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

Andrés Julián Gutiérrez-Escobar Corresponding Author, 1, 2, Gina Méndez-Callejas 3, Orlando Acevedo 4, Maria Mercedes Bravo 5

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 Investigaciones Biomédicas y Genética Humana Aplicada - GIBGA. Programa de medicina, Universidad de Ciencias Aplicadas y Ambientales U.D.C.A., Bogotá, Colombia
4 Grupo de Biofísica y Bioquímica Estructural, Facultad de Ciencias, Pontifica Universidad Javeriana, Bogotá, Colombia
5 Grupo de Investigación en Biología del Cáncer, Instituto Nacional de Cancerología de Colombia, Bogotá, Colombia

To be able to survive, *H. 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.
Rapid evolution of the *Helicobacter pylori* AlpA adhesin in a high cancer gastric risk region from Colombia

Andrés Julián Gutiérrez-Escobar¹*, Gina Méndez-Callejas¹, Orlando Acevedo², María Mercedes Bravo³

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.

* Doctorado en Ciencias Biológicas, Pontificia Universidad Javeriana, Bogotá, Colombia.

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

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

Corresponding Author:
Email address: andresjulian1981@gmail.com
Abstract

To be able to survive, \textit{H. 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 \textit{alpA} have not been described in Colombia. The study aim was to describe the variation patterns and microevolutionary process of \textit{alpA} in Colombian clinical isolates of \textit{H. pylori}. The existing polymorphisms, which are deviations from the neutral model of molecular evolution, and the genetic differentiation of the \textit{alpA} gene from Colombian clinical isolates of \textit{H. pylori} were determined. The analysis shows that gene conversion and purifying selection have shaped the evolution of three different variants of \textit{alpA} in Colombia.

1. Introduction

\textit{H. pylori}, a Gram-negative bacterium, has persistently colonized the stomach of half of the human population (Pérez - Pérez GI et al., 2004; Khalifa MM et al., 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 E, Resnick MB, 2013).

Gastric cancer is the fifth most common cancer worldwide (Jemal et al., 2010; Bertuccio et al., 2009; Forman et al., 2014), and it is the second leading cause of cancer deaths (Ferlay et al., 2013); an infection with \textit{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 & Piazuelo, 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 B et al., 2007). The bacterium has followed a similar dispersal pattern as its human host. Currently, seven populations of \textit{H. pylori} have been identified: hpEurope, hpNEAfrica, hpAfrica1, hpAfrica2, hpAsia2, hpSahul and hpEastAsia (Yamaoka Y, 2009; Falush D et al., 2003; Achtman M et al., 1999; Moodley Y et al., 2009; Devi SM 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 \textit{H. pylori}, which caused the disappearance of 80% of the native population (Bianchine PJ, Russo TA, 1992; Parrish CR et al., 2008). It has been reported that \textit{H. pylori} has followed unique evolutionary pathways in Latin-America (Gutiérrez-Escobar AJ et al., 2017; Muñoz-Ramírez ZY et al., 2017) and that the strains followed rapid adaptive processes in different countries of the region (Thorell K et al., 2017), establishing local independent lineages.

The bacterium has virulence factors that correlate with the risk of developing gastric diseases (Parsonnet J et al., 1997; Nomura AMY et al., 2002; Mahdavi J et al., 2002; Yamaoka Y et al.,

2002). *H. pylori* attach to the gastric epithelium thorough adhesins that contribute to the initial steps of the infection (Matsuo Y, Kido Y, 2017). AlpA (~56 kDa) is an adhesin that is encoded by the locus *alpAB* (Alm R.A et al., 2000), which is essential for adherence to the human gastric epithelium (Odenbreit S et al., 1999). This adhesin is expressed by all clinical isolates (Odenbreit S, 2009) and is recognized in sera from infected patients (Xue J, 2005), and in addition, it induces the secretion of interleukin-8 (IL-8) (Lu H 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 Z.L 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.

2. Materials and methods

2.1. 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, 5 of Gastritis concomitant with Duodenal Ulcer (G-DU) and 2 of Intestinal Metaplasia concomitant with Duodenal Ulcer (IM-DU) (Gutiérrez-Escobar AJ 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.

2.2. 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 RC, 2004); the evolutionary model and the phylogenetic reconstruction was determined using MEGA V 7 (Kumar S et al., 2016) and the NJ algorithm (Saitou N, Nei M, 1987) with 1000 bootstrap repetitions for statistical robustness.

2.3. 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 (θw) and recombination events (Rm) analyses were performed using DnaSP v 5.10 (Librado P, Rozas J, 2009). Deviations of the neutral model of molecular evolution were tested using the Tajima test.
(Tajima F, 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 Junkes-Cantor correction. The variance of the difference was computed using the bootstrap method (1000 replicates) using Mega v 7 (Kumar S et al., 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 P, Rozas J, 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 1000 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, 1987) using the software DnaSP v 5.10 (Librado P, Rozas J, 2009).

2.4. Gene conversion analysis of alpA

To test whether gene conversion generates genetic diversity in the alpA, the Betran’s method (Betran E 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 S, 1989). Only conversion tracks with $p < 0.05$ were considered.

2.5. 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 J et al., 2015) and tested using the MetaMQAPII server (Pawlowski M et al., 2008). DIVERGE V 3 software (Gu X 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 X, 1999). The software tested whether a significant change in the evolution rate has occurred by calculating the coefficient of divergence ($\theta_D$). 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 RC, 2004), and the alignment file was pre-processing by screening for recombination breakpoints using the GARD algorithm implemented by the HyPhy software (Kosakovsky P et al., 2006a, b). Then, the processed file was tested for selection using the FEL and IFEL algorithms using the datamonkey server (Kosakovskyi PSL, Frost SD, 2005). Episodic diversifying selection was detected using the MEME algorithm (Murrell B 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 (Kosakovskyi PSL, Frost SD, 2005).
3. 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 (Figure 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 (Figure 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 (Figure 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 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 \( \omega \) value above 1, which suggests that episodic positive diversifying selection is also shaping the microevolutionary patterns of this gene in Colombia (Figure 4).

The test of genetic differentiation showed that the Colombian population was well-differentiated from the reference pool, but as expected, the Nm obtained from FsT was 3.63, which indicates a moderate gene flow between the Colombian isolates (Slatkin, 1985, 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 6 tracks of gene conversion between the sequences that belong exclusively to the Colombian isolates (Figure 5).

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 θD > 0.8 (Figure 3B), which indicates that the protein presents sites with different evolutionary rates. The analysis of positive and negative selection covered 540 codon sites of 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 (Figure 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).

4. 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 P et al., 2005; Maeda S et al., 1998; Oleastro M et al., 2009a; Van Doorn LJ 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 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 G and Romero D, 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 MS 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 Y et al., 2002) 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 and Berg, 2001; Blaser and 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 ZY et al., 2017), and the strains have
followed rapid adaptive processes in different countries of the region (Gutiérrez-Escobar AJ et
al., 2017; Thorell K 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 A et al., 1999; Oleastro M 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 MS et al., 2013; Dubois A et al., 1999; Oleastro
M et al., 2009a; Oleastro M et al., 2009b; Santoyo G Romero D, 2005; Yamaoka Y et al., 2002),
sabB and omp27 (Talarico S et al., 2012) and *babA* (Hennig EE et al., 2006), and a Latin-
American type of *babA* has also been recently reported (Thorell K et al., 2016).

The interplay between positive selection and recombination has been detected in bacterial
genomes (Lefebure & Stanhope, 2007; Joseph et al., 2011; Orsi et al., 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 0D 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 M, Conery JS, 2000) or lead to evolutionary novelties by positive
selection (Wertheim JO et al., 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, \( p = 5.48 \times 10^{-10}\),
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 N 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 Q et al., 2015; Furuta Y 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 AR, 2007).

6. 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.

5. Acknowledgements

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

6. References

Achtman M, Azuma T, Berg DE, Ito Y, Morelli G, Pan ZJ, Suerbaum S, Thompson SA, van der Ende A and van Doorn LJ. 1999. Recombination and clonal groupings within Helicobacter pylori from different geographical regions. Mol Microbiol 32(3): 459-70.

Alm RA, Bina J, Andrews, BM, Doig P, Hancock RE and Trust TJ. 2000. Comparative genomics of Helicobacter pylori: Analysis of the outer membrane protein families. Infect Immun 68(7): 4155–68.
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, Int. J. Cancer 125 (3): 666– 673.

Betran E, Rozas J, Navarro A and 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 and Russo TA. 1992. The role of epidemic infectious diseases in the discovery of America. Allergy Proc 13(5): 225-32.

Blaser MJ & Atherton JC. 2004. Helicobacter pylori persistence: biology and disease. J. Clin.Invest. 113, 321–333.

Blaser MJ & Berg DE. 2001. Helicobacter pylori genetic diversity and risk of human disease. J. Clin. Invest. 107, 767–773.

Cao P, Lee KJ, Blaser MJ and Cover TL. 2005. Analysis of hopQ alleles in East Asian and Western strains of Helicobacter pylori. FEMS Microbiol Lett 251(1): 37-43.

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 and Zhang J. 2015. Progressive genomic convergence of two Helicobacter pylori strains during mixed infection of a patient with chronic gastritis. Gut 64(4): 554-61.

Correa P & Piazuelo B. (2010). Gastric cancer: The colombian enigma. Revista Colombiana de Gastroenterología, 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.

Dorer MS, Cohen IE, Sessler TH, Fero J and Salama NR. 2013. Natural competence promotes Helicobacter pylori chronic infection. Infect Immun 81(1): 209-15.

Dubois, A, Berg DE, Incecik ET, Fiala N, Heman-Ackah LM, Del Valle J, Yang M, Wirth HP, Perez-Perez GI and Blaser MJ. 1999. Host specificity of Helicobacter pylori strains and host responses in experimentally challenged non-human primates. Gastroenterology 116(1): 90–6.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5): 1792-7.

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 and Suerbaum S. 2003. Traces of human migrations in Helicobacter pylori populations. Science 299(5612): 1582-5.
Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. 2013. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC Cancer Base No.11, International Agency for Research on Cancer, Available from: URL: 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.

Furuta Y, Konno M, Osaki T, Yonezawa H, Ishige T, Imai M, Shiwa Y, Shibata-Hatta M, Kanesaki Y, Yoshikawa H, Kamiya S and Kobayashi I. 2015. Microevolution of virulence-related genes in Helicobacter pylori familial infection. PLoS One 10(5): e0127197.

Gu X, Zou Y, Su Z, Huang W, Zhou Z, Arendsee Z and Zeng Y. 2013. An update of DIVERGE software for functional divergence analysis of protein family. Mol Biol Evol 30(7): 1713–9.

Gu X. Statistical methods for testing functional divergence after gene duplication.1999. Mol Biol Evol 16(12): 1664–74.

Gutiérrez-Escobar AJ, Trujillo E, Orlando E. Acevedo S, Bravo MM. 2017. Phylogenomics of Colombian Helicobacter pylori isolates. Gut Pathog, 9:52–61.

Hartl DL & Clark GC. 1997. Principles of Population Genetics. Sinauer Associates,Sunderland.

Hennig EE, Allen JM and Cover TL. 2006. Multiple chromosomal loci for the babA gene in Helicobacter pylori. Infect Immun 74(5): 3046–3051.

Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends, Cancer Epidemiol. Biomarker Prev. 19 (8) (2010) 1893–1907.

Joseph SJ, Didelot X. Gandhi K, Dean D, Read TD. 2011. Interplay of recombination and selection in the genomes of Chlamydia trachomatis, Biol. Direct., 6, 28.

Khalifa MM, Sharaf RR and Aziz RK. 2010. Helicobacter pylori: a poor man's gut pathogen? Gut Pathog 2: 2.

Kodaman N, Pazos A, Schneider BG, Piazuelo MB, Mera R, Sobota RS, Sicinschi LA, Shaffer CL, Romero-Gallo J, de Sablet T, Harder RH, Bravo LE, Peek RM Jr, Wilson KT, Cover TL, Williams SM and Correa P. 2014. Human and Helicobacter pylori coevolution shapes the risk of gastric disease. Proc Natl Acad Sci U S A 111(4): 1455-1460.
Kosakovsky Pond SL and Frost SD. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol 22(5): 1208–22.

Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. 2006. GARD: a genetic algorithm for recombination detection. Bioinformatics 22: 3096–3098.

Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol 23 (10):1891-901.

Kumar S Stecher G, Tamura K . 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger data sets. Mol Biol Evol. 33(7):1870–1874.

Lefebure T. & Stanhope MJ. 2007. Evolution of the core and pangenome of Streptococcus: positive selection, recombination, and genome composition, Genome Biol., 8, R71.

Librado P and Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25(11): 1451-2.

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 and Achtman M. 2007. An African origin for the intimate association between humans and Helicobacter pylori. Nature 445(7130): 915-918.

Lu H, Wu JY, Beswick EJ, Ohno T, Odenbreit S, Haas R, Reyes VE, Kita M, Graham DY and Yamaoka Y. 2007. Functional and intracellular signalling differences associated with the Helicobacter pylori AlpAB adhesin from Western and East Asian strains. J Biol Chem 282(9): 6242–54.

Lynch M and Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science 290(5494): 1151-5.

Maeda S, Ogura K, Yoshida H, Kanai F, Ikenoue T, Kato N, Shiratori Y and Omata M. 1998. Major virulence factors, VacA and CagA, are commonly positive in Helicobacter pylori isolates in Japan. Gut 42(3): 338-43.

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, Arnvist A, Hammarström L and Borén T. 2002. Helicobacter pylori SabA adhesin in persistent infection and chronic inflammation. Science 297(5581): 573-8.

Matsuo Y, Kido Y, Yamaoka Y. 2017. Helicobacter pylori outer membrane protein-related pathogenesis. Toxins 9(3): 101.
Matta AJ, Pazos AJ, Bustamante-Rengifo JA, Bravo LE. Genomic variability of Helicobacter pylori isolates of gastric regions from two Colombian populations. World J Gastroenterol. 2017 Feb 7;23(5):800-809.

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 and Achtman M. 2009. The peopling of the Pacific from a bacterial perspective. Science 323(5913): 527-30.

Muñoz-Ramírez ZY, Mendez-Tenorio A, Kato I, Bravo MM, Rizzato C, Thorell K, Torres R, Aviles-Jimenez F, Camorlinga M, Canzian F and Torres J. 2017. Whole genome sequence and phylogenetic analysis show Helicobacter pylori strains from Latin America have followed a unique evolution pathway. Front Cell Infect Microbiol 7: 50.

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky and Pond SL. 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet 8(7): e1002764.

Nomura AMY, Perez Perez GI, Lee J, Stemmermann G, Blaser MJ. 2002. Relation between Helicobacter pylori cagA status and risk of peptic ulcer disease. Am J Epidemiol 155(11): 1054-9.

Odenbreit S, Swoboda K, Barwig I, Ruhl S, Borén T, Koletzko S and Haas R. 2009. Outer membrane protein expression profile in Helicobacter pylori clinical isolates. Infect Immun 77(9): 3782–90.

Odenbreit S, Till M, Hofreuter D, Faller G and Haas R. 1999. Genetic and functional characterization of the alpAB gene locus essential for the adhesion of Helicobacter pylori to human gastric tissue. Mol Microbiol 31(5): 1537–48.

Oleastro MR, Cordeiro R, Ménard A, Yamaoka Y, Queiroz D, Mégraud F and Monteiro L. 2009a. Allelic diversity and phylogeny of homB, a novel co-virulence marker of Helicobacter pylori. BMC Microbiol 9: 248.

Oleastro MR, Cordeiro Y, Yamaoka D, Queiroz F, Mégraud L, Monteiro and Ménard A. 2009b. Disease association with two Helicobacter pylori duplicate outer membrane protein genes, homB and homA. Gut Pathog 1(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 Evol. Biol., 8, 233.

Parrish CR, Holmes EC, Morens DM, Park EC, Burke DS, Calisher CH, Laughlin CA, Saif LJ and Daszak P. 2008. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiol Mol Biol Rev 72(3): 457-470.

Parsonnet J, Friedman GD, Orentreich N and Vogelman H. 1997. Risk for gastric cancer in people with CagA positive or CagA negative Helicobacter pylori infection. Gut 40(3): 297-301.
Pawlowski M, Gajda MJ, Matlak R and Bujnicki JM. 2008. MetaMQAP: a meta-server for the quality assessment of protein models. BMC Bioinformatics 9: 403.

Perez-Perez GI, Rothenbacher D and Brenner H. 2004. Epidemiology of Helicobacter pylori infection. Helicobacter 9 (Suppl. 1): 1-6.

Saitou N and Nei M. 1987. The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4): 406-25.

Santoyo G and Romero D. 2005. Gene conversion and concerted evolution in bacterial genomes. FEMS Microbiol Lett 29(2): 169-83.

Sawyer S. 1989. Statistical tests for detecting gene conversion. Mol Biol Evol 6(5): 526-38.

Slatkin M. 1985. Gene flow in natural populations. Ann. Rev. Ecol. Syst. 16:393-430.

Slatkin M. 1987. Gene flow and the geographic structure of natural populations. Science 236:787-792.

Suerbaum S. and Josenhans C. 2007. Helicobacter pylori evolution and phenotypic diversification in a changing host, Nat. Rev. Microbiol., 5, 441–52.

Sun ZL, Bi YW, Bai CM, Gao DD, Li ZH, Dai ZX, Li JF and 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): 203–6.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123(3): 585-95.

Talarico S, Whitefield SE, Fero J, Haas R and Salama NR. 2012. Regulation of Helicobacter pylori adherence by gene conversion. Mol Microbiol 84(6): 1050-61.

Templeton AR. 2007. Shared history of humans and gut bacteria: evolutionary togetherness: coupled evolution of humans and a pathogen. Heredity 98(6): 337–8.

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. Infect.Immun. 72, 4668–4679.

Thorell K, Hosseini S, Palacios Gonzáles RV, Chaotham C, Graham DY, Paszat L, Rabeneck L, Lundin SB, Nookaew I and Sjöling Å. 2016. Identification of a Latin American-specific BabA adhesin variant through whole genome sequencing of Helicobacter pylori patient isolates from Nicaragua. BMC Evol Biol 16: 53.
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 and Falush D. 2017. Rapid evolution of distinct Helicobacter pylori subpopulations in the Americas. PLoS Genet 13(2): e1006546.

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.

Van Doorn LJ, Figueiredo C, Mégraud F, Pena S, Midolo P, Queiroz DM, Carneiro F, Vanderborght B, Pegado MD, Sanna R, De Boer W, Schneeberger PM, Correa P, Ng EK, Atherton J, Blaser MJ and Quint WG. 1999. Geographic distribution of vacA allelic types of Helicobacter pylori. Gastroenterology 116(4):823-30.

Wertheim JO, Murrell B, Smith MD, Kosakovsky-Pond SL and Scheffler K. 2015. RELAX: detecting relaxed selection in a phylogenetic framework. Mol Biol Evol 32(3): 820–32.

Yang J, Yan R, Roy A, Xu D, Poisson J and Zhang Y. 2015. The I-TASSER Suite: Protein structure and function prediction. Nat Methods 12(1): 7-8.
Figure 1

Phylogenetic tree of AlpA from reference strains 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 (1000 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.
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 (1000 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 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.
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 (1000 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 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 θD > 0.8 indicating a strong signal of functional divergence.
Figure 4

Sliding windows analysis of $\omega$ rate.

Evolutionary rate $\omega = \frac{dN}{dS}$ 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.
Figure 5

Gene conversion tracks identified by the GENECONV method.

Each bar is a representation of a gene. 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.
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.
Table 1 (on next page)

Genetic diversity statistics for alpA.

The genetic diversity statistics were calculated by using thirty-four sequences from reference pool and 108 sequences from Colombian isolates of *Helicobacter pylori*. 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, θw: Watterson estimator, π: nucleotide diversity per site and RM: recombination events.
### Single Nucleotide Polymorphism

|        | n  | Sites | Ss  | k   | H   | Hd  | θw | π   | Rm |
|--------|----|-------|-----|-----|-----|-----|-----|-----|----|
| Reference | 34 | 1324  | 774 | 200.868 | 33 | 0.998 | 0.143 | 0.152 | 70 |
| Colombian | 108 | 1384  | 306 | 53.335 | 102 | 0.998 | 0.042 | 0.039 | 91 |
| Colombian and references | 142 | 1223  | 754 | 81.411 | 134 | 0.999 | 0.112 | 0.067 | 98 |
**Table 2** (on next page)

Deviations of the neutral model of molecular evolution.

The analysis involved 142 nucleotide sequences. Tajima test: Codon positions included were 1st. All positions containing gaps and missing data were eliminated. 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, \( ps = S/n, \) \( \Theta = ps/a1, \) \( \pi = \) 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 (*\( p < 0.05, **p < 0.01, ***p < 0.001; \) ns: not significative). 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 (1000 replicates). Analyses were conducted using the Nei-Gojobori method. All positions containing gaps and missing data were eliminated.
Table 2 Deviations of the neutral model of molecular evolution

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

The analysis involved 142 nucleotide sequences. Tajima test: Codon positions included were 1st. All positions containing gaps and missing data were eliminated. 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 = p_s/a_1, 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 < 0.05, **p < 0.01, ***p < 0.001; ns: not significantative). 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 (1000 replicates). Analyses were conducted using the Nei-Gojobori method. All positions containing gaps and missing data were eliminated.
Genetic heterogeneity and genetic flow were detected using the following tests: Hst, Kst, Kst*, Z, Z*. The HBK, Snn and chi squared tests were performed using the haplotype frequencies under the permutation of 1000 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 significative.
| Genetic differentiation | Genetic flow |
|-------------------------|-------------|
| Chi2                  | Gst         |
| Hst                   | GammaST    |
| Kst                   | Nst         |
| Kst*                  | Fst         |
| Z                     | 0.003***    |
| Z*                    | 0.090**     |
| Snn                   | 0.034***    |
| 1501.4 ns             | 4011.8***   |
| 0.316***              | 7.837***    |
| 0.011 (44.82)         | 0.011***    |
| 0.178 (2.30)          | 0.178 (2.30)|
| 0.114 (3.88)          | 0.114 (3.88)|
| 0.120 (3.63)          | 0.120 (3.63)|
**Table 4** (on next page)

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

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). Significative FsT values are highlighted in bold.
### Populations

|        | Amerind | FsT  | GammaST |
|--------|---------|------|---------|
| 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

|        | Europe | FsT  | GammaST |
|--------|---------|------|---------|
| 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

|        | Asia | FsT  | GammaST |
|--------|------|------|---------|
| 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

|        | Africa | FsT  | GammaST |
|--------|--------|------|---------|
| 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|