

Allelic Diversity among *Helicobacter pylori* Outer Membrane Protein Genes *homB* and *homA* Generated by Recombination[∇]

Mónica Oleastro,^{1,2*} Rita Cordeiro,¹ Armelle Ménard,^{2,3} and João Paulo Gomes¹

Department of Infectious Diseases, National Institute of Health, Lisbon, Portugal¹; INSERM U853, 33076 Bordeaux, France²; and Laboratoire de Bactériologie, Université Victor Segalen Bordeaux 2, Bordeaux, France³

Received 7 April 2010/Accepted 22 May 2010

Recombination is one of the main mechanisms contributing to *Helicobacter pylori* genomic variability. *homB* and *homA* are paralogous genes coding for *H. pylori* outer membrane proteins (OMPs). Both genes display allelic variation yielded by polymorphisms of the genes' middle regions, with six different alleles. This study used bioinformatic and statistical analyses to evaluate whether the allelic diversity of *homB* and *homA* is generated by recombination. A detailed molecular analysis of the most prevalent *homB* allelic variant was also performed to establish its molecular profile. The two most prevalent *homB* and *homA* allelic variants resulted from interallelic homologous recombination between the rarest allelic variants of each gene, with a crossover point localized in the middle of the genes, containing the allelic region. Molecular analysis of the most prevalent *homB* allele revealed a geographic partition among Western and East Asian strains, more noticeable for the 5' and 3' *homB* regions than for the middle allelic regions. In conclusion, the diversity of the 5' and 3' *homB* regions reflect the strains' geographical origin, and variants likely occur via the accumulation of single nucleotide polymorphisms. On the other hand, homologous recombination seems to play an important role in the diversification of the highly polymorphic *homB* and *homA* allele-defining regions, where the most prevalent alleles worldwide result from genomic exchange between the rarest variants of each gene, suggesting that the resulting combinations confer biological advantages to *H. pylori*. This phenomenon illustrates an evolutionary scenario in which recombination appears to be associated with ecological success.

The Gram-negative bacterium *Helicobacter pylori* colonizes the stomachs of more than half of the world's population, and this infection is associated with diverse gastroduodenal diseases (6, 16, 27). The risk of developing severe gastric diseases is influenced by *H. pylori* strain-dependent factors, among which are several outer membrane proteins (OMPs) that have been identified as virulence factors such as the *babA/babB*, *hopZ*, *hopQ*, *sabA*, *oipA*, and *homB* proteins (8, 17, 24, 28, 31, 48).

The *homB* gene and its highly similar (90% similarity at the nucleotide level) paralogous gene, *homA*, are members of the *H. pylori* OMP gene family (2). Both genes can be present in the *H. pylori* genome at two conserved loci, as a single copy or as two copies (*homA/homA*, *homB/homB*, *homA/homB*, or *homB/homA*) (28, 29). The *homB* gene has been proposed as a *H. pylori* virulence candidate, since its presence was correlated with severe gastroduodenal disease (19, 30) and with corpus inflammation and atrophy, as demonstrated by pathological analyses (19). Moreover, a putative role as an adhesin was suggested for the corresponding protein HomB, since *homB* knockout mutant strains showed significantly reduced binding (28). HomB was also shown to be antigenic and implicated in the *in vitro* activation of interleukin-8 secretion (28).

In contrast, the *homA* gene has been correlated with nonulcer gastritis (28, 30), and histological analyses did not show any association between the presence of the *homA* gene and in-

flammation or atrophy (19). The corresponding protein was also shown to be antigenic in humans (29).

Similar to other *H. pylori* OMP-encoding genes, such as *babA/babB*, *hopM/hopN* and *hopQ* (8, 20, 32), *homB* and *homA* genes display allelic variation which, in this case, was shown to occur in 300-bp regions localized in the middle of these genes, with *homB* displaying greater allelic diversity than *homA* (29). This highly polymorphic region, spanning from approximately 750 to 1,050 bp in *homB* ($\approx 2,007$ bp) and from 720 to 980 bp in *homA* ($\approx 1,980$ bp), was identified by the analysis of the similarity plot of each gene (29). Based on these plots, three segments were thus defined for those genes, where segments 1, 2, and 3 correspond to the regions preceding, matching, and following the allelic zone, respectively (see Fig. 2C) (29). Moreover, detailed sequence analyses, at nucleotide and amino acid levels, of *homB* and *homA* segment 2 were also performed, revealing the existence of six distinct and well-conserved allelic variants denominated AI to AVI. Five distinct alleles were observed for *homB* (AI, AII, AIII, AV, and AVI), while three allelic variants were observed for the *homA* gene (AII, AIII, and AIV). The variants AI, AV, and AVI were exclusively observed in *homB*, whereas AIV was present only in *homA*. Each gene displayed a predominant worldwide allelic variant (AI and AII for *homB* and *homA*, respectively), which was present in up to 80% of the clinical strains (29).

H. pylori displays one of the highest levels of genomic variability known among bacteria, and the high frequency of recombination is one of the main mechanisms contributing to that diversity (11, 39–41). Intragenomic recombination involving some of the OMP paralogous genes was also known to take place during infection, probably reflecting the selective pres-

* Corresponding author. Mailing address: Department of Infectious Diseases, National Institute of Health, Av. Padre Cruz, 1649-016 Lisbon, Portugal. Phone: (351) 217519231. Fax: (351) 217526400. E-mail: monica.oleastro@insa.min-saude.pt.

[∇] Published ahead of print on 4 June 2010.

sure for adhesion, which may differ across different hosts as well as within an individual over time (5, 20, 38).

Accordingly, the present study aimed to identify recombination phenomena that may have occurred within and between *homB* and *homA* allelic variants, which may underlie the genetic diversity and generation of alleles. Furthermore, considering that *homB* presents a worldwide dispersion (*homA* is rarely found among East Asian strains) and that the AI allele is the predominant *homB* variant worldwide (29), a detailed molecular analysis of this allele was performed in order to establish its molecular profile.

MATERIALS AND METHODS

***homB* and *homA* sequences.** The 107 *homB* sequences and 30 *homA* sequences used in the present study belong to 92 *H. pylori* strains isolated from patients from seven Western countries (Brazil, Colombia, France, Germany, Portugal, Sweden, and the United States) and from two East Asian countries (Japan and Korea), as previously reported (28, 29). These sequences are available under the GenBank accession numbers EF648320 to EF648325, EF648331 to EF648354, EF648374 to EF648379, EU363366 to EU363460, and EU910189 to EU910194 (28, 29).

Analyses of recombination events. A search for potential mosaic structures within the *homA* and *homB* genes was performed using the bioinformatic application SimPlot 3.5.1 (<http://sray.med.som.jhmi.edu/SCRSoftware>) to study the similarity of one putative recombinant sequence (query) to the other sequences. This software calculates and plots the nucleotide similarity of a query sequence against a panel of sequences under study, in a sliding window fashion, along the alignment. A sliding window size of 200 bp moved across the alignment using a step size of 20 bp. Nucleotide pairwise distances were calculated using the Jukes-Cantor parameter method (18). SimPlot analysis was performed for the 20 possible combinations involving the 6 *homB/homA* allelic variants (alleles AI to AVI). Whenever a putative recombination event was detected, a bootscan analysis (36) was also performed in which the likely recombinant sequence was compared with sequences from the probable parental strain(s) and a known outgroup sequence, generating four-member trees. The bootscan analysis is also a sliding window-based analysis that combines phylogeny and bootstrapping. At each nucleotide window range, a phylogenetic analysis was performed using neighbor-joining topology on the basis of pairwise genetic distances (Jukes-Cantor parameter method). For more accurate phylogenies, the software pairwise deletion option was chosen in order to remove all sites containing alignment gaps for each pair of sequences under comparison and not on an overall basis. Bootstrap confidence levels were determined by 100 replicates. Significant changes in phylogenetic relationships from window to window resulted in changes in bootscan values that were indicative of probable recombination events. SimPlot was also used to identify informative sites (34) where two sequences share one specific nucleotide but the other two share a different nucleotide. Each informative site supports one of the three possible phylogenetic relationships generated by bootscanning among the four taxa. The likelihood that the observed distribution of sites favoring specific phylogenetic groupings might occur randomly was assessed using the maximum χ^2 test. The most likely crossover region occurred where the observed distribution is less likely to occur randomly (maximum χ^2 value). The null hypothesis for recombination was that informative sites are drawn from a single distribution.

A *P* value for any specified crossover was determined by Fisher's exact test, using contingency tables. These included the number of the informative sites shared by the recombinant allele and each of the two putative parental strains, upstream and downstream of the crossovers, and also the lengths of these two regions.

To support the recombination regions detected by SimPlot and to accurately infer the phylogenetic relationships of the putative recombinant strains, the defined crossover regions were used to create two separate alignments, each corresponding to the left or right region of the crossover, and phylogenetic trees were inferred by neighbor-joining (Kimura two-parameter) analysis (22).

Ten different pairs of parental strains and recombinant sequences were tested for *homB* and *homA*, in order to confirm the crossover regions.

Molecular analysis of the *homB* AI allele. A total of 90 *homB* AI allele sequences, 70 from Western *H. pylori* strains and 20 from East Asian *H. pylori* strains, were used for molecular analysis of the *homB* AI allele. The nucleotide variability within *homB* allelic variants and within the *homB* AI allele was inferred by the determination of the variable sites (a site that contains at least two

types of nucleotides or amino acids), nonsynonymous (K_a) sites (a nucleotide site in which one or more changes are nonsynonymous), and parsimony-informative sites (a variable site occurring in at least two sequences), using MEGA 4.0.2 (<http://www.megasoftware.net>). By analyzing parsimony-informative sites rather than variable sites, a conservative approach was used in order to avoid biased bioinformatic analyses, because some sequences presented a high number of nucleotide changes that were specific to a single strain (likely due to the limited number of sequences analyzed in the present study, mainly from the East Asian group).

To study the nucleotide variability within *homB* allelic variants, a consensus sequence for each of the five variants was constructed, based on 10 *homB* AI, 10 AII, 3 AIII, 1 AV, and 3 AVI allele sequences. The number of sequences is proportional to the prevalence of each allele (29), in order to create consensus sequences that could accurately represent the frequency of each allele and, thus, its variability.

In order to evaluate the differences within the East Asian and Western strains, dispersion plots were constructed based on the estimated pairwise distances of the 90 *homB* AI allele sequences (70 from Western strains and 20 from East Asian strains) against *homB* AI East Asian consensus and *homB* AI Western consensus sequences. All of the consensus sequences, based on the five allelic *homB* variants and on both the *homB* AI East Asian and Western groups, were generated by sequence alignment using MegAlign software (Lasergene; DNASTar, Madison, WI) (7).

Evolutionary parameters were determined using MEGA 4.0.2 for all of the *homB* AI allele sequences and for *homB* AI East Asian and Western consensus sequences. Molecular distances were estimated using the Kimura two-parameter method (22), while the overall mean values of synonymous (K_s) and nonsynonymous (K_a) substitution rates were determined using the Nei-Gojobori method (25). The type of selection operating at the amino acid level can be detected by comparing the synonymous and nonsynonymous substitution rates. Positive selection refers to selection in favor of nonsynonymous substitutions at the DNA level, where the evolutionary distance based on nonsynonymous substitutions is expected to be greater than that based on synonymous substitutions ($K_a > K_s$); purifying selection refers to selection against nonsynonymous substitutions, where the evolutionary distance based on synonymous substitutions is expected to be greater than the distance based on nonsynonymous substitutions ($K_a < K_s$). The codon-based one-tailed Z-test of selection (26) was used to evaluate the probability of rejection of the null hypothesis of evolution ($K_a = K_s$) in favor of the positive or the purifying selection. The level of significance was set at 5%.

Considering that the percentage of G+C is a good indicator of genetic exchange among species (1, 3), the G+C percentage was also determined using EditSeq software (Lasergene; DNASTar) (7).

RESULTS

Analyses of recombination events. The potential existence of recombination events underlying *homB* and *homA* allelic diversity was evaluated with recombination analysis tools which have been successfully applied to analysis of this type of recombination in several other pathogens (13–15, 36). In a window-based analysis, the putative recombination crossover lies in the position where the plots of similarity values cross each other (35), as can be observed in the similarity plots between each *homB* and *homA* recombinant sequence (query sequence) and the respective parental strains (Fig. 1A1 and B1, respectively). Upstream and downstream of this position, the query sequence shows similarity to distinct sequences, evidencing a mosaic structure. Thus, the potential mosaic structures for the 20 possible combinations of the six *homB/homA* allelic variants were evaluated by the SimPlot analysis, suggesting the existence of homologous recombination in the following two cases: (i) the predominant *homB* allele AI (78.9%) seems to result from recombination between *homB* allelic variants AV and AVI (Fig. 1A1), the two less frequent variants (1.4% and 4.2%, respectively), and (ii) the predominant *homA* AII allele (84.9%) seems to result from recombination between the less

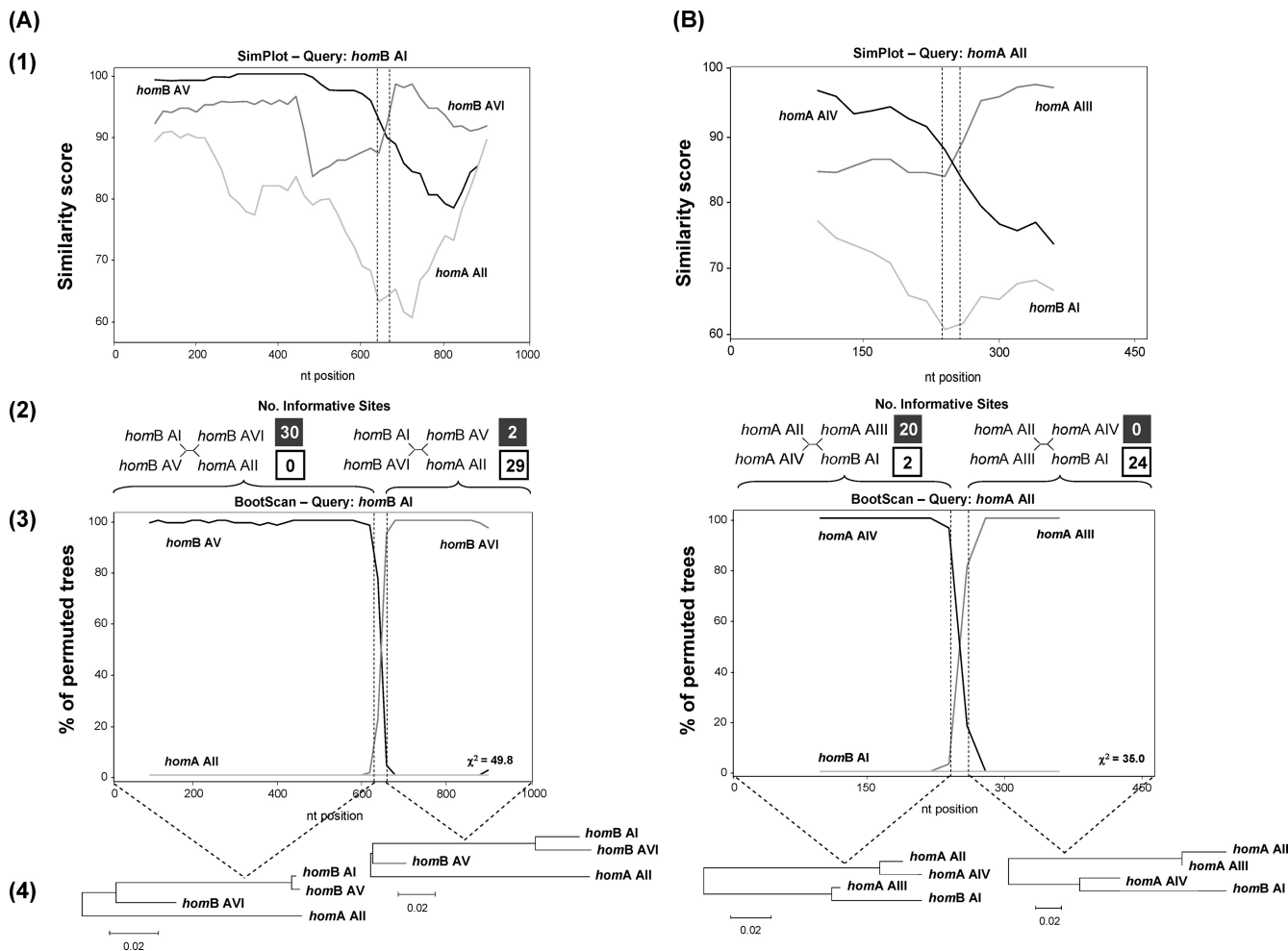


FIG. 1. SimPlot and bootscan analyses of the recombination events generating the two most prevalent *Helicobacter pylori* *homB AI* (A) and *homA AII* (B) allelic variants. (A) The *homB AV* allele parental strain and the *homB AVI* allele generate the recombinant (query) *homB AI* allele, and the *homA AII* allele was used as the outgroup. (B) The *homA AIV* allele parental strain and the *homA AIII* allele generate the recombinant (query) *homA AII* allele, and the *homB AI* allele was used as the outgroup. (A1, B1) Similarity plots between each recombinant sequence and the respective parental strains (window size, 200 bp; step size, 20 bp). (A2 and B2) Four-member trees representing the number of informative sites shared by the recombinant sequence and the parental strains. The dark boxes show the number of informative sites shared by the recombinant allele and the first parental strain (left), and the white boxes show the number of informative sites shared by the recombinant allele and the second parental strain (right). (A3, B3) Bootscan analyses (window size, 200 bp; step size, 20 bp) showing the phylogenetic relatedness (percentage of permuted trees) between the recombinant and the parental sequences. (A1 and 3, B1 and 3) The crossover regions are located between the vertical lines (represented in Fig. 2), and the nucleotide (nt) positions at the bottom of each plot correspond to alignment positions of the genomic regions. (A4, B4) Phylogenetic reconstructions for each specific region bounded by the crossover region supporting each putative crossover. The sequences used to represent the *homB AI*, *AV*, and *AVI* alleles and the *homA AII*, *AIII*, and *AIV* alleles are available under GenBank accession numbers EF648377, EU363409, EF648345, EU363372, EF648340, EU363449, EF648321, and EF648336.

frequent *homA* allelic variants AIV and AIII (3.8% and 11.3%, respectively) (Fig. 1B1) (29).

The bootscan analysis was used to statistically confirm the results obtained using the SimPlot analysis. In this analysis, “four-member” trees are generated, based on the similarity of the query (recombinant) sequence to the two probable parental strains, and a tree permutation is expected to occur whenever a recombination phenomenon takes place (35). A more detailed analysis using bootscan analysis confirmed the above-described recombination events (Fig. 1A3 and B3). In fact, for the *homB AI* allele, a putative crossover region was found, for which 30 informative sites upstream of that region support the similarity with the parental AV allele, whereas 29 informative

sites downstream of that region support the similarity with the parental AVI allele (Fig. 1A2). Similarly, for the *homA AII* allele, 20 informative sites upstream of the crossover region support the similarity with the parental AIV allele, and 24 informative sites downstream of that region support the similarity with the parental AIII allele (Fig. 1B2). These results were confirmed by the two incongruent phylogenetic reconstructions presented in Fig. 1A4 and B4, which show that the clustering of the *homB* and *homA* recombinant alleles alternates between the two putative parental alleles, respectively. The crossover regions of these putative homologous recombination events involve gene segments of 16 and 36 bp for *homB* and *homA*, respectively (Fig. 2A and B) and are located in the

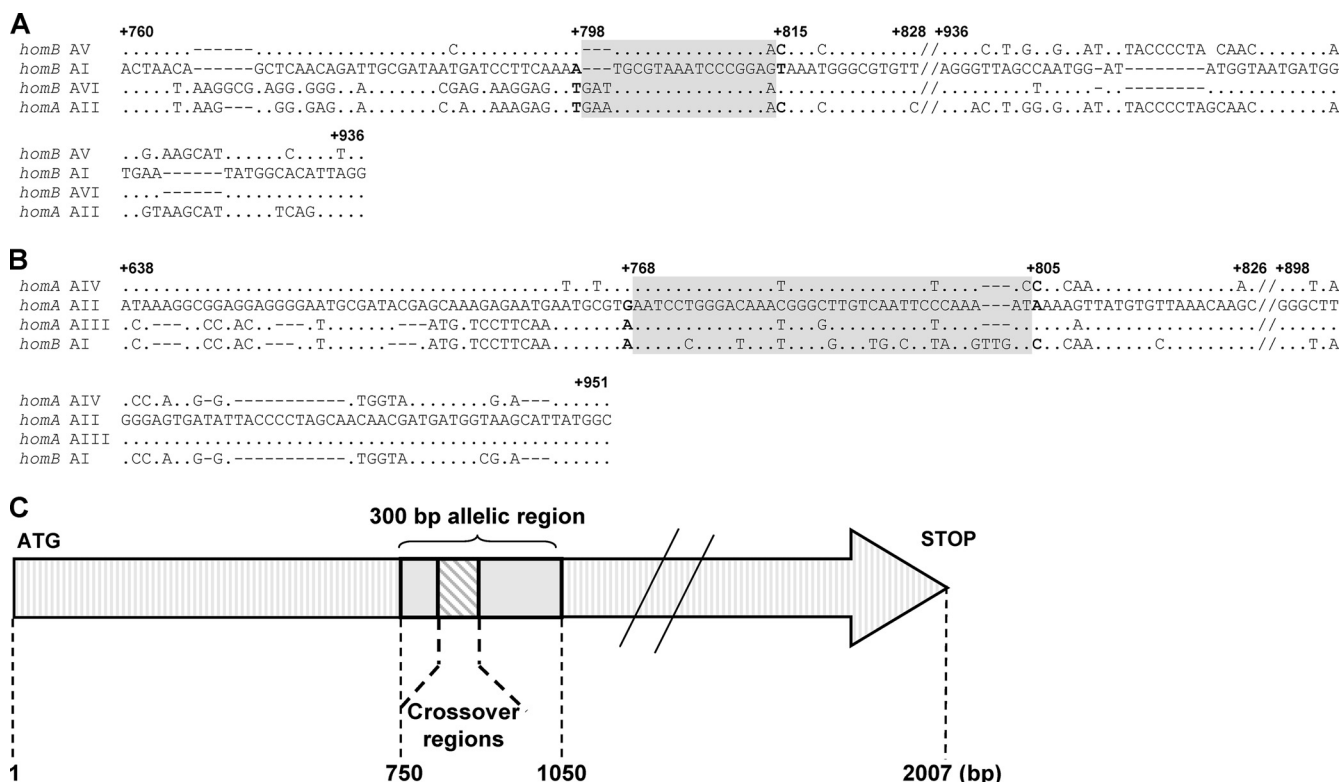


FIG. 2. (A, B) Nucleotide sequences of crossovers for *Helicobacter pylori* *homB* (A) and *homA* (B) genes. Numbers represent positions relative to the start codon of *homB* or *homA*, dots represent positions where nucleotides match between the recombinant and the parental sequences, and hyphens represent deletions. Crossover regions are highlighted in gray, bordered by two informative sites (in boldface), obtained from SimPlot/bootscan analyses. Between the informative sites, the recombinant allele displays no specific similarity with any of the parental alleles. (C) Localization of the crossover regions within the allelic regions for *homB* and *homA*.

middle of the genes within their 300-bp allelic regions (Fig. 2C). Within the crossover regions, no specific similarity was observed between the query and any parental strain (i.e., no informative sites were found). On the other hand, the difference between the numbers of informative sites shared by the recombinant allele and one parental strain with regard to the other, upstream and downstream of the crossover, was statistically significant, validating the two putative crossovers (Table 1).

Molecular analysis of the *homB* AI allele. About 40% of the nucleotide changes for the Western strains and 60% of those for the East Asian strains occur in a single strain (data not shown), which could lead to a biased analysis, since nucleotide changes spread among the population would be overestimated due to changes that are likely deleterious. Thus, a conservative

approach employing parsimony informative sites was used to estimate the number of genetic polymorphisms within *homB* (Fig. 3). When comparing the consensus sequences of each of the five alleles, it is interesting to note that 41.5% (88/212) of the polymorphisms are located in segment 2 of the gene. Furthermore, almost half (44.3%; 94/212) of all the polymorphisms observed are nonsynonymous. As expected, 48.9% (46/94) of these nonsynonymous polymorphisms are also concentrated in segment 2 of the *homB* gene (Fig. 3A). Contrary to the nucleotide interallelic variability, the nucleotide distribution within the *homB* AI allele seems to be random for both Western strains (Fig. 3B) and East Asian strains (Fig. 3C). Although the former presented a higher degree of polymorphism than the latter, this is likely due to the higher number of Western strains analyzed in the present study. The same pattern of random distribution occurs at the protein level, as shown by the location of the nonsynonymous nucleotide substitutions. This distribution pattern was confirmed by the molecular distance analysis of 90 *homB* AI allele sequences, presented in Table 2, where similar values for the whole gene and for the three gene segments were found. Furthermore, the evolutionary analysis of the *homB* AI allele revealed no significant differences among the three segments studied, neither with the rate of synonymous or nonsynonymous polymorphisms nor with the ratio of nonsynonymous to synonymous substitutions (Table 2). These results contrast with those ob-

TABLE 1. Statistical significance of the crossover regions in *H. pylori* *homB* AI and *homA* AII alleles

Region	Statistical significance of crossover regions ^a	
	<i>homB</i> AI	<i>homA</i> AII
Upstream of the crossover	1 × 10 ⁻⁷	1.9 × 10 ⁻⁶
Downstream of the crossover	1.6 × 10 ⁻⁴	1.4 × 10 ⁻⁶

^a Based on the *P* values calculated for the regions upstream and downstream of the two crossovers.

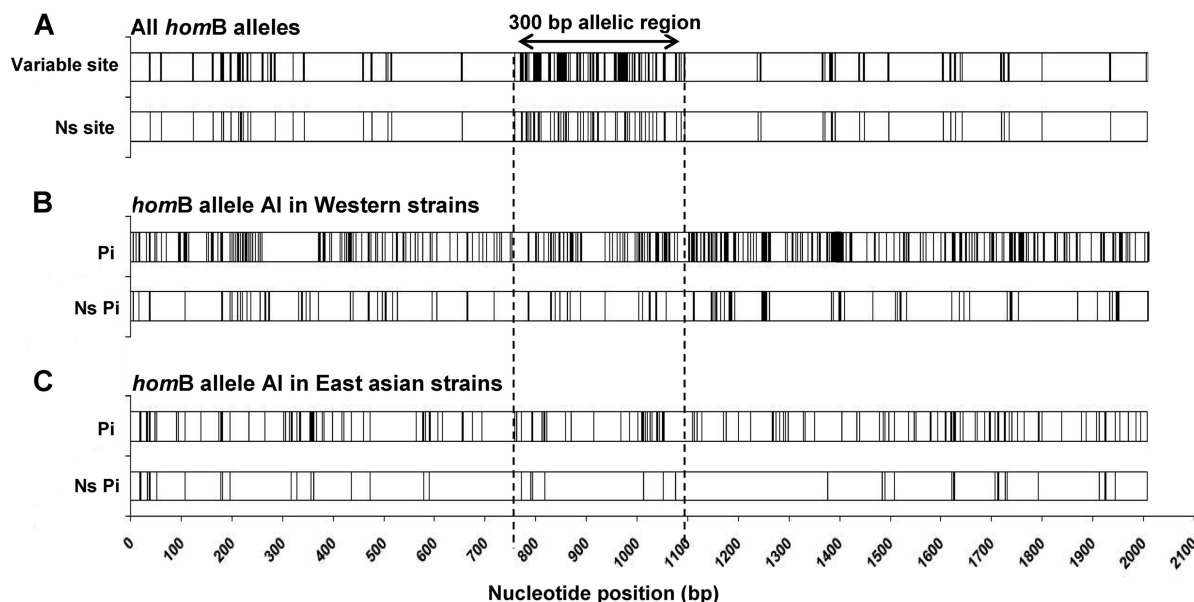


FIG. 3. Genetic polymorphisms within *Helicobacter pylori* *homB*. The polymorphism distribution was determined by comparing the consensus sequences of each of the five allelic variants at nucleotide (variable site) and protein (nonsynonymous site) levels (A), the *homB* AI allele in Western strains ($n = 70$) at nucleotide (parsimony-informative site) and protein (nonsynonymous parsimony-informative site) levels (B), and the *homB* AI allele in East Asian strains ($n = 20$) at nucleotide (parsimony-informative site) and protein (nonsynonymous parsimony-informative site) levels (C). Given that some sequences presented a high number of nucleotide changes that were exclusive of a single strain, parsimony-informative sites rather than variable sites were used. Ns, nonsynonymous; Pi, parsimony informative.

tained when analyzing *homB* sequences, including all five *homB* alleles, for which the molecular distance, the rate of nonsynonymous polymorphisms, and the ratio of nonsynonymous to synonymous substitutions were higher for segment 2 than for the entire gene or segments 1 and 3 (29). The G+C content revealed no differences for the entire gene (38.03% [standard error {SE, 0.34}]) or for segments 1, 2, and 3 (35.8% [SE, 0.494], 35.6% [SE, 1.283], and 40.3% [SE, 0.538], respectively), presenting values similar to that of the *H. pylori* J99 strain genome content ($\approx 38\%$) (3).

Interestingly, the evolutionary parameters calculated for two consensus sequences, *homB* AI allele sequences from East Asian and Western strains, showed that the ratio of nonsyn-

onymous to synonymous substitutions was approximately 8- to 14-fold lower for segment 2, the allele-defining region, than for segments 3 and 1, respectively (Table 2). The ratio of nonsynonymous to synonymous substitutions was 1.84 for *homB* segment 1 (Table 2), suggesting a positive selection, i.e., a selection toward amino acid change (45), but this hypothesis was not statistically supported by the codon-based Z-test of selection ($P_{Z\text{-test}} = 0.081$), thus a neutral evolution for this segment cannot be excluded. On the other hand, the null hypothesis ($K_a = K_s$) was rejected, with statistical support in favor of the purifying selection ($K_a < K_s$) for segment 2 ($P_{Z\text{-test}} < 0.001$). With regard to segment 3, the value obtained for the ratio of nonsynonymous to synonymous sub-

TABLE 2. Evolutionary parameters of the *H. pylori homB* AI allele^c

Sequence (n)	Molecular distance ^a	K_s rate	K_a rate	K_a/K_s ratio ^d
<i>homB</i> AI allele (90) ^b				
Entire gene	0.070 ± 0.003	0.194 ± 0.0100	0.028 ± 0.003	0.144 ± 0.017
Segment 1	0.064 ± 0.005	0.150 ± 0.013	0.029 ± 0.004	0.193 ± 0.031
Segment 2	0.052 ± 0.008	0.147 ± 0.028	0.024 ± 0.007	0.163 ± 0.057
Segment 3	0.071 ± 0.005	0.205 ± 0.014	0.031 ± 0.005	0.151 ± 0.026
Western and East Asian consensus				
Entire gene	0.063 ± 0.006	0.160 ± 0.017	0.032 ± 0.005	0.200 ± 0.038
Segment 1	0.037 ± 0.007	0.025 ± 0.012	0.046 ± 0.009	1.84 ± 0.954
Segment 2	0.084 ± 0.02	0.235 ± 0.059	0.031 ± 0.014	0.132 ± 0.068
Segment 3	0.062 ± 0.007	0.063 ± 0.017	0.064 ± 0.010	1.02 ± 0.317

^a Based on the Kimura two-parameter method.

^b Includes 70 sequences from Western *H. pylori* strains and 20 sequences from East Asian *H. pylori* strains.

^c Values ± SE.

^d Results in boldface highlights the K_a/K_s ratio values that diverge from one, suggesting that gene segments 1 and 2 are under selective pressure, although for segment 1, this hypothesis was not statistically significant, according to the codon-based Z-test of selection ($P_{Z\text{-test}} = 0.081$).

stitutions (1.02 ± 0.317) does not suggest any kind of pressure in this gene region.

The evaluation of the dispersion of the individual *homB* AI sequences against the East Asian and Western consensus sequences showed a clear separation between the East Asian and Western strains (Fig. 4A), which is consistent with the previously described phylogenetic reconstruction of *homB*, showing the existence of two predominant clusters corresponding to East Asian and Western countries (29). This dispersion was also noted for each gene segment, 1, 2, and 3 (Fig. 4B to D), although an intersection of the two clusters of strains was observed for segment 2, most likely reflecting a smaller influence of the geographic background of the strains in the evolution of this gene segment (Fig. 4C).

DISCUSSION

Homologous recombination is crucial for the long-term survival of bacterial cells being implicated in the evolution of genomes (14, 37, 43, 44) and, more specifically, in the evolution of paralogous genes such as OMP-encoding genes (2). Indeed, intragenomic recombination between members of the OMPs may represent a means of regulating protein expression for adaptation to host selective pressures. This mechanism was shown to occur in *H. pylori* *in vivo* for several OMPs, such as BabA/BabB and HopM/HopN (20, 38), allowing *H. pylori* to dynamically change the adherence and antigenic properties of OMPs in response to host immune and inflammatory pressures, as was previously demonstrated for BabA/BabB (38). The genes *homB* and *homA*, analyzed in the present study, seem to somehow mirror *babA* and *babB* at the genomic level, as they also code for OMPs and they are present in both a single- or two-copy fashion, exchanging positions between themselves in conserved loci (28, 29). Furthermore, as described for *babA/babB* and for other *H. pylori* OMP genes (*hopQ* and *hopM/hopN*) (8, 20, 33), allelic variation has been reported for *homB* and *homA* (29). It has been suggested that intragenomic recombination is the likely mechanism underlying the generation of allelic groups (20, 28, 29, 33).

In the present study, the analysis of the *homB/homA* interallelic recombination events showed that the rarest circulating alleles constitute the parental sequences of the most prevalent circulating alleles, a phenomenon observed for both *homB* (the AI allele results from recombination between AV and AVI) and *homA* (AV and AVI yield the AII allele). Assuming that there is no DNA sequence affecting homologous recombination frequency in this area, it is reasonable to speculate that the events promoting the recombination between the two rarest alleles, yielding the most prevalent one, confer a biological advantage to the *H. pylori* strain. This is supported by the fact that this phenomenon occurs in *H. pylori* strains from both East Asian and Western countries. In this context, if a role as an adhesin is hypothesized for HomB (28), then one can speculate that the protein coded by the *homB* AI allele is the best adapted to adhere to human gastric mucosa. On the other hand, and considering the antigenicity of HomB (28), it is likely that the HomB AI protein would be a better fit to mediate the host-bacteria interaction.

Considering that (i) both *homA* and *homB* can be present in one *H. pylori* genome and/or mixed infections can occur and

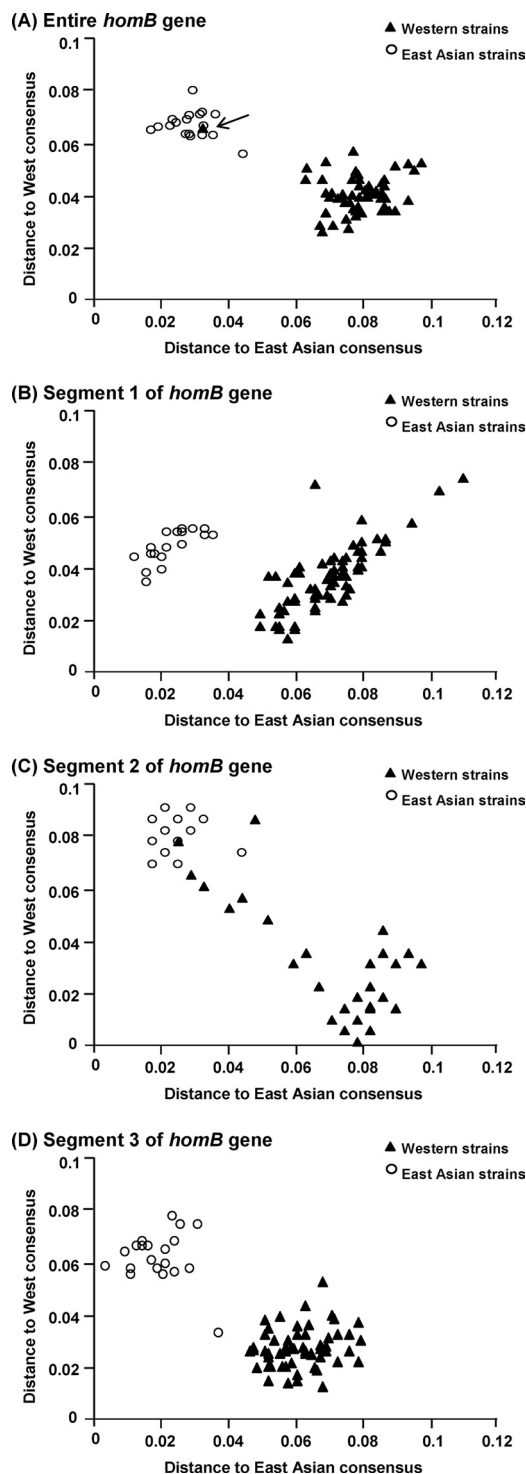


FIG. 4. Dispersion plots of the *Helicobacter pylori* *homB* AI allele individual sequences against the Western and East Asian AI allele consensus sequences. Dispersion plots were determined using the AI allele consensus sequences, based on the pairwise distances calculated for 70 *homB* sequences from *H. pylori* Western strains and 20 *homB* sequences from East Asian strains, considering the entire gene (A) and each of the gene segments, i.e., segments 1 (B), 2 (C), and 3 (D). (A) The arrow refers to an *H. pylori* clinical strain isolated from a U.S. citizen of Asian origin. (C) The geographical separation (East Asia versus West) was less noticeable for segment 2, most likely reflecting a smaller influence of the geographic background of the strains in the evolution of this gene segment.

that (ii) some alleles are gene specific, whereas some alleles are common to both genes, consequently a common allele may also be involved in the phenomenon of homologous recombination, although the exact parental origin of the allele cannot be determined (*homA* or *homB*). As an example, the predominant AII variant of the *homA* gene could also arise by recombination between the specific AIV allele of *homA* and the common AIII allele of *homB*. Therefore, and although the results point to recombination involving alleles present in the same gene (*homB* or *homA*), alleles located in different genes (*homB* and *homA*), whether belonging to the same strain or not, are also most likely involved in these recombination events, unless it can be demonstrated otherwise.

The molecular analysis of the *homB* AI allele showed that the polymorphic sites were randomly distributed throughout the gene for both East Asian and Western strains (Fig. 3B and C), while the interallelic variability, previously determined for all five *homB* allelic variants, was observed mostly in the middle region (segment 2) of the gene (Fig. 3A), which corresponds to the allele-defining region (29). Indeed, when sequences of all five *homB* allelic variants were analyzed, segment 2 of this gene displayed both the largest molecular distance and the highest nonsynonymous substitution rate, even compared to those of the entire gene; a sliding window analysis also showed that the nonsynonymous substitution rate in that region was about 5-fold higher than that in the rest of the gene (29). This is not the case for the *homB* AI allele, showing that within each allele, segment 2 is the most conserved segment (Table 2). Considering that all five "allelic" HomB proteins were shown to be antigenic, these results suggest that segment 2 is the segment most implicated in promoting antigenic variability (29). Thus, it would be interesting to evaluate the antigenicity of each of the three HomB segments, in order to clarify each one's role.

The geographical segregation observed for the *homB* AI variant has also been described for several *H. pylori* gene categories, such as housekeeping and virulence genes (1, 21, 23, 46). Moreover, as described for the two paralogous genes *babA/babB* (33), phylogenetic analysis of *homB* revealed that strains are grouped first by their geographic origin, with two main clusters being observed, the Western cluster and the East Asian cluster. Then, within each main cluster, a separation according to the allelic variant could also be detected (29).

Though the factors that lie beneath the geographic partitioning of *H. pylori* genes are not known, one can speculate that genetic factors underlying human population heterogeneity will select for divergence among *H. pylori* strains. This heterogeneity can include specificity and strength of immune and inflammatory responses, as well as availability and distribution of receptors to which *H. pylori* can adhere (10, 38). Previous results indicate the role of HomB in *H. pylori* adhesion (28); therefore, by analogy with the BabA adhesin for which functional adaptation to geographically predominant blood groups was reported (4), HomB likely is subject to that kind of pressure. Furthermore, several studies suggest that *H. pylori* polymorphisms reflect human phylogeography and historical migrations, constituting a reliable biological marker of host-pathogen coevolution (9, 12, 23, 42, 47) which is facilitated by the long-term contact between the infecting strain and the host. Thus, it is likely that the strains infecting humans from

different geographical regions display unique characteristics that reflect distinctive human population traits on behalf of the never-ending coevolutionary process that takes place during infection.

However, it is interesting to observe that, when solely analyzing the *homB* AI allele, the geographical separation is less noticeable for segment 2 than for segments 1 and 3 (Fig. 4). Consequently, the existence of an allelic region less dependent on the geographic origin of the strain, flanked by regions that show strong evidence of geographic segregation, indicates that segments 1 and 3 are more likely to be influenced by the host selective pressure. This is clearly supported by evolutionary parameters, as the ratio of nonsynonymous to synonymous substitutions obtained by comparing the East Asian and Western *homB* AI consensus sequences (Table 2) reveals 8- to 14-fold-higher values for segments 3 and 1, respectively, than for segment 2. Moreover, the evidence that the latter segment is the most polymorphic among all of the *homB* allelic variants (Fig. 3A) (29) but the most conserved within *homB* AI allele sequences, regardless of the strain's geographic origin (Table 2 and Fig. 4C), strongly supports its choice as the definition of the allelic variants. Lastly, the predominance of the *homB* AI allele among East Asian and Western strains (29) suggests that this allele may confer some biological advantages to *H. pylori* strains. To clarify this issue, further experiments are necessary to evaluate the binding ability and antigenicity strength of *H. pylori* strains according to their *homB* allelic variants.

In summary, by using bioinformatic analysis on sequences taken from human biopsy samples, this study presents statistically supported evidence that the allelic variability among the *H. pylori* OMP-encoding genes *homB* and *homA* is generated by homologous recombination events, in addition to the accumulation of point mutations. Finally, the fact that the most prevalent alleles worldwide seem to be chimeras of the rarest alleles suggests that the resulting genetic combinations are critical for the ecological success of *H. pylori* strains.

ACKNOWLEDGMENTS

We thank the Sociedade Portuguesa de Gastroenterologia and the Programme de Cooperation Scientifique et Technique Franco-Portugais, offered by the French Embassy in Portugal, for supporting this project. We thank the Fundação para a Ciência e Tecnologia, project PPCDT/SAL-IMI/57297/2004, for financial support.

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. Alm, R. A., J. Bina, B. M. Andrews, P. Doig, R. E. W. Hancock, and T. J. Trust. 2000. Comparative genomics within *Helicobacter pylori*: analysis of the outer membrane protein families. *Infect. Immun.* **68**:4155–4168.
3. Alm, R. A., L. S. L. Ling, D. T. Moir, B. L. King, E. D. Brown, P. C. Doig, D. R. Smith, B. Noonan, B. C. Guild, B. L. deJonge, G. Carmel, P. J. Tummino, A. Caruso, M. UriaNickelsen, D. M. Mills, C. Ives, R. Gibson, D. Merberg, S. D. Mills, Q. Jiang, D. E. Taylor, G. F. Vovis, and T. J. Trust. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **397**:176–180.
4. Aspholm-Hurtig, M., G. Dailide, M. Lahmann, A. Kalia, D. Ilver, N. Roche, S. Vikstrom, R. Sjostrom, S. Linden, A. Backstrom, C. Lundberg, A. Arnavqvist, J. Mahdavi, U. J. Nilsson, B. Velapatino, R. H. Gilman, M. Gerhard, T. Alarcon, M. Lopez Brea, T. Nakazawa, J. G. Fox, P. Correa, M. G. Dominguez Bello, G. I. Perez Perez, M. J. Blaser, S. Normark, I. Carlstedt, S. Oscarson, S. Teneberg, D. E. Berg, and T. Boren. 2004. Functional adaptation of BabA, the *H. pylori* ABO blood group antigen binding adhesin. *Science* **305**:519–522.

5. Backstrom, A., C. Lundberg, D. Kersulyte, D. E. Berg, T. Boren, and A. Arnqvist. 2004. Metastability of *Helicobacter pylori* bab adhesin genes and dynamics in Lewis b antigen binding. *Proc. Natl. Acad. Sci. U. S. A.* **101**:16923–16928.
6. Blaser, M. J. 1992. Hypotheses on the pathogenesis and natural history of *Helicobacter pylori*-induced inflammation. *Gastroenterology* **102**:720–727.
7. Burland, T. G. 2000. DNASTAR's Lasergene sequence analysis software. *Methods Mol. Biol.* **132**:71–91.
8. Cao, P., and T. L. Cover. 2002. Two different families of *hopQ* alleles in *Helicobacter pylori*. *J. Clin. Microbiol.* **40**:4504–4511.
9. Disotell, T. 2003. Discovering human history from stomach bacteria. *Genome Biol.* **4**:213.
10. 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.
11. Falush, D., C. Kraft, N. S. Taylor, P. Correa, J. G. Fox, M. Achtman, and S. Suerbaum. 2001. Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. *Proc. Natl. Acad. Sci. U. S. A.* **98**:15056–15061.
12. 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. Y. Wang, M. Achtman, and S. Suerbaum. 2003. Traces of human migrations in *Helicobacter pylori* populations. *Science* **299**:1582–1585.
13. Fang, G., B. Weiser, C. Kuiken, S. M. Philpott, S. Rowland-Jones, F. Plummer, J. Kimani, B. Shi, R. Kaul, J. Bwayo, O. Anzala, and H. Burger. 2004. Recombination following superinfection by HIV-1. *AIDS* **18**:153–159.
14. Gomes, J. P., W. J. Bruno, A. Nunes, N. Santos, C. Florindo, M. J. Borrego, and D. Dean. 2007. Evolution of *Chlamydia trachomatis* diversity occurs by widespread interstrain recombination involving hotspots. *Genome Res.* **17**:50–60.
15. He, C. Q., N. Z. Ding, J. G. Chen, and Y. L. Li. 2007. Evidence of natural recombination in classical swine fever virus. *Virus Res.* **126**:176–185.
16. IARC. 1994. Schistosomes, liver flukes and *Helicobacter pylori*. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7–14 June 1994. IARC Monogr. Eval. Carcinog. Risks Hum. **61**:1–241.
17. Ilver, D., A. Arnqvist, J. Ögren, I.-M. Frick, D. Kersulyte, E. T. Incecik, D. E. Berg, A. Covacci, L. Engstrand, and T. Boren. 1998. *Helicobacter pylori* adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science* **279**:373–377.
18. Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules, p. 21–132. In H. N. Munro (ed.), *Mammalian protein metabolism*. Academic Press, New York, NY.
19. Jung, S., M. Sugimoto, D. Graham, and Y. Yamaoka. 2009. *homB* status of *Helicobacter pylori* as a novel marker to distinguish gastric cancer from duodenal ulcer. *J. Clin. Microbiol.* **47**:3241–3245.
20. Kersulyte, D., H. Chalkauskas, and D. E. Berg. 1999. Emergence of recombinant strains of *Helicobacter pylori* during human infection. *Mol. Microbiol.* **31**:31–43.
21. Kersulyte, D., A. K. Mukhopadhyay, B. Velapatino, W. W. Su, Z. J. Pan, C. Garcia, V. Hernandez, Y. Valdez, R. S. Mistry, R. H. Gilman, Y. Yuan, H. Gao, T. Alarcon, M. Lopez Brea, G. B. Nair, A. Chowdhury, S. Datta, M. Shirai, T. Nakazawa, R. Ally, I. Segal, B. C. Y. 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.
22. Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**:111–120.
23. Linz, B., F. Balloux, Y. Moodley, A. Manica, H. Liu, P. Roumagnac, D. Falush, C. Stamer, F. Prugnolle, S. W. van der Merwe, Y. Yamaoka, D. Y. Graham, E. Perez-Trallero, T. Wadstrom, S. Suerbaum, and M. Achtman. 2007. An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* **445**:915–918.
24. Mahdavi, J., B. Sonden, M. Hurtig, F. O. Olfat, L. Forsberg, N. Roche, J. Angstrom, T. Larsson, S. Teneberg, K. A. Karlsson, S. Altraia, T. Wadström, D. Kersulyte, D. E. Berg, A. Dubois, C. Petersson, K. E. Magnusson, T. Norberg, F. Lindh, B. B. Lundskog, A. Arnqvist, L. Hammarstrom, and T. Boren. 2002. *Helicobacter pylori* SabA adhesin in persistent infection and chronic inflammation. *Science* **297**:573–578.
25. Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**:418–426.
26. Nei, M., and S. Kumar. 2000. Synonymous substitutions and nonsynonymous nucleotide substitutions, p. 52–61. In M. Nei (ed.), *Molecular evolution and phylogenetics*, vol. 1. Oxford University Press, New York, NY.
27. NIH Consensus Development Panel on *Helicobacter pylori* in Peptic Ulcer Disease. 1994. NIH Consensus Conference. *Helicobacter pylori* in peptic ulcer disease. *JAMA* **272**:65–69.
28. Oleastro, M., R. Cordeiro, J. Ferrand, B. Nunes, P. Lehours, I. Carvalho-Oliveira, A. I. Mendes, D. Penque, L. Monteiro, F. Mégraud, and A. Ménard. 2008. Evaluation of the clinical significance of *homB*, a novel candidate marker of *Helicobacter pylori* strains associated with peptic ulcer disease. *J. Infect. Dis.* **198**:1379–1387.
29. Oleastro, M., R. Cordeiro, A. Ménard, D. Queiroz, F. Mégraud, and L. Monteiro. 2009. Allelic diversity and phylogeny of *homB*, a novel co-virulence marker of *Helicobacter pylori*. *BMC Microbiol.* **9**:248.
30. Oleastro, M., R. Cordeiro, Y. Yamaoka, D. Queiroz, F. Mégraud, L. Monteiro, and A. Ménard. 2009. Disease association with two *Helicobacter pylori* duplicate outer membrane protein genes, *homB* and *homA*. *Gut Pathog.* **1**:12.
31. Peck, B., M. Ortkamp, K. D. Diehl, E. Hundt, and B. Knapp. 1999. Conservation, localization and expression of HopZ, a protein involved in adhesion of *Helicobacter pylori*. *Nucleic Acids Res.* **27**:3325–3333.
32. Pride, D. T., and M. J. Blaser. 2002. Concerted evolution between duplicated genetic elements in *Helicobacter pylori*. *J. Mol. Biol.* **316**:629–642.
33. Pride, D. T., R. J. Meinersmann, and M. J. Blaser. 2001. Allelic variation within *Helicobacter pylori* *babA* and *babB*. *Infect. Immun.* **69**:1160–1171.
34. Robertson, D. L., B. H. Hahn, and P. M. Sharp. 1995. Recombination in AIDS viruses. *J. Mol. Evol.* **40**:249–259.
35. Salminen, M. 2003. Detecting recombination in viral sequences: theory, p. 348–361. In M. Salemi and A.-M. Vandamme (ed.), *The phylogenetic handbook: a practical approach to DNA and protein phylogeny*. Cambridge University Press, Cambridge, United Kingdom.
36. Salminen, M. O., J. K. Carr, D. S. Burke, and F. E. McCutchan. 1995. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res. Hum. Retroviruses* **11**:1423–1425.
37. Santoyo, G., and D. Romero. 2005. Gene conversion and concerted evolution in bacterial genomes. *FEMS Microbiol. Rev.* **29**:169–183.
38. Solnick, J. V., L. M. Hansen, N. R. Salama, J. K. Boonjakuakul, and M. Syvanen. 2004. Modification of *Helicobacter pylori* outer membrane protein expression during experimental infection of rhesus macaques. *Proc. Natl. Acad. Sci. U. S. A.* **101**:2106–2111.
39. Suerbaum, S. 2000. Genetic variability within *Helicobacter pylori*. *Int. J. Med. Microbiol.* **290**:175–181.
40. Suerbaum, S., and M. Achtman. 1999. Evolution of *Helicobacter pylori*: the role of recombination. *Trends Microbiol.* **7**:182–184.
41. Suerbaum, S., J. M. Smith, K. Bapumia, G. Morelli, N. H. Smith, E. Kunstmann, I. Dyrek, and M. Achtman. 1998. Free recombination within *Helicobacter pylori*. *Proc. Natl. Acad. Sci. U. S. A.* **95**:12619–12624.
42. Templeton, A. R. 2007. Shared history of humans and gut bacteria: evolutionary togetherness: coupled evolution of humans and a pathogen. *Heredity* **98**:337–338.
43. Tsuru, T., M. Kawai, Y. Mizutani-Ui, I. Uchiyama, and I. Kobayashi. 2006. Evolution of paralogous genes: reconstruction of genome rearrangements through comparison of multiple genomes within *Staphylococcus aureus*. *Mol. Biol. Evol.* **23**:1269–1285.
44. Tsuru, T., and I. Kobayashi. 2008. Multiple genome comparison within a bacterial species reveals a unit of evolution spanning two adjacent genes in a tandem paralog cluster. *Mol. Biol. Evol.* **25**:2457–2473.
45. Vandamme, A.-M. 2003. Basic concepts of molecular evolution, p. 1–23. In M. Salemi and A.-M. Vandamme (ed.), *The phylogenetic handbook: a practical approach to DNA and protein phylogeny*. Cambridge University Press, Cambridge, United Kingdom.
46. Van Doorn, L.-J., C. Figueiredo, F. Mégraud, S. Pena, P. Midolo, D. M. Queiroz, F. Carneiro, B. Vanderborcht, M. D. Pegado, R. Sanna, W. De Boer, P. M. Schneeberger, P. Correa, E. K. Ng, J. Atherton, M. J. Blaser, and W. G. V. Quint. 1999. Geographic distribution of *vacA* allelic types of *Helicobacter pylori*. *Gastroenterology* **116**:823–830.
47. Wirth, T., A. Meyer, and M. Achtman. 2005. Deciphering host migrations and origins by means of their microbes. *Mol. Ecol.* **14**:3289–3306.
48. Yamaoka, Y., S. Kikuchi, H. M. T. El-Zimaity, O. Gutierrez, M. S. Osato, and D. Y. Graham. 2002. Importance of *Helicobacter pylori* *oipA* in clinical presentation, gastric inflammation, and mucosal interleukin 8 production. *Gastroenterology* **123**:414–424.