.8% for six chromosomal genes [Dailide and Berg, unpublished data] but see GenBank accession numbers AY152917 to AY152929, AY152970 to AY152986, AY153290 to AY153309, AY153217 to AY153238, AY153368 to AY153388, and AY153164 to AY153177). In contrast, the nucleotide diversity in IS*607* group II ($\pi = 1.8\%$) was ~1.3- to 3-fold less than that in Indo-European strains (π ranged from 2.4 to 4.6% for six chromosomal genes [Dailide and Berg, unpublished] but see GenBank accession numbers AY152881 to AY152893, AY152930 to AY152941, AY153261 to AY153289, AY153204 to AY153216, AY153338 to AY1533367, and AY153151 to AY153263). The Indo-European *H. pylori* population is far more complex than the Japanese population; further sampling of IS*607* and *H. pylori* populations would help us judge better how reduced the nucleotide diversity in group II is and how much of the diversity stems from genetic drift.

In contrast to IS*607*, the nucleotide diversity within East Asian (group I $\pi = 3.3\%$) and Indo-European (group II $\pi = 3.4\%$) IS*605* elements (Table 4) was similar to that in normal chromosomal genes from corresponding populations. However, the data also suggested greater recombination among IS*605* subpopulations than the recombination observed with IS*607* (Fig. 3 and Table 6). Recombination between different subpopulations can increase the net nucleotide divergence within a given subpopulation (13) and can also reduce genetic differentiation between populations (12). This is seen in the F_{ST} values for IS*605* populations ($F_{ST} = 0.281$; PM test *P* values, <0.01) relative to those of IS*607* ($F_{ST} = 0.648$; PM test *P* values, <0.001) (Tables 4 and 5).

In most studies of IS element population genetics the workers have focused on elements from model enteric bacterial species (*E. coli*, *Salmonella*) and have used strains from North American or European collections. We anticipate, however, that additional sampling of Asian enteric strains would reveal little, if any, geographic clustering. The strong geographic clustering of IS element sequences seen in *H. pylori* is ascribed to the different population genetic structures and biology of the bacterial host species (*H. pylori* versus *E. coli* and *Salmonella*).

Strains of *E. coli* containing any one IS element were far more likely to contain additional, unrelated elements than would be expected by chance (26). This suggested that enteric IS elements were often transferred on high-capacity plasmid vectors. In contrast, only 9 (6%) of the 157 *H. pylori* isolates of the 488 isolates that we screened that carried either IS*607* or IS*605* carried both elements. This predominance of strains carrying only single elements would be explained if such high-capacity mobilizable plasmid vectors are uncommon in *H. pylori* populations.

Natural selection and IS element evolution. We sought evidence for positive selection in IS genes based on the premise that IS element adaptation to its host should accelerate the accumulation of new divergent alleles (4, 66). ML analysis of ratios of nonsynonymous to synonymous nucleotide changes provided a measure of selective pressures acting along a protein-encoding sequence, and the results suggested that the IS*605* and IS*607* genes had been subjected to different selective pressures at different sites (Tables 7 to 10). Purifying selection dominated at most codon sites of IS elements, but positive (diversifying) selection was evident at specific codons in IS*605 orfA* (12.9% of the codons with an ω_1 of 1.125 and 1.9% of the codons with an ω_2 of 5.842), in IS*605 orfB* (1.5% of the codons with an ω_2 of 2.613), and in IS*607 orfB* (10% of the codons with an ω_2 of 1.665). In contrast, only purifying selection and neutral selection were detected in IS*607 orfA* (ω_0 was 0 in 75% of the codons, and ω_1 was 1 in ~25% of the codons); no positive selection was detected.

The finding that there was positive selection in ~15% of IS*605 orfA* (transposase) codons but not in IS*607 orfA* suggests that the unrelated transposases evolved under different selective regimens. We also detected positive selection ($\omega_2 = 1.63$) at ~10% of the codons in the transposase of IS*Hp608* (data not shown), which is related to the transposase of IS*605* but not to the transposase of IS*607* (37). Several bacterial IS incorporate host proteins in their transposition complexes (10, 15); given the genetic variability in natural *H. pylori* populations, we suggest that IS*605* and IS*Hp608* transposases adapted to interact with more than one accessory factor (or common factor with polymorphic features), whose abundance varies geographically. This proposed difference in selective forces acting on the unrelated transposases of IS*605* and IS*607* is consistent with the distinct insertion specificities (34, 35) and the possibility of fundamentally different mechanisms of transposition. Differences in selective pressures, recombination frequency, and random genetic drift (see above) together suggest that IS*605* and IS*607* have distinct evolutionary dynamics.

The dominance of purifying selection seen here, which is also evident in eukaryotic mobile elements (18, 40, 65), is consistent with a model in which transposition causes IS element proliferation in the host gene pool (38). Bacterial IS elements tend to be more active when they are introduced into a naïve cell (free of the homologous IS element) than after they are established (3, 10, 57). Counteracting the short-term selective advantage of any element that transposes frequently would be host defenses that minimize potentially deleterious consequences of transposition (usually gene inactivation or genome rearrangements) plus negative regulating mechanisms of the element itself. Several different mechanisms for regulating transposition have been identified in various elements (3, 10, 15).

Mutational studies have shown that wild-type transposases themselves are less than maximally active and that certain amino acid substitutions can increase their transposition activity (10, 17, 57). Accordingly, an alternate hypothesis for the observed positive selection involves selection for optimal (but not necessarily maximal) IS*605* transposition activity in different host backgrounds. Changes that reduce transposition of established elements would often be selected because they decrease deleterious effects of transposition on host fitness (see above). Compensatory mutations that increase transposition might be selected at other times (e.g., under conditions that favor spread of the element to other genetic lineages or, in cases of transposition, that contribute to host fitness by increasing genome plasticity) (45, 48, 53). Much as in other cases of evolution in rugged landscapes (7), independent histories of mutations that cause decreased and then compensatory increased transposition activities would result in different evolutionary trajectories and manifest as positive selection (i.e., for amino acid change) at specific sites, as observed here.

Because IS *orfB* genes are not needed for transposition, at least in an *E. coli* model, selection (purifying and positive) of IS *orfB* genes suggests that the proteins that are encoded have other roles. The amino acid sequences of IS*605* and IS*607* OrfB proteins are ~25 to 30% identical to those of the *Salmonella* phage-encoded GipA protein, which promotes *Salmonella*'s growth and survival in the Peyer's patches of the intestine (60). A single C-terminal Zn(II) binding tetracysteine motif, $CX_2CX_{15-16}CX_2C$ (C4-type zinc finger), is well conserved among GipA and *H. pylori* IS OrfB proteins (60) and potentially would facilitate DNA or RNA binding or protein-protein interactions (42). It is thus appealing to consider the possibility that IS OrfB proteins may regulate chromosomal gene expression and/or perhaps IS element transposition and thereby contribute to *H. pylori* host adaptation, growth, and/or survival.

Concluding remarks. The present study of sequences from two representative IS elements of *H. pylori* provided a new perspective on the evolutionary dynamics of IS elements in bacterial populations, which was not evident from studies of the unrelated IS elements of enteric bacterial models (5, 26, 41). The distinctive geographic clustering of *H. pylori* IS element sequences likely reflects the extreme genetic diversity and the nonclonal, geographically partitioned population structure of *H. pylori* hosts and the probable paucity of readily transferable, high-capacity plasmid vectors in this gastric pathogen. Some of the patterns observed here may also result from features of these IS elements themselves (such as *orfB*) that distinguish them from the previously studied enteric elements. Long-term association of transposable elements with their hosts can result in coevolution (38, 43). Apparently parallel evolutionary histories of IS*605*, IS*607*, and *H. pylori* populations and adaptive evolution in IS genes suggest such coevolutionary dynamics. Understanding when and how IS*605* and IS*607* affect bacterial fitness and any reciprocal effects of host factors on the fitness of IS elements themselves (the give and take of any evolutionary process) may in turn provide new insights into *H. pylori*'s transmission, colonization, and virulence mechanisms and the evolution of its genome.

ACKNOWLEDGMENTS

We thank many colleagues for the *H. pylori* strains used in this study.

This study was supported by research grants RO1 AI38166, RO3 AI49161, RO1 AI63041, RO1 DK53727, RO1 DK057826, and P30 DK52574 from the National Institutes of Health.

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. **Bandelt, H. J., and A. W. Dress.** 1992. Split decomposition: a new and useful approach to phylogenetic analysis of distance data. Mol. Phylogenet. Evol. **1:**242–252.
3. **Berg, D. E., and M. M. Howe.** 1989. Mobile DNA. ASM Press, Washington, D.C.
4. **Bergelson, J., M. Kreitman, E. A. Stahl, and D. Tian.** 2001. Evolutionary dynamics of plant R-genes. Science **292:**2281–2285.
5. **Bisercic, M., and H. Ochman.** 1993. The ancestry of insertion sequences common to *Escherichia coli* and *Salmonella typhimurium*. J. Bacteriol. **175:** 7863–7868.
6. **Blaser, M. J., and D. E. Berg.** 2001. *Helicobacter pylori* genetic diversity and risk of human disease. J. Clin. Investig. **107:**767–773.
7. **Burch, C. L., and L. Chao.** 1999. Evolution by small steps and rugged landscapes in the RNA virus phi6. Genetics **151:**921–927.
8. **Cavalli-Sfroza, L. L., and W. F. Bodmer.** 1999. The genetics of human populations. Dover Publications Inc., New York, N.Y.
9. **Censini, S., C. Lange, Z. Xiang, J. E. Crabtree, P. Ghiara, M. Borodovsky, R. Rappuoli, and A. Covacci.** 1996. *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. Proc. Natl. Acad. Sci. USA **93:**14648–14653.
10. **Chandler, M., and J. Mahillon.** 2002. Insertion sequences revisited, p. 305–367. *In* N. C. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz (ed.), Mobile DNA II. ASM Press, Washington, D.C.
11. **Chao, L., C. Vargas, B. B. Spear, and E. C. Cox.** 1983. Transposable elements as mutator genes in evolution. Nature **303:**633–635.
12. **Cherry, J. L.** 2004. Selection, subdivision, and extinction and recolonization. Genetics **166:**1105–1114.
13. **Cohan, F. M.** 2002. Sexual isolation and speciation in bacteria. Genetica **116:**359–370.
14. **Cover, T. L., D. E. Berg, M. J. Blaser, and H. L. T. Mobely.** 2001. *Helicobacter pylori* pathogenesis, p. 509–558. *In* E. A. Groisman (ed.), Principles of bacterial pathogenesis. Academic Press, New York, N.Y.
15. **Craig, N. C., R. Craigie, M. Gellert, and A. M. Lambowitz (ed.).** 2002. Mobile DNA II. ASM Press, Washington, D.C.
16. **Curcio, M. J., and K. M. Derbyshire.** 2003. The outs and ins of transposition: from mu to kangaroo. Nat. Rev. Mol. Cell Biol. **4:**865–877.
17. **Derbyshire, K. M., and N. D. Grindley.** 1996. *cis* preference of the IS903 transposase is mediated by a combination of transposase instability and inefficient translation. Mol. Microbiol. **21:**1261–1272.

18. **Doak, T. G., D. J. Witherspoon, C. L. Jahn, and G. Herrick.** 2003. Selection on the genes of *Euplotes crassus Tec1* and *Tec2* transposons: evolutionary appearance of a programmed frameshift in a Tec2 gene encoding a tyrosine family site-specific recombinase. Eukaryot. Cell **2:**95–102.

19. **Doolittle, W. F., and C. Sapienza.** 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature **284:**601–603.

20. **Dubois, A., D. E. Berg, E. T. Incecik, N. Fiala, L. M. Heman-Ackah, J. Del Valle, M. Yang, H. P. Wirth, G. I. Perez-Perez, and M. J. Blaser.** 1999. Host specificity of *Helicobacter pylori* strains and host responses in experimentally challenged nonhuman primates. Gastroenterology **116:**90–96.

21. **Falush, D., C. Kraft, N. S. Taylor, P. Correa, J. G. Fox, M. Achtman, and S. Suerbaum.** 2002. Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. Proc. Natl. Acad. Sci. USA **98:**15056–15061.

22. **Falush, D., T. Wirth, B. Linz, J. K. Pritchard, M. Stephens, M. Kidd, M. J. Blaser, D. Y. Graham, S. Vacher, G. I. Perez-Perez, Y. Yamaoka, F. Megraud, K. Otto, U. Reichard, E. Katzowitsch, X. Wang, M. Achtman, and S. Suerbaum.** 2003. Traces of human migrations in *Helicobacter pylori* populations. Science **299:**1582–1585.

23. **Ghose, C., G. I. Perez-Perez, M. G. Dominguez-Bello, D. T. Pride, C. M. Bravi, and M. J. Blaser.** 2002. East Asian genotypes of *Helicobacter pylori* strains in Amerindians provide evidence for its ancient human carriage. Proc. Natl. Acad. Sci. USA **99:**15107–15111.

24. **Guedon, G., F. Bourgoin, M. Pebay, Y. Roussel, C. Colmin, J. M. Simonet, and B. Decaris.** 1995. Characterization and distribution of two insertion sequences, IS1191 and iso-IS981, in *Streptococcus thermophilus*: does intergeneric transfer of insertion sequences occur in lactic acid bacteria cocultures? Mol. Microbiol. **16:**69–78.

25. **Hartl, D. L., D. E. Dykhuizen, R. D. Miller, L. Green, and J. de Framond.** 1983. Transposable element IS50 improves growth rate of *E. coli* cells without transposition. Cell **35:**503–510.

26. **Hartl, D. L., and S. A. Sawyer.** 1988. Why do unrelated insertion sequences occur together in the genome of *Escherichia coli*. Genetics **118:**537–541.

27. **Hartl, D. L., and A. G. Clark.** 1997. Principles of population genetics. Sinauer Associates Inc., Sunderland, Mass.

28. **Hey, J.** 1991. The structure of genealogies and the distribution of fixed differences between DNA sequence samples from natural populations. Genetics **128:**831–840.

29. **Hook-Nikanne, J., D. E. Berg, R. M. Peek, Jr., D. Kersulyte, M. K. Tummuru, and M. J. Blaser.** 1998. DNA sequence conservation and diversity in transposable element IS*605* of *Helicobacter pylori*. Helicobacter **3:**79–85.

30. **Hudson, R. R., D. D. Boos, and N. L. Kaplan.** 1992. A statistical test for detecting geographic subdivision. Mol. Biol. Evol. **9:**138–151.

31. **Hudson, R. R.** 2000. A new statistic for detecting genetic differentiation. Genetics **155:**2011–2014.

32. **Huelsenbeck, J. P., and B. Rannala.** 1997. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. Science **276:**227–232.

33. **Huson, D. H.** 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics **14:**68–73.

34. **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*: chimaeric transposable elements of *Helicobacter pylori*. Gene **223:**175–186.

35. **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.

36. **Kersulyte, D., A. K. Mukhopadhyay, B. Velapatino, W. Su, Z. 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.

37. **Kersulyte, D., B. Velapatino, G. Dailide, A. K. Mukhopadhyay, Y. Ito, L. Cahuayme, A. J. Parkinson, R. H. Gilman, and D. E. Berg.** 2002. Transposable element IS*hp608* of *Helicobacter pylori*: nonrandom geographic distribution, functional organization, and insertion specificity. J. Bacteriol. **184:**992–1002.

38. **Kidwell, M. G., and D. R. Lisch.** 2002. Perspective: transposable elements, parasitic DNA, and genome evolution. Evol. Int. J. Org. Evol. **55:**1–24.

39. **Kiss, J., M. Szabo, and F. Olasz.** 2003. Site-specific recombination by the DDE family member mobile element IS30 transposase. Proc. Natl. Acad. Sci. USA **100:**15000–15005.

40. **Lampe, D. J., D. J. Witherspoon, F. N. Soto-Adames, and H. M. Robertson.** 2003. Recent horizontal transfer of mellifera subfamily mariner transposons into insect lineages representing four different orders shows that selection acts only during horizontal transfer. Mol. Biol. Evol. **20:**554–562.

41. **Lawrence, J. G., H. Ochman, and D. L. Hartl.** 1992. The evolution of insertion sequences within enteric bacteria. Genetics **131:**9–20.

42. **Leon, O., and M. Roth.** 2000. Zinc fingers: DNA binding and protein-protein interactions. Biol. Res. **33:**21–30.

43. **Levin, B. R.** 1993. The accessory genetic elements of bacteria: existence conditions and (co)evolution. Curr. Opin. Genet. Dev. **3:**849–854.

44. **Maynard Smith, J., and N. H. Smith.** 1998. Detecting recombination from gene trees. Mol. Biol. Evol. **15:**590–599.

45. **Modi, R. I., L. H. Castilla, S. Puskas-Rozsa, R. B. Helling, and J. Adams.** 1992. Genetic changes accompanying increased fitness in evolving populations of *Escherichia coli*. Genetics **130:**241–249.

46. **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.

47. **Mukhopadhyay, A. K., J. Y. Jeong, D. Dailidiene, P. S. Hoffman, and D. E. Berg.** 2003. The *fdxA* ferredoxin gene can down-regulate *frxA* nitroreductase gene expression and is essential in many strains of *Helicobacter pylori*. J. Bacteriol. **185:**2927–2935.

48. **Naas, T., M. Blot, W. M. Fitch, and W. Arber.** 1995. Dynamics of IS-related genetic rearrangements in resting *Escherichia coli* K-12. Mol. Biol. Evol. **12:**198–207.

49. **Page, R. D. M.** 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. **12:**357–358.

50. **Page, R. D. M., and E. C. Holmes.** 1998. Molecular evolution: a phylogenetic approach. Blackwell Science Inc., Malden, Mass.

51. **Palys, T., L. K. Nakamura, and F. M. Cohan.** 1997. Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. Int. J. Syst. Bacteriol. **47:**1145–1156.

52. **Pan, Z. J., D. E. Berg, R. W. van der Hulst, W. W. Su, A. Raudonikiene, S. D. Xiao, J. Dankert, G. N. Tytgat, and A. van der Ende.** 1998. Prevalence of vacuolating cytotoxin production and distribution of distinct *vacA* alleles in *Helicobacter pylori* from China. J. Infect. Dis. **178:**220–226.

53. **Papadopoulos, D., D. Schneider, J. Meier-Eiss, W. Arber, R. E. Lenski, and M. Blot.** 1999. Genomic evolution during a 10,000-generation experiment with bacteria. Proc. Natl. Acad. Sci. USA **96:**3807–3812.

54. **Picardeau, M., T. J. Bull, and V. Vincent.** 1997. Identification and characterization of IS-like elements in *Mycobacterium gordonae*. FEMS Microbiol. Lett. **154:**95–102.

55. **Posada, D., and K. A. Crandall.** 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14:**817–818.

56. **Reynolds, A. E., J. Felton, and A. Wright.** 1981. Insertion of DNA activates the cryptic *bgl* operon in *E. coli* K12. Nature **293:**625–629.

57. **Reznikoff, W. S.** 2003. Tn5 as a model for understanding DNA transposition. Mol. Microbiol. **47:**1199–1206.

58. **Rozas, J., and R. Rozas.** 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:**174–175.

59. **Smith, M. C., and H. M. Thorpe.** 2002. Diversity in the serine recombinases. Mol. Microbiol. **44:**299–307.

60. **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.

61. **Suerbaum, S., J. M. Smith, K. Bapumia, G. Morelli, N. H. Smith, E. Kunstmann, I. Dyrek, and M. Achtman.** 1998. Free recombination in *Helicobacter pylori*. Proc. Natl. Acad. Sci. USA **95:**12619–12624.

62. **Swofford, D. L., G. J. Olsen, P. J. Wadell, and D. M. Hillis.** 1996. Phylogenetic inference, p. 407–514. *In* D. M. Hillis, C. Moritc, and B. K. Mable (ed.), Molecular systematics. Sinauer Associates Inc., Sunderland, Mass.

63. **Swofford, D. L.** 2003. PAUP. Phylogenetic analysis using parsimony (and other methods), version 4. Sinauer Associates, Sunderland, Mass.

64. **Wakeley, J., and J. Hey.** 1997. Estimating ancestral population parameters. Genetics **145:**847–855.

65. **Witherspoon, D. J.** 1999. Selective constraints on P element evolution. Mol. Biol. Evol. **16:**472–478.

66. **Woolhouse, M. E., J. P. Webster, E. Domingo, B. Charlesworth, and B. R. Levin.** 2003. Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nat. Genet. **32:**569–577.

67. **Yang, Z.** 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13:**555–556.

68. **Yang, Z., R. Nielsen, N. Goldman, and A. M. Pedersen.** 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155:**431–449.