



Rapid evolution of the *Helicobacter pylori* AlpA adhesin in a high gastric cancer risk region from Colombia

Andrés Julián Gutiérrez-Escobar^{1,2}, Gina Méndez-Callejas¹, Orlando Acevedo³ and Maria Mercedes Bravo⁴

¹ Grupo de Investigaciones Biomédicas y Genética Humana Aplicada—GIBGA, Programa de medicina, Universidad de Ciencias Aplicadas y Ambientales U.D.C.A., Bogotá, Colombia

² Doctorado en Ciencias Biológicas, Pontificia Universidad Javeriana, Bogotá, Colombia

³ Grupo de Biofísica y Bioquímica Estructural, Facultad de Ciencias, Pontificia Universidad Javeriana, Bogotá, Colombia

⁴ Grupo de Investigación en Biología del Cáncer, Instituto Nacional de Cancerología de Colombia, Bogotá, Colombia

ABSTRACT

To be able to survive, *Helicobacter pylori* must adhere to the gastric epithelial cells of its human host. For this purpose, the bacterium employs an array of adhesins, for example, AlpA. The adhesin AlpA has been proposed as a major adhesin because of its critical role in human stomach colonization. Therefore, understanding how AlpA evolved could be important for the development of new diagnostic strategies. However, the genetic variation and microevolutionary patterns of *alpA* have not been described in Colombia. The study aim was to describe the variation patterns and microevolutionary process of *alpA* in Colombian clinical isolates of *H. pylori*. The existing polymorphisms, which are deviations from the neutral model of molecular evolution, and the genetic differentiation of the *alpA* gene from Colombian clinical isolates of *H. pylori* were determined. The analysis shows that gene conversion and purifying selection have shaped the evolution of three different variants of *alpA* in Colombia.

Submitted 2 March 2018

Accepted 6 May 2018

Published 25 May 2018

Corresponding author

Andrés Julián Gutiérrez-Escobar,
andregutierrez@udca.edu.co,
andresjulian1981@gmail.com

Academic editor

Joseph Gillespie

Additional Information and
Declarations can be found on
page 13

DOI 10.7717/peerj.4846

© Copyright

2018 Gutiérrez-Escobar et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Bioinformatics, Evolutionary Studies, Genetics, Microbiology

Keywords Rapid evolution, Gene convergence, *Helicobacter pylori*, Purifying selection, AlpA adhesin.

INTRODUCTION

Helicobacter pylori, a Gram-negative bacterium, has persistently colonized the stomach of half of the human population (Perez-Perez, Rothenbacher & Brenner, 2004; Khalifa, Sharaf & Aziz, 2010). This infection produces an asymptomatic inflammation of the gastric epithelium, but in some patients, it progresses toward a more severe clinical disease, such as ulcers and gastric cancer (Yakirevich & Resnick, 2013).

Gastric cancer is the fifth most common cancer worldwide (Jemal et al., 2010; Bertuccio et al., 2009; Forman & Sierra, 2014), and it is the second leading cause of cancer deaths (Ferlay et al., 2013); an infection with *H. pylori* is the strongest factor risk for its development (*Helicobacter and Cancer Collaborative Group*, 2001). In Colombia, the prevalence of this infection is universally high (Matta et al., 2017). However, the gastric cancer risk increases

along with the altitudinal gradient (Torres et al., 2013). Thus, it is higher in the Andes region than along the Pacific coast (Kodaman et al., 2014); this phenomenon has been called the Colombian enigma (Correa & Piazuolo, 2010).

The bacterium has coevolved with its human host for a period of 60,000 years, since the first major migratory wave outside Africa (Linz et al., 2007). The bacterium has followed a similar dispersal pattern as its human host. Currently, seven populations of *H. pylori* have been identified: hpEurope, hpNEAfrica, hpAfrica1, hpAfrica2, hpAsia2, hpSahul and hpEastAsia (Yamaoka, 2009; Falush et al., 2003; Achtman et al., 1999; Moodley et al., 2009; Devi et al., 2007). A recent human migratory event was the colonization of the Americas 525 years ago. During this colonization, new pathogens arrived to the American continent, including new strains of *H. pylori*, which caused the disappearance of 80% of the native population (Bianchine & Russo, 1992; Parrish et al., 2008). It has been reported that *H. pylori* has followed unique evolutionary pathways in Latin-America (Gutiérrez-Escobar et al., 2017; Muñoz Ramírez et al., 2017) and that the strains followed rapid adaptive processes in different countries of the region (Thorell et al., 2017), establishing local independent lineages.

The bacterium has virulence factors that correlate with the risk of developing gastric diseases (Parsonnet et al., 1997; Nomura et al., 2002; Mahdavi et al., 2002; Yamaoka et al., 2002a; Yamaoka et al., 2002b). *H. pylori* attach to the gastric epithelium thorough adhesins that contribute to the initial steps of the infection (Matsuo, Kido & Yamaoka, 2017). AlpA (~56 kDa) is an adhesin that is encoded by the locus *alpAB* (Alm et al., 2000), which is essential for adherence to the human gastric epithelium (Odenbreit et al., 1999). This adhesin is expressed by all clinical isolates (Odenbreit et al., 2009) and is recognized in sera from infected patients (Xue et al., 2005), and in addition, it induces the secretion of interleukin-8 (IL-8) (Lu et al., 2007). AlpA represents an emerging virulence factor of *H. pylori* that has been gaining attention because of its potential as a vaccine target (Sun et al., 2010).

The aim of the present study was to describe the genetic diversity and microevolution of the adhesin AlpA at the population level in the high gastric cancer risk region of Colombia. Population genetics statistics and phylogenetic methods were performed to compare *H. pylori* strains from Colombian clinical isolates against strains from different geographical backgrounds.

MATERIALS AND METHODS

DNA and protein sequences

The DNA (*alpA*) and protein sequences (AlpA) were obtained from 115 genomes from the *H. pylori* stock collection, which was previously sequenced by our research group at the Instituto Nacional de Cancerología in Bogotá. The isolates belonged to patients with different types of gastric pathologies associated with *H. pylori* infection, as follows: 30 cases of Gastritis (G), 20 of Gastric Adenocarcinoma (GA), 28 of Atrophic Gastritis (AG), 30 of Intestinal Metaplasia, five of Gastritis concomitant with Duodenal Ulcer (G-DU) and two of Intestinal Metaplasia concomitant with Duodenal Ulcer (IM-DU) (Gutiérrez-Escobar et al., 2017).

The reference pool sequences were obtained from 34 *H. pylori* strains, as follows: HspAmerind: Cuz20, PeCan4, Puno135, Sat464, Shi112, Shi169, Shi417, Shi470, v225d; HpEurope: 26695, B8, G27, HPAG1, ELS37, Lithuania75, SJM180; HspAsia: F57, XZ274, 51, 52, 35A, F32, F16, F30, 83; HspAsia2: SNT49, India7; HspWestAfrica: J99, 908, PeCan18, 2017, 2018, Gambia94/24 and HpSouthAfrica: SouthAfrica7.

Phylogenetic analysis of *alpA*

A total of 142 protein sequences for AlpA (108 sequences from Colombian isolates and 34 from reference) were aligned using the software Muscle V 3.8.31 ([Edgar, 2004](#)); the evolutionary model and the phylogenetic reconstruction was determined using MEGA V 7 ([Kumar, Stecher & Tamura, 2016](#)) and the NJ algorithm ([Saitou & Nei, 1987](#)) with 1,000 bootstrap repetitions for statistical robustness.

Genetic diversity, natural selection tests and differentiation analysis of *alpA*

The following population statistics: number of haplotypes (H), haplotype diversity (Hd), nucleotide diversity (Pi), average number of nucleotide differences (k), Theta estimator (θ_w) and recombination events (Rm) analyses were performed using DnaSP v 5.10 ([Librado & Rozas, 2009](#)). Deviations of the neutral model of molecular evolution were tested using the Tajima test ([Tajima, 1989](#)) and the Z-test in which the average number of synonymous substitutions per synonymous site (dS) and the average number of non-synonymous substitutions per non-synonymous site (dN) were calculated using the modified Nei-Gojobori method with the Jukes-Cantor correction. The variance of the difference was computed using the bootstrap method (1,000 replicates) using Mega v 7 ([Kumar, Stecher & Tamura, 2016](#)). Finally, a sliding window analysis was applied to detect the evolutionary rate $\omega = dN/dS$ throughout the gene to identify specific regions under natural selection using the software DnaSP v 5.10 ([Librado & Rozas, 2009](#)).

The DNA sequences of Colombian isolates were grouped according to the gastric pathology of the patients: Gastritis (G), Gastric Adenocarcinoma (GA), Atrophic Gastritis (AG), Intestinal Metaplasia (IM), Gastritis concomitant with Duodenal Ulcer (GDU) and Intestinal Metaplasia concomitant with Duodenal Ulcer (IM-DU); the reference sequences were grouped according to their geographic origin. To detect the genetic heterogeneity and genetic flow, the following tests were applied: Hst, Kst, Kst*, Z, Z*. The HBK, Snn and chi squared tests were performed using the haplotype frequencies under the permutation of 1,000 repetitions, as well as the tests for the haplotype diversity of Gst, Nst, Fst, and Da, and the gene flow (Nm), a measure of the genetic interaction, was estimated from FsT ([Slatkin, 1985](#); [Slatkin, 1987](#)) using the software DnaSP v 5.10 ([Librado & Rozas, 2009](#)).

Gene conversion analysis of *alpA*

To test whether gene conversion generates genetic diversity in the *alpA*, the Betran's method ([Betran et al., 1997](#)) implemented in the DnaSP v 5.10 software was used to compare Colombian isolates and the reference pool populations. RDP3 v 3.4 software was used to test gene conversion in the overall population using the GENECONV algorithm ([Sawyer, 1989](#)). Only conversion tracks with $p < 0.05$ were considered.

Type I functional divergence and site specific positive selection analyses

The 3D structure of the AlpA protein was predicted using the server I-TASSER (Yang *et al.*, 2015) and tested using the MetaMQAPII server (Pawlowski *et al.*, 2008). DIVERGE V 3 software (Gu *et al.*, 2013) was used to estimate type-I functional divergence, which detects functional changes in a protein based on site-specific shifts in the evolutionary rates (Gu, 1999). The software tested whether a significant change in the evolution rate has occurred by calculating the coefficient of divergence (θD). Positive and negative selection was evaluated as the proportions of synonymous to non-synonymous substitution rates. The DNA sequences were aligned using the software Muscle V 3.8.31 (Edgar, 2004), and the alignment file was pre-processing by screening for recombination breakpoints using the GARD algorithm implemented by the HyPhy software (Kosakovsky Pond *et al.*, 2006a; Kosakovsky Pond *et al.*, 2006b). Then, the processed file was tested for selection using the FEL and IFEL algorithms using the datamonkey server (Kosakovsky Pond & Frost, 2005). Episodic diversifying selection was detected using the MEME algorithm (Murrell *et al.*, 2012). A $p < 0.01$ was considered to be statistically significant for all of the selection tests. The RELAX test to detect relaxed selection on the codon-based phylogenetic framework of *alpA* from Colombian isolates was performed using the datamonkey server (Kosakovsky Pond & Frost, 2005).

RESULTS

A total of 142 sequences were used in this study: 86 were obtained from Colombian patients with different gastric diseases, and 34 sequences were obtained from reference strains from GenBank. The phylogenetic tree of AlpA using only the reference pool showed three clades: one clustering sequences from HspAmerind and HspAsia populations, the second clustering sequences from HpEurope and HspSouthIndia, and the last clustering sequences exclusively from HspWestAfrica. The tree does not reflect a clear separation between HspAmerind and HspAsia nor between HpEurope and HspSouthIndia, which suggests that recombination and gene conversion gave rise to its evolutionary pattern (Fig. 1).

The phylogenetic tree including Colombian isolates showed an intricate pattern of clades. Five major clades were detected: (1) cluster sequences from HspWestAfrica, HspColombia and HpEurope; (2) cluster sequences from HpEurope, HspSouthIndia and HspColombia; (3) cluster sequences from HspColombia; (4) cluster sequences from HspAsia, HspAmerind; and (5) cluster sequences from HpEurope, HspAsia and HspColombia (Fig. 2). The phylogenetic tree of the Colombian isolates showed three major clades, called Col1, Col2 and Col3; a more detailed analysis allowed us to show that the phylogenetic tree had seven subclades, called 1 to 7, which indicates that an intense evolutionary process is taking place between Colombian isolates with respect to AlpA (Fig. 3).

The analysis of the nucleotide diversity (2.5-fold) and the average number of nucleotide differences (3.7-fold) showed them to be higher in the reference pool than in the Colombian

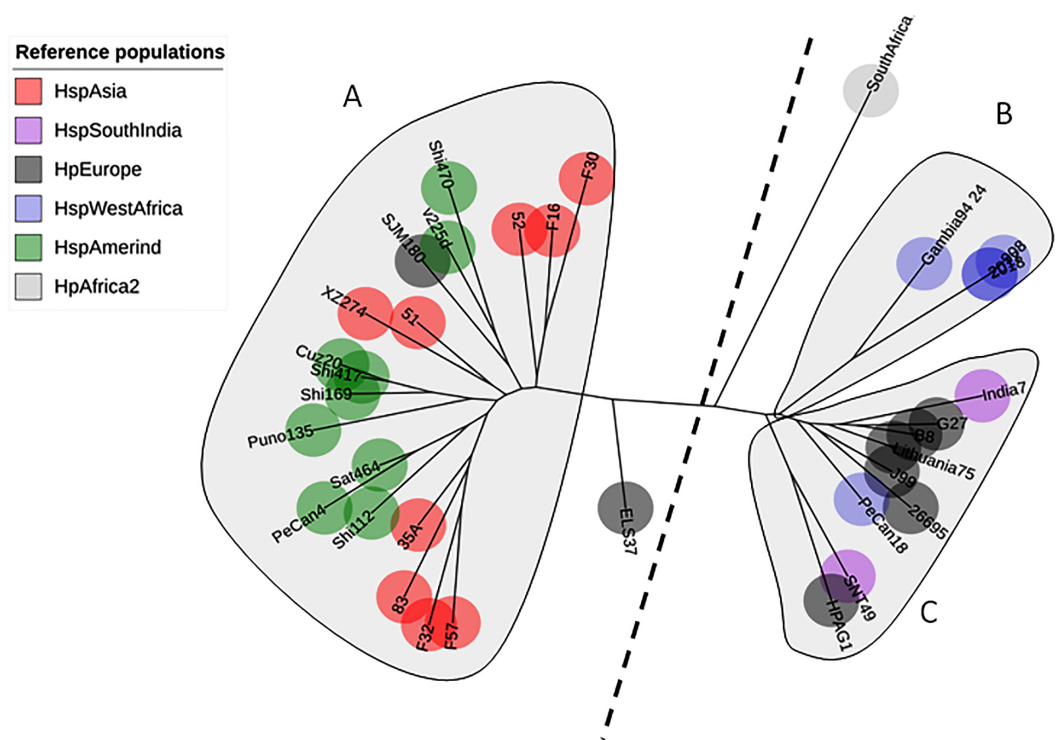


Figure 1 Phylogenetic tree of *AlpA* from reference strains of *H. pylori*. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.84663616 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates); significant consensus tree branches are showed. The evolutionary distances were computed using the JTT matrix-based method and are in the units of the number of amino acid substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 2). The analysis involved 34 amino acid sequences. (A) Clade East. (B) Clade Western. All positions containing gaps and missing data were eliminated. There were a total of 451 positions in the final dataset.

Full-size DOI: [10.7717/peerj.4846/fig-1](https://doi.org/10.7717/peerj.4846/fig-1)

isolates. Similarly, the theta estimator showed that the reference pool was significantly more diverse than the Colombian counterparts. The number of haplotypes was 3-fold higher in the Colombian isolates than that observed in the reference pool, but the haplotype diversity was similarly higher in both populations; the total number of haplotypes was 134, with an extreme value for the haplotype diversity. Recombination events were 1.3-fold higher in the Colombian isolates than in the reference pool (Table 1).

To test the deviations of the neutral model of molecular evolution, the Tajima and *Z*-test were employed. The Tajima's *D* test was negative and the *Z*-test showed significant results for neutrality, indicating that in overall average the synonymous substitution rate was similar the non-synonymous substitution rate ($p = 0.005$) (Table 2). However, the sliding windows analysis showed that several regions of the *alpA* has an ω value above 1, which suggests that episodic positive diversifying selection is also shaping the microevolutionary patterns of this gene in Colombia (Fig. 4).

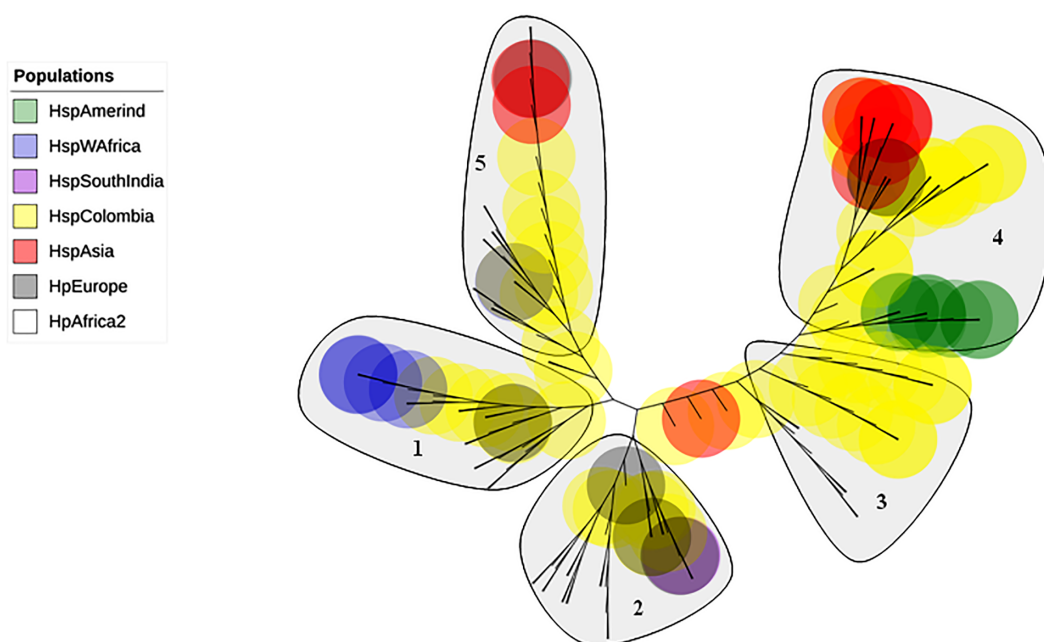


Figure 2 Phylogenetic tree of AlpA proteins of *Helicobacter pylori*. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.84663616 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates); significant consensus tree branches are shown. The evolutionary distances were computed using the JTT matrix-based method and are in the units of the number of amino acid substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 2). The analysis involved 142 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 451 positions in the final dataset. Five major clades were detected: (1) Cluster sequences from HspWestAfrica, HspColombia and HpEurope; (2) Cluster sequences from HpEurope, HspSouthIndia and HspColombia; (3) Cluster sequences from HspColombia; (4) Cluster sequences from HpAsia, HspAmerind and HspColombia; and (5) Cluster sequences from HpEurope, HpAsia and HspColombia.

Full-size DOI: 10.7717/peerj.4846/fig-2

The test of genetic differentiation showed that the Colombian population was well-differentiated from the reference pool, but as expected, the N_m obtained from F_{ST} was 3.63, which indicates a moderate gene flow between the Colombian isolates (Slatkin, 1985; Slatkin, 1987) (Table 3). To assess the isolation between pairs of populations, the Colombian isolates were organized into seven groups based on the histopathological diagnosis, and then, the *alpA* DNA sequences were compared pairwise with the reference pool clustered according to their phylogeographic classification. Population isolation was found between the Colombian populations and all subpopulations (Table 4). Two gene conversion algorithms were applied to the population to test whether this process was involved in the evolutionary pattern of *alpA*. Betran's algorithm found 54 conversion tracks between the sequences: 20.5% for the reference pool and 44.4% for the Colombian isolates, and the genconv algorithm found six tracks of gene conversion between the sequences that belong exclusively to the Colombian isolates (Fig. 5).

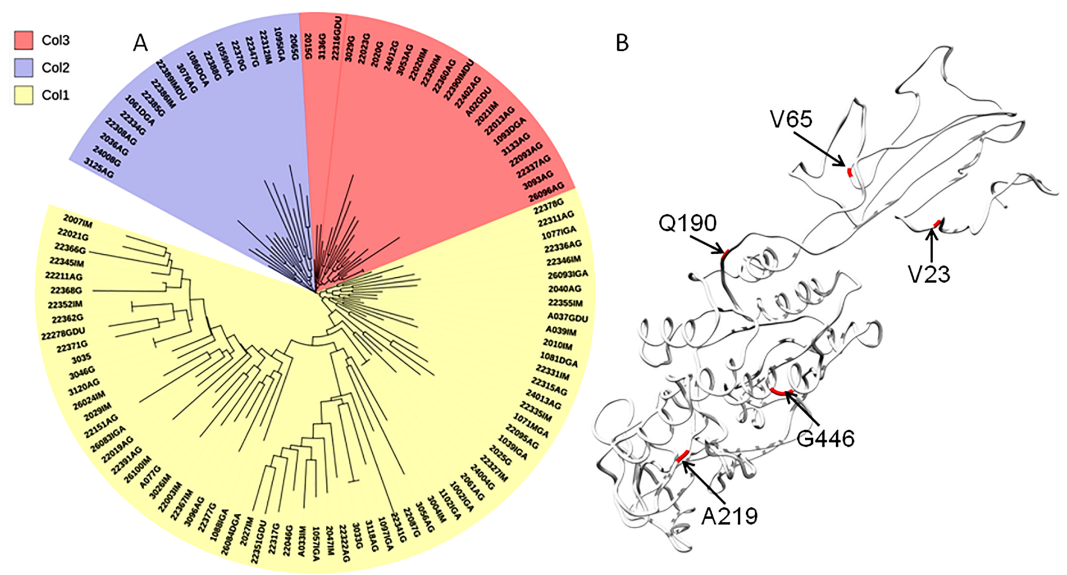


Figure 3 Phylogenetic tree of AlpA proteins from Colombian isolates of *Helicobacter pylori*. (A) The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.84663616 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates); significant consensus tree branches are shown. The evolutionary distances were computed using the JTT matrix-based method and are in the units of the number of amino acid substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 2). The analysis involved 108 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 451 positions in the final dataset. Three major clades were detected: (1) Col1; (2) Col2 and (3) Col3. (B) Comparison all-vs-all between Colombian phylogenetic tree clusters reveals that Col1/Col3 clades have five sites with a $\theta D > 0.8$ indicating a strong signal of functional divergence.

Full-size DOI: [10.7717/peerj.4846/fig-3](https://doi.org/10.7717/peerj.4846/fig-3)

Table 1 Genetic diversity statistics for alpA. The genetic diversity statistics were calculated by using 34 sequences from reference pool and 108 sequences from Colombian isolates of *Helicobacter pylori*.

<i>n</i>	Sites	Single nucleotide polymorphism						
		<i>Ss</i>	<i>k</i>	<i>H</i>	<i>Hd</i>	θ_w	π	Rm
<i>Reference</i>								
34	1,324	774	200.868	33	0.998	0.143	0.152	70
<i>Colombian</i>								
108	1,384	306	53.335	102	0.998	0.042	0.039	91
<i>Colombian and references</i>								
142	1,223	754	81.411	134	0.999	0.112	0.067	98

Notes.

The estimators were presented in three groups: references, Colombian and both together.

n, number of isolates; sites, total number of sites analyzed (excluding gaps); *Ss*, number of segregant sites; *k*, average number of nucleotide differences by sequence pairs; *H*, number of haplotypes; *Hd*, Haplotype diversity; θ_w , Watterson estimator; π , nucleotide diversity per site and; Rm, recombination events.

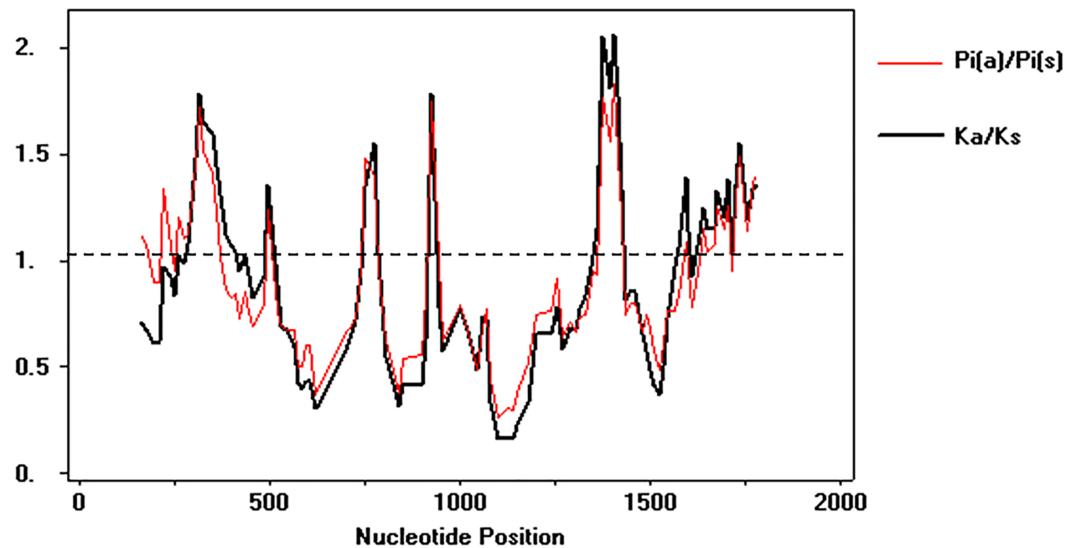


Figure 4 Sliding windows analysis of ω rate Evolutionary rate $\omega = dN/dS$ throughout the gene to identify specific regions under natural selection. The analysis involved 142 DNA sequences, 108 from Colombian isolates and 34 as phylogeographic references. The analysis was performed using DnaSP v 5.10.

Full-size [DOI: 10.7717/peerj.4846/fig-4](https://doi.org/10.7717/peerj.4846/fig-4)

Table 2 Deviations of the neutral model of molecular evolution. The analysis involved 142 nucleotide sequences. Tajima test: codon positions included were first. All positions containing gaps and missing data were eliminated.

Tajima test						Codon Based Z-test of selection		
<i>m</i>	<i>S</i>	<i>p_s</i>	Θ	π	<i>D</i>	dN = dS	dN > dS	dN < dS
142	217	0.604	0.109	0.059	-1.476	-2.878***	-0.213 ns	0.214 ns

Notes.

There were a total of 359 positions in the final dataset, *m*, number of sequences, *n*, total number of sites; *S*, Number of segregating sites; *p_s*, *S/n*; Θ , *ps/a1*; π , nucleotide diversity, and *D* is the Tajima statistic. Z-Test: Codon-based test of neutrality for analysis averaging over all sequence pairs. The probability of rejecting the null hypothesis of strict-neutrality (dN = dS) is shown. Values of *P* less than 0.05 are considered significant at the 5% level and are highlighted (****p* < 0.001; ns, not significant). The test statistics: dN < dS, neutrality; dN > dS positive selection; dN < dS purifying selection are shown. dS and dN are the numbers of synonymous and nonsynonymous substitutions per site, respectively. The variance of the difference was computed using the bootstrap method (1,000 replicates). Analyses were conducted using the Nei-Gojobori method. All positions containing gaps and missing data were eliminated.

Table 3 Genetic heterogeneity and genetic flow of *alpA*. Genetic heterogeneity and genetic flow were detected using the following tests: Hst, Kst, Kst*, Z, Z*.

Genetic differentiation							Genetic flow			
<i>Chi2</i>	<i>Hst</i>	<i>Kst</i>	<i>Kst*</i>	<i>Z</i>	<i>Z*</i>	<i>Snn</i>	<i>Gst</i>	<i>GammaST</i>	<i>Nst</i>	<i>Fst</i>
1501.4 ns	0.003**	0.090**	0.034***	4011.8***	7.837***	0.316***	0.011 (44.82)	0.178 (2.30)	0.114 (3.88)	0.120 (3.63)

Notes.

The HBK, Snn and chi squared tests were performed using the haplotype frequencies under the permutation of 1,000 repetitions, as well as the tests for haplotype diversity of Gst, Nst, Fst. In parenthesis, the gene flow (Nm) value was estimated from FsT using the software DnaSP 5.10.

Statistical significance: **p* < 0.05. ***p* < 0.01. ****p* < 0.001. ns, not significant.

Table 4 Pairwise analysis of differentiation and genetic flow in the populations of *alpA*.

Populations	FsT	GammaST
Amerind		
G vs. hspAmerind	0.256	0.113
GA vs. hspAmerind	0.325	0.151
IM vs. hspAmerind	0.276	0.141
IGA vs. hspAmerind	0.281	0.212
DGA vs. hspAmerind	0.379	0.328
G/DU vs. hspAmerind	0.246	0.243
IM/DU vs. hspAmerind	0.504	0.373
Europe		
G vs. HpEurope	0.125	0.143
GA vs. HpEurope	0.125	0.148
IM vs. HpEurope	0.131	0.155
IGA vs. HpEurope	0.133	0.160
DGA vs. HpEurope	0.119	0.128
G/DU vs. HpEurope	0.11	0.123
IM/DU vs. HpEurope	0.153	0.093
Asia		
G vs. HpAsia	0.119	0.123
GA vs. HpAsia	0.135	0.137
IM vs. HpAsia	0.123	0.132
IGA vs. HpAsia	0.132	0.133
DGA vs. HpAsia	0.153	0.118
G/DU vs. HpAsia	0.11	0.099
IM/DU vs. HpAsia	0.204	0.086
Africa		
G vs. HpAfrica	0.128	0.134
GA vs. HpAfrica	0.123	0.137
IM vs. HpAfrica	0.131	0.148
IGA vs. HpAfrica	0.124	0.163
DGA vs. HpAfrica	0.099	0.138
G/DU vs. HpAfrica	0.111	0.145
IM/DU vs. HpAfrica	0.165	0.121

Notes.

The FsT and GammaST differentiation parameters are shown. The populations of Colombian isolates are denoted as follows: G, Gastritis; GA, Gastric Adenocarcinoma; IM, Intestinal Metaplasia; IGA, Intestinal Gastric Adenocarcinoma; DGA, Diffuse Gastric Adenocarcinoma; G/DU, Gastritis + Duodenal ulcer; IM/DU, Metaplasia + Duodenal ulcer. A total of 142 sequences were used for the analyses using DnaSP V 5.10. The significance of genetic differentiation was assumed as follow: <0.05 = little, 0.05–0.15 = moderate, 0.15–0.25 = high and >0.25 = complete (Hartl & Clark, 1997). Significant FsT values are highlighted in bold.

Functional divergence analysis among the AlpA proteins sequences based on the three major clades found in the phylogenetic tree of the Colombian isolates was performed using Gu's type-I method. The pairwise comparisons between the Col1/Col3 clades showed five sites with a $\theta D > 0.8$ (Fig. 3B), which indicates that the protein presents sites with different evolutionary rates. The analysis of positive and negative selection covered 540 codon sites of

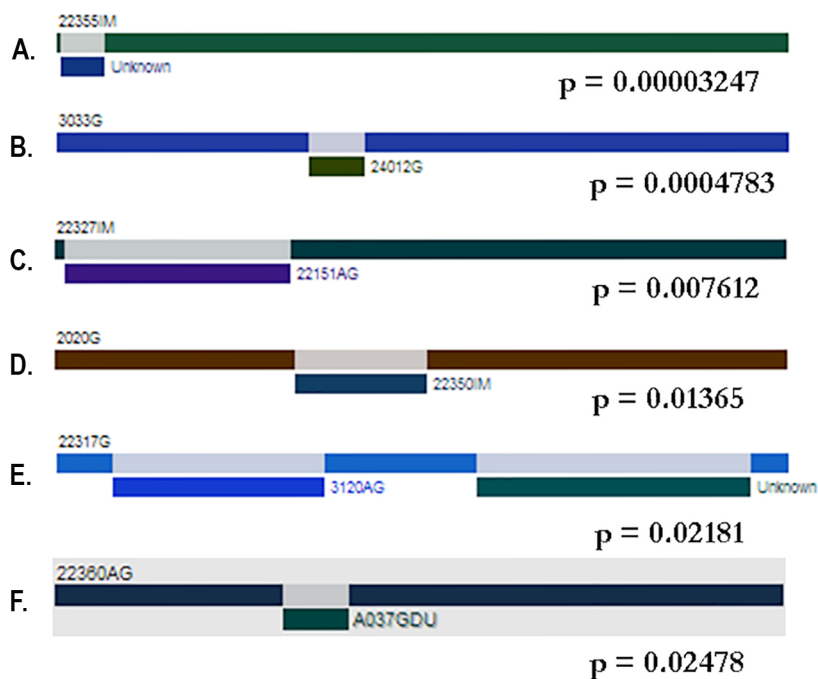


Figure 5 Gene conversion tracks identified by the GENECONV method. (A–F) are bars that represent genes. The colored rectangles represent the gene fragments or minor parents involved in a gene conversion event. The tracks were located predominantly at the 5' region of the gene; however some events were also detected at 3' region.

Full-size DOI: 10.7717/peerj.4846/fig-5

the AlpA protein. A similar number of sites were detected by the FEL and IFEL algorithms. FEL identified that 3.3% of the sites were under positive selection and 21.4% evolve under purifying selection. IFEL showed that 2.7% of the sites evolve under positive selection, with 12.2% under purifying selection, and finally, MEME showed that 5% of the adhesin was under episodic diversifying selection (Fig. 6). The relax test showed that when the internal branches were compared against the external branches, a significant pattern of natural selection intensification was detected ($K = 29.55$, $p = 5.48e^{-10}$, $LR = 38.50$).

DISCUSSION

In this study, we describe the genetic diversity and microevolution of AlpA in Colombian isolates of *H. pylori* obtained from a high gastric cancer region composed of the cities of Bogotá, Tunja and the surrounding towns. The phylogenetic tree using the reference AlpA sequences revealed that the strains clustered according to their geographic origin; two main clades were observed: the Eastern and the Western clades. Then, within each main clade, there was no separation between the *H. pylori* subpopulations. This geographical segregation has been observed for other virulence factors of *H. pylori* (Cao *et al.*, 2005; Maeda *et al.*, 1998; Oleastro *et al.*, 2009a; Van Doorn *et al.*, 1999) and it is an indicative of the pre-colonization period. However, when the sequences of Colombian isolates are added to the analysis, the main East and Western clades fade, which indicates that the sequences share

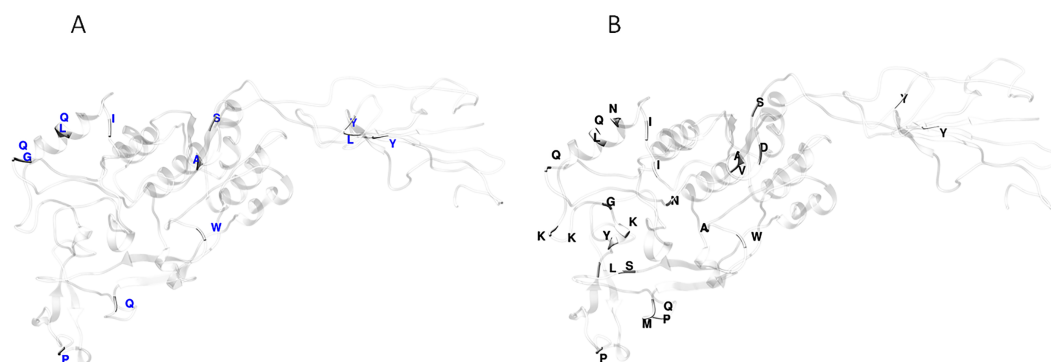


Figure 6 Protein model of AlpA showing the positive episodic diversified selected sites after the SBP correction thereby recombination. (A) Positive selected sites were obtained from the tests FEL (18 sites) and IFEL (14 sites) showed only common sites at the protein structure in red. (B) Episodic diversifying selective sites were obtained using the test MEME (27 sites) in black, and both analyses were performed using the Datamonkey server from the codon alignment file. Only sites with a $p < 0.01$ are shown.

Full-size DOI: [10.7717/peerj.4846/fig-6](https://doi.org/10.7717/peerj.4846/fig-6)

information by recombination and diversifying selection induced by the post-colonization period, when the host and bacterial populations mix together in Latin-America, and for the induction of different divergence rates between the paralogous/homologous members of different strains as a consequence of gene conversion ([Santoyo & Romero, 2005](#)).

The phylogenetic tree of Colombian isolates showed tree dominant clades entitled Col1, Col2 and Col3. The clades Col1 and Col2 have few branching events with a low number of members, but the Col3 clade showed five subclades. There was no association between the disease state and belonging to a specific clade ($p = 0.245$ Chi-square 2.81), which indicates that the *H. pylori* strains interchange DNA randomly between the Colombian strains. It is a well-established fact that the bacterium is naturally competent ([Dorer et al., 2013](#)).

The nucleotide diversity and the average number of nucleotide differences was lower in the Colombian *alpA* alleles that in the reference pool. However, the Colombian isolates showed a higher number of haplotypes and recombination events. The *alpA* gene from Colombian isolates has shown strong patterns of genetic differentiation, but it maintains the genetic flow that has produced new allelic variants of *alpA* in Colombia. The arrival of the HpEurope to Latin-America induced the replacement of the native HspAmerind population from urban zones ([Yamaoka et al., 2002a](#); [Yamaoka et al., 2002b](#)) and perhaps produced a selective bottleneck. After the bottleneck, a subsequent population expansion emerged with the new recombinant *H. pylori* subtypes in Colombia. It has been proposed that recombination among *H. pylori* strains can induce the evolution of different subclones and genotypes ([Blaser & Berg, 2001](#); [Blaser & Atherton, 2004](#)).

H. pylori has one of the highest mutation and recombination rates between bacterial species ([Suerbaum & Josenhans, 2007](#)). It has been identified that *H. pylori* has followed unique evolutionary pathways in Latin-America ([Muñoz Ramírez et al., 2017](#)), and the strains have followed rapid adaptive processes in different countries of the region ([Gutiérrez-Escobar et al., 2017](#); [Thorell et al., 2017](#)). Another factor that contributes to the rapid evolution of the Colombian *H. pylori alpA* subtypes was the mestizo host. Perhaps

the immune and inflammatory responses of this new host and its genetic heterogeneity represented by the variability and distribution of receptors to which the bacterium can adhere could be a selective factor for the divergence among Colombian *H. pylori* strains (Dubois et al., 1999; Oleastro et al., 2009b). In fact, the host imposes a selective pressure that induces variation within the bacterium (Thompson et al., 2004).

Gene conversion was a major process that gave rise to the allelic variation of *alpA* in Colombia, and the paralogous interchange of the DNA fragments close to the 3' region of the gene was considerable higher between the Colombian isolates than in the reference pool. In total, 60 recombination events were detected by the two applied algorithms, and despite the total number of sequences used in the analysis, only Colombian *alpA* alleles showed statistical signals of this type of recombination. Gene conversion has been identified previously in other *H. pylori* adhesins, for example, *homB* and *homA* (Dorer et al., 2013; Dubois et al., 1999; Oleastro et al., 2009a; Oleastro et al., 2009b; Santoyo & Romero, 2005; Yamaoka et al., 2002a; Yamaoka et al., 2002b), *sabB* and *omp27* (Talarico et al., 2012) and *babA* (Hennig, Allen & Cover, 2006), and a Latin-American type of *babA* has also been recently reported (Thorell et al., 2016).

The interplay between positive selection and recombination has been detected in bacterial genomes (Lefebure & Stanhope, 2007; Joseph et al., 2011; Orsi, Sun & Wiedmann, 2008). The analysis of functional divergence analysis of the protein AlpA based on the comparison of the tree clades Col1, Col2 and Col3 showed that the protein residues 23V, 65V, 190Q, 219A and 446G have a significant evolutionary site variation rate with a θD value of 0.8, which confers functional differences between the members of the clades Col1/Col3. In Colombia, positive, episodic diversifying, purifying selection and recombination of *alpA* has given rise to the presence of rare alleles and new haplotypes in the emerging population, which have been maintained at the protein level by natural selection.

The action of natural selection might purge diversified genes from the population by strong purifying selection (Lynch & Conery, 2000) or lead to evolutionary novelties by positive selection (Wertheim, Murrell & Smith, 2015). When the internal branches of the phylogenetic tree of Colombian *alpA* alleles were compared against the external ones, we determined that there was a significant natural selection intensification operating in the *alpA* alleles ($K = 29.55$, $p = 5.48e^{-10}$, $LR = 38.50$), which means that positive selection is leading the microevolution at the high gastric cancer risk region of Colombia.

The Mestizo populations in the mountain zones in Colombia have an admixture of European and Amerind ancestries in a similar fashion that the bacterium mirroring the colonization process (Kodaman et al., 2014). When the Colombian *H. pylori* strains and human host ancestries are compared, a difference in the intensification of the disease aggressiveness is observed. Deleterious duplicated alleles of *alpA* were purged out of the population, but those strains with fixed *alpA* alleles under positive selection give advantages to the new types of strains in this region.

H. pylori is a bacterium that can display very fast local adaptive processes via mutation and recombination (Cao et al., 2015; Furuta et al., 2015). The phylogenetics, population genetics and protein evolutionary analysis suggest that *alpA* in Colombia has functionally divergent variants that are the result of gene conversion in a staggeringly short period

of time. AlpA proteins from the Colombian population show evidence of functional divergence, positive selection and episodic positive selection at specific sites. It is possible that the polymorphism of this adhesin in Colombia reflects the phylogeography and historical generation of Mestizos in Latin-America because *H. pylori* is a reliable biological marker of human migratory events (*Templeton, 2007*).

CONCLUSION

The molecular evolution of virulence factors of *H. pylori* is currently gaining attention in the scientific community due to the new genomics and evolutionary findings around the world. Currently, Latin-American countries have emerged as evolutionary laboratories for *H. pylori*. To our knowledge, this study is the first study that presents statistically-supported evidence that *alpA* alleles from a high gastric cancer risk area from Colombia owe their variation patterns to gene conversion and purifying selection. In addition, a fast process of gene diversification followed by positive and relaxed selection has shaped three protein variants for the AlpA adhesin in Colombia.

ACKNOWLEDGEMENTS

Thank to the anonymous reviewers for the constructive comments and guidelines.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by Colciencias Grant Contract No. 599-2014 to Gina Méndez and Andrés Julian Gutiérrez-Escobar and a Grant 41030610588 from Instituto Nacional de Cancerología to Maria Mercedes Bravo. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

Colciencias Grant Contract: 599-2014.

Instituto Nacional de Cancerología: 41030610588.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Andrés Julián Gutiérrez-Escobar conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper.
- Gina Méndez-Callejas analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.

- Orlando Acevedo and Maria Mercedes Bravo analyzed the data, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The sequences are provided as [Supplemental Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.4846#supplemental-information>.

REFERENCES

- Achtman M, Azuma T, Berg DE, Ito Y, Morelli G, Pan ZJ, Suerbaum S, Thompson SA. 1999. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Molecular Microbiology* 32(3):459–470 DOI 10.1046/j.1365-2958.1999.01382.x.
- Alm RA, Bina J, Andrews BM, Doig P, Hancock RE, Trust TJ. 2000. Comparative genomics of *Helicobacter pylori*: analysis of the outer membrane protein families. *Infection and Immunity* 68(7):4155–4168 DOI 10.1128/IAI.68.7.4155-4168.2000.
- Bertuccio P, Chatenoud L, Levi F, Praud D, Ferlay J, Negri E, Malvezzi M, La Vecchia C. 2009. Recent patterns in gastric cancer: a global overview. *International Journal of Cancer* 125(3):666–673 DOI 10.1002/ijc.24290.
- Betran E, Rozas J, Navarro A, Barbadilla A. 1997. The estimation of the number and the length distribution of gene conversion tracts from population DNA sequence data. *Genetics* 146(1):89–99.
- Bianchine PJ, Russo TA. 1992. The role of epidemic infectious diseases in the discovery of America. *Allergy Proceedings* 13(5):225–232 DOI 10.2500/108854192778817040.
- Blaser MJ, Atherton JC. 2004. *Helicobacter pylori* persistence: biology and disease. *Journal of Clinical Investigation* 113:321–333 DOI 10.1172/JCI20925.
- Blaser MJ, Berg DE. 2001. *Helicobacter pylori* genetic diversity and risk of human disease. *Journal of Clinical Investigation* 107:767–773 DOI 10.1172/JCI12672.
- Cao P, Lee KJ, Blaser MJ, Cover TL. 2005. Analysis of hopQ alleles in East Asian and Western strains of *Helicobacter pylori*. *FEMS Microbiology Letters* 251(1):37–43 DOI 10.1016/j.femsle.2005.07.023.
- Cao Q, Didelot X, Wu Z, Li Z, He L, Li Y, Ni M, You Y, Lin X, Li Z, Gong Y, Zheng M, Zhang M, Liu J, Wang W, Bo X, Falush D, Wang S, Zhang J. 2015. Progressive genomic convergence of two *Helicobacter pylori* strains during mixed infection of a patient with chronic gastritis. *Gut* 64(4):554–561 DOI 10.1136/gutjnl-2014-307345.
- Correa P, Piazuelo B. 2010. Gastric cancer: the colombian enigma. *Revista Colombiana de Gastroenterologia* 25(4):334–337.

- Devi SM, Ahmed I, Francalacci P, Hussain MA, Akhter Y, Alvi A, Sechi LA, Mégraud F, Ahmed N. 2007. Ancestral European roots of *Helicobacter pylori* in India. *BMC Genomics* 8:184 DOI 10.1186/1471-2164-8-184.
- Dorer MS, Cohen IE, Sessler TH, Fero J, Salama NR. 2013. Natural competence promotes *Helicobacter pylori* chronic infection. *Infection and Immunity* 81(1):209–215 DOI 10.1128/IAI.01042-12.
- Dubois A, Berg DE, Incecik ET, Fiala N, Heman-Ackah LM, Del Valle J, Yang M, Wirth HP, Perez-Perez GI, Blaser MJ. 1999. Host specificity of *Helicobacter pylori* strains and host responses in experimentally challenged non-human primates. *Gastroenterology* 116(1):90–96 DOI 10.1016/S0016-5085(99)70232-5.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5):1792–1797 DOI 10.1093/nar/gkh340.
- Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, Blaser MJ, Graham DY, Vacher S, Perez-Perez GI, Yamaoka Y, Mégraud F, Otto K, Reichard U, Katzowitsch E, Wang X, Achtman M, Suerbaum S. 2003. Traces of human migrations in *Helicobacter pylori* populations. *Science* 299(5612):1582–1585 DOI 10.1126/science.1080857.
- Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. 2013. GLOBOCAN 2012 v10, Cancer Incidence and Mortality Worldwide: IARC Cancer Base No.11, International Agency for Research on Cancer. Available at <http://globocan.iarc.fr>.
- Forman D, Sierra MS. 2014. The current and projected global burden of gastric cancer. *Helicobacter pylori* Eradication as a Strategy for Preventing Gastric Cancer. IARC *Helicobacter pylori* Working Group. Lyon, France: International Agency for Research on Cancer (IARC Working Group Reports, No. 8), 5–15. Available at <http://www.iarc.fr/en/publications/pdfs-online/wrk/wrk8/index.php>.
- Furuta Y, Konno M, Osaki T, Yonezawa H, Ishige T, Imai M, Shiwa Y, Shibata-Hatta M, Kanesaki Y, Yoshikawa H, Kamiya S, Kobayashi I. 2015. Microevolution of virulence-related genes in *Helicobacter pylori* familial infection. *PLOS ONE* 10(5):e0127197 DOI 10.1371/journal.pone.0127197.
- Gu X. 1999. Statistical methods for testing functional divergence after gene duplication. *Molecular Biology and Evolution* 16(12):1664–1674 DOI 10.1093/oxfordjournals.molbev.a026080.
- Gu X, Zou Y, Su Z, Huang W, Zhou Z, Arendsee Z, Zeng Y. 2013. An update of DIVERGE software for functional divergence analysis of protein family. *Molecular Biology and Evolution* 30(7):1713–1719 DOI 10.1093/molbev/mst069.
- Gutiérrez-Escobar AJ, Trujillo E, Orlando E, Acevedo S, Bravo MM. 2017. Phylogenomics of Colombian *Helicobacter pylori* isolates. *Gut Pathogens* 9:52–61 DOI 10.1186/s13099-017-0201-1.
- Hartl DL, Clark GC. 1997. *Principles of Population Genetics*. Sunderland: Sinauer Associates.

- Helicobacter and Cancer Collaborative Group. 2001.** Gastric cancer and *Helicobacter pylori*: a combined analysis of 12 case control studies nested within prospective cohorts. *Gut* **49**:347–353 DOI [10.1136/gut.49.3.347](https://doi.org/10.1136/gut.49.3.347).
- Hennig EE, Allen JM, Cover TL. 2006.** Multiple chromosomal loci for the babA gene in *Helicobacter pylori*. *Infection and Immunity* **74**(5):3046–3051 DOI [10.1128/IAI.74.5.3046-3051.2006](https://doi.org/10.1128/IAI.74.5.3046-3051.2006).
- Jemal A, Center MM, DeSantis C, Ward EM. 2010.** Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiology Biomarkers & Prevention* **19**(8):1893–1907 DOI [10.1158/1055-9965.EPI-10-0437](https://doi.org/10.1158/1055-9965.EPI-10-0437).
- Joseph SJ, Didelot X, Gandhi K, Dean D, Read TD. 2011.** Interplay of recombination and selection in the genomes of *Chlamydia trachomatis*. *Biology Direct* **6**:28 DOI [10.1186/1745-6150-6-28](https://doi.org/10.1186/1745-6150-6-28).
- Khalifa MM, Sharaf RR, Aziz RK. 2010.** *Helicobacter pylori*: a poor man's gut pathogen? *Gut Pathogens* **2**:2 DOI [10.1186/1757-4749-2-2](https://doi.org/10.1186/1757-4749-2-2).
- Kodaman N, Pazos A, Schneider BG, Piazuolo MB, Mera R, Sobota RS, Sicinschi LA, Shaffer CL, Romero-Gallo J, De Sablet T, Harder RH, Bravo LE, Peek Jr RM, Wilson KT, Cover TL, Williams SM, Correa P. 2014.** Human and *Helicobacter pylori* coevolution shapes the risk of gastric disease. *Proceedings of the National Academy of Sciences of the United States of America* **111**(4):1455–1460 DOI [10.1073/pnas.1318093111](https://doi.org/10.1073/pnas.1318093111).
- Kosakovsky Pond SL, Frost SD. 2005.** Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution* **22**(5):1208–1222 DOI [10.1093/molbev/msi105](https://doi.org/10.1093/molbev/msi105).
- Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. 2006a.** GARD: a genetic algorithm for recombination detection. *Bioinformatics* **22**:3096–3098 DOI [10.1093/bioinformatics/btl474](https://doi.org/10.1093/bioinformatics/btl474).
- Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. 2006b.** Automated phylogenetic detection of recombination using a genetic algorithm. *Molecular Biology and Evolution* **23**(10):1891–1901 DOI [10.1093/molbev/msl051](https://doi.org/10.1093/molbev/msl051).
- Kumar S, Stecher G, Tamura K. 2016.** MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger data sets. *Molecular Biology and Evolution* **33**(7):1870–1874 DOI [10.1093/molbev/msw054](https://doi.org/10.1093/molbev/msw054).
- Lefebure T, Stanhope MJ. 2007.** Evolution of the core and pangenome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome Biology* **8**:R71 DOI [10.1186/gb-2007-8-5-r71](https://doi.org/10.1186/gb-2007-8-5-r71).
- Librado P, Rozas J. 2009.** DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**(11):1451–1452 DOI [10.1093/bioinformatics/btp187](https://doi.org/10.1093/bioinformatics/btp187).
- Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P, Falush D, Stamer C, Prugnolle F, Van der Merwe SW, Yamaoka Y, Graham DY, Perez-Trallero E, Wadstrom T, Suerbaum S, Achtman M. 2007.** An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* **445**(7130):915–918 DOI [10.1038/nature05562](https://doi.org/10.1038/nature05562).

- Lu H, Wu JY, Beswick EJ, Ohno T, Odenbreit S, Haas R, Reyes VE, Kita M, Graham DY, Yamaoka Y. 2007. Functional and intracellular signalling differences associated with the *Helicobacter pylori* AlpAB adhesin from Western and East Asian strains. *Journal of Biological Chemistry* 282(9):6242–6254 DOI 10.1074/jbc.M611178200.
- Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290(5494):1151–1155 DOI 10.1126/science.290.5494.1151.
- Maeda S, Ogura K, Yoshida H, Kanai F, Ikenoue T, Kato N, Shiratori Y, Omata M. 1998. Major virulence factors, VacA and CagA, are commonly positive in *Helicobacter pylori* isolates in Japan. *Gut* 42(3):338–343 DOI 10.1136/gut.42.3.338.
- Mahdavi J, Sondén B, Hurtig M, Olfat FO, Forsberg L, Roche N, Angstrom J, Larsson T, Teneberg S, Karlsson KA, Altraja S, Wadström T, Kersulyte D, Berg DE, Dubois A, Petersson C, Magnusson KE, Norberg T, Lindh F, Lundskog BB, Arnqvist A. 2002. *Helicobacter pylori* SabA adhesin in persistent infection and chronic inflammation. *Science* 297(5581):573–578 DOI 10.1126/science.1069076.
- Matsuo Y, Kido Y, Yamaoka Y. 2017. *Helicobacter pylori* outer membrane protein-related pathogenesis. *Toxins* 9(3):101 DOI 10.3390/toxins9030101.
- Matta AJ, Pazos AJ, Bustamante-Rengifo JA, Bravo LE. 2017. Genomic variability of *Helicobacter pylori* isolates of gastric regions from two Colombian populations. *World Journal of Gastroenterology* 23(5):800–809 DOI 10.3748/wjg.v23.i5.800.
- Moodley Y, Linz B, Yamaoka Y, Windsor HM, Breurec S, Wu JY, Maady A, Bernhöft S, Thiberge JM, Phuanukoonnon S, Jobb G, Siba P, Graham DY, Marshall BJ, Achtman M. 2009. The peopling of the Pacific from a bacterial perspective. *Science* 323(5913):527–530 DOI 10.1126/science.1166083.
- Muñoz Ramírez ZY, Mendez-Tenorio A, Kato I, Bravo MM, Rizzato C, Thorell K, Torres R, Aviles-Jimenez F, Camorlinga M, Canzian F, Torres J. 2017. Whole genome sequence and phylogenetic analysis show *Helicobacter pylori* strains from Latin America have followed a unique evolution pathway. *Frontiers in Cellular and Infection Microbiology* 7:50 DOI 10.3389/fcimb.2017.00050.
- Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky, Pond SL. 2012. Detecting individual sites subject to episodic diversifying selection. *PLOS Genetics* 8(7):e1002764 DOI 10.1371/journal.pgen.1002764.
- Nomura AMY, Perez Perez GI, Lee J, Stemmermann G, Blaser MJ. 2002. Relation between *Helicobacter pylori* cagA status and risk of peptic ulcer disease. *American Journal of Epidemiology* 155(11):1054–1059 DOI 10.1093/aje/155.11.1054.
- Odenbreit S, Swoboda K, Barwig I, Ruhl S, Borén T, Koletzko S, Haas R. 2009. Outer membrane protein expression profile in *Helicobacter pylori* clinical isolates. *Infection and Immunity* 77(9):3782–3790 DOI 10.1128/IAI.00364-09.
- Odenbreit S, Till M, Hofreuter D, Faller G, Haas R. 1999. Genetic and functional characterization of the alpAB gene locus essential for the adhesion of *Helicobacter pylori* to human gastric tissue. *Molecular Microbiology* 31(5):1537–1548 DOI 10.1046/j.1365-2958.1999.01300.x.

- Oleastro MR, Cordeiro R, Ménard A, Yamaoka Y, Queiroz D, Mégraud F, Monteiro L. 2009a.** Allelic diversity and phylogeny of homB, a novel co-virulence marker of *Helicobacter pylori*. *BMC Microbiology* **9**:248 DOI [10.1186/1471-2180-9-248](https://doi.org/10.1186/1471-2180-9-248).
- Oleastro MR, Cordeiro Y, Yamaoka D, Queiroz F, Mégraud L, Monteiro, Ménard A. 2009b.** Disease association with two *Helicobacter pylori* duplicate outer membrane protein genes, homB and homA. *Gut Pathogens* **1**(1):12 DOI [10.1186/1757-4749-1-12](https://doi.org/10.1186/1757-4749-1-12).
- Orsi RH, Sun Q, Wiedmann M. 2008.** Genome-wide analyses reveal lineage specific contributions of positive selection and recombination to the evolution of *Listeria monocytogenes*. *BMC Evolutionary Biology* **8**:233 DOI [10.1186/1471-2148-8-233](https://doi.org/10.1186/1471-2148-8-233).
- Parrish CR, Holmes EC, Morens DM, Park EC, Burke DS, Calisher CH, Laughlin CA, Saif LJ, Daszak P. 2008.** Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiology and Molecular Biology Reviews* **72**(3):457–470 DOI [10.1128/MMBR.00004-08](https://doi.org/10.1128/MMBR.00004-08).
- Parsonnet J, Friedman GD, Orentreich N, Vogelman H. 1997.** Risk for gastric cancer in people with CagA positive or CagA negative *Helicobacter pylori* infection. *Gut* **40**(3):297–301 DOI [10.1136/gut.40.3.297](https://doi.org/10.1136/gut.40.3.297).
- Pawlowski M, Gajda MJ, Matlak R, Bujnicki JM. 2008.** MetaMQAP: a meta-server for the quality assessment of protein models. *BMC Bioinformatics* **9**:403 DOI [10.1186/1471-2105-9-403](https://doi.org/10.1186/1471-2105-9-403).
- Perez-Perez GI, Rothenbacher D, Brenner H. 2004.** Epidemiology of *Helicobacter pylori* infection. *Helicobacter* **9**(Suppl. 1):1–6.
- Saitou N, Nei M. 1987.** The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**(4):406–425.
- Santoyo G, Romero D. 2005.** Gene conversion and concerted evolution in bacterial genomes. *FEMS Microbiology Letters* **29**(2):169–183.
- Sawyer S. 1989.** Statistical tests for detecting gene conversion. *Molecular Biology and Evolution* **6**(5):526–538.
- Slatkin M. 1985.** Gene flow in natural populations. *Annual Review of Ecology and Systematics* **16**:393–430 DOI [10.1146/annurev.es.16.110185.002141](https://doi.org/10.1146/annurev.es.16.110185.002141).
- Slatkin M. 1987.** Gene flow and the geographic structure of natural populations. *Science* **236**:787–792 DOI [10.1126/science.3576198](https://doi.org/10.1126/science.3576198).
- Suerbaum S, Josenhans C. 2007.** *Helicobacter pylori* evolution and phenotypic diversification in a changing host. *Nature Reviews. Microbiology* **5**:441–452.
- Sun ZL, Bi YW, Bai CM, Gao DD, Li ZH, Dai ZX, Li JF, Xu WM. 2010.** Expression of *Helicobacter pylori* alpA gene in *Lactococcus lactis* and its immunogenicity analysis. *Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi* **26**(3):203–206.
- Tajima F. 1989.** Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**(3):585–595.
- Talarico S, Whitefield SE, Fero J, Haas R, Salama NR. 2012.** Regulation of *Helicobacter pylori* adherence by gene conversion. *Molecular Microbiology* **84**(6):1050–1061 DOI [10.1111/j.1365-2958.2012.08073.x](https://doi.org/10.1111/j.1365-2958.2012.08073.x).

- Templeton AR. 2007.** Shared history of humans and gut bacteria: evolutionary togetherness: coupled evolution of humans and a pathogen. *Heredity* **98**(6):337–338 DOI [10.1038/sj.hdy.6800977](https://doi.org/10.1038/sj.hdy.6800977).
- Thompson LJ, Danon SJ, Wilson JE, O'Rourke JL, Salama NR, Falkow S, Mitchell H, Lee A. 2004.** Chronic *Helicobacter pylori* infection with Sydney strain 1 and a newly identified mouse-adapted strain (Sydneystrain2000) in C57BL/6 and BALB/c mice. *Infection and Immunity* **72**:4668–4679 DOI [10.1128/IAI.72.8.4668-4679.2004](https://doi.org/10.1128/IAI.72.8.4668-4679.2004).
- Thorell K, Hosseini S, Palacios Gonzáles RV, Chaotham C, Graham DY, Paszat L, Rabeneck L, Lundin SB, Nookaew I, Sjöling Å. 2016.** Identification of a Latin American-specific BabA adhesin variant through whole genome sequencing of *Helicobacter pylori* patient isolates from Nicaragua. *BMC Evolutionary Biology* **16**:53 DOI [10.1186/s12862-016-0619-y](https://doi.org/10.1186/s12862-016-0619-y).
- Thorell K, Yahara K, Berthenet E, Lawson DJ, Mikhail J, Kato I, Mendez A, Rizzato C, Bravo MM, Suzuki R, Yamaoka Y, Torres J, Sheppard SK, Falush D. 2017.** Rapid evolution of distinct *Helicobacter pylori* subpopulations in the Americas. *PLOS Genetics* **13**(2):e1006546 DOI [10.1371/journal.pgen.1006546](https://doi.org/10.1371/journal.pgen.1006546).
- Torres J, Correa P, Ferreccio C, Hernandez-Suarez G, Herrero R, Cavazza-Porro M, Dominguez R, Morgan D. 2013.** Gastric cancer incidence and mortality is associated with altitude in the mountainous regions of Pacific Latin America. *Cancer Causes Control* **24**(2):249–256 DOI [10.1007/s10552-012-0114-8](https://doi.org/10.1007/s10552-012-0114-8).
- Van Doorn LJ, Figueiredo C, Mégraud F, Pena S, Midolo P, Queiroz DM, Carneiro F, Vanderborcht B, Pegado MD, Sanna R, De Boer W, Schneeberger PM, Correa P, Ng EK, Atherton J, Blaser MJ, Quint WG. 1999.** Geographic distribution of vacA allelic types of *Helicobacter pylori*. *Gastroenterology* **116**(4):823–830 DOI [10.1016/S0016-5085\(99\)70065-X](https://doi.org/10.1016/S0016-5085(99)70065-X).
- Wertheim JO, Murrell B, Smith MD. 2015.** RELAX: detecting relaxed selection in a phylogenetic framework. *Molecular Biology and Evolution* **32**(3):820–832 DOI [10.1093/molbev/msu400](https://doi.org/10.1093/molbev/msu400).
- Xue J, Bai Y, Chen Y, Wang JD, Zhang ZS, Zhang YL, Zhou DY. 2005.** Expression of *Helicobacter pylori* AlpA protein and its immunogenicity. *World Journal of Gastroenterology* **11**(15):2260–2263 DOI [10.3748/wjg.v11.i15.2260](https://doi.org/10.3748/wjg.v11.i15.2260).
- Yakirevich E, Resnick MB. 2013.** Pathology of gastric cancer and its precursor lesions. *Gastroenterology Clinics of North America* **42**(2):261–284 DOI [10.1016/j.gtc.2013.01.004](https://doi.org/10.1016/j.gtc.2013.01.004).
- Yamaoka Y. 2009.** *Helicobacter pylori* typing as a tool for tracking human migration. *Clinical Microbiology and Infection* **15**(9):829–834 DOI [10.1111/j.1469-0691.2009.02967.x](https://doi.org/10.1111/j.1469-0691.2009.02967.x).
- Yamaoka Y, Kikuchi S, El-Zimaity HMT, Gutierrez O, Osato MS, Graham DY. 2002a.** Importance of *Helicobacter pylori* oipA in clinical presentation, gastric inflammation, and mucosal interleukin 8 production. *Gastroenterology* **123**(2):414–424 DOI [10.1053/gast.2002.34781](https://doi.org/10.1053/gast.2002.34781).
- Yamaoka Y, Orito E, Mizokami M, Gutierrez O, Saitou N, Kodama T, Osato MS, Kim JG, Ramirez FC, Mahachai V, Graham DY. 2002b.** *Helicobacter pylori*

in North and South America before Columbus. *FEBS Letters* **517(1–3)**:180–184
[DOI 10.1016/S0014-5793\(02\)02617-0](https://doi.org/10.1016/S0014-5793(02)02617-0).

Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER Suite: protein structure and function prediction. *Nature Methods* **12(1)**:7–8
[DOI 10.1038/nmeth.3213](https://doi.org/10.1038/nmeth.3213).