The *Helicobacter pylori* Genome Evolution in Different Gastric Cancer Risk Colombian Populations

**Abstract**

**Background:** The *Helicobacter pylori* (*H. pylori*) has evolved with its human host by nearly 110,000 years. Despite that *H. pylori* has been considered as a main factor for gastric cancer (GC) development, the pathogenesis depends on its hosts evolutive relations.

**Objective:** In this study we analyzed the *H. pylori* evolutive relations of two populations with different GC risk in Colombia.

**Materials and methods:** We study 10 human genomes and same number of *H. pylori* genomes from Tuquerres: high GC risk population, and 9 genomes from Tumaco: low GC risk population. The evolutive analysis was performed using MLST, *vacA* virulence gene and *alpA* adhesine gene for *H. pylori* and human ancestry by phylogenomic analyzes.

**Results:** We found that the studied people from Tumaco had marked African and Amerindian ancestry and in minor proportion European ancestry. In contrast, the studied human population from Tuquerres had mainly Amerindian and European ancestry. The *H. pylori* phylogenomic trees from Tumaco were grouped with African strains (*hspWAfrica y hpAfrica2*) in 56% and with the Colombian evolutive group (*hspColombia*) in 44%. We found that the *H. pylori* genomes from Tumaco are in major proportion in co-evolution with its human host genomes. In Tuquerres the phylogenomic trees grouped in 80% with local *H. pylori* strains (*hspColombia*) and the 20% of genomes grouped with *hspWAfrica* ancestors. Also, we found that *H. pylori* from Tuquerres were in minor proportion in co-evolution with its human host genomes. In Tuquerres the *H. pylori* *vacA* and *alpA* genes showed phylogenetic relationship with Amerindian strains (*hspAmerindian*) and European (*hpEurope*), and in minor proportion with African strains (*hspWAfrica y hpAfrica2*) and Asian (*hpEAsia*).

**Conclusion:** The marked difference of GC risk in Colombian populations could be explained by the genome coevolution time between *Helicobacter pylori* and human host genomes.

**Keywords:** Evolution; Colombian populations; Gastric cancer; *H. pylori*; Phylogenomics

---

**Introduction**

The *Helicobacter pylori* (*H. pylori*) is a Gram-negative bacterium that has colonized half of the world’s population and which induces a chronic inflammatory process [1,2]. *H. pylori* is the main factor in the development of gastric cancer, which was classified as a type I carcinogen by World Health Organization (WHO) in 1994 [3]. The infection by *H. pylori* in the human host occurred 88,000-116,000 years ago [4], and has coevolved with human beings since the first migration from East Africa approximately 60,000 years ago [5].

The high genomic diversity of *H. pylori* is a product of intra and intergenomic mixing process from multiple strains that provides
its adaptation and colonization skills in human populations [6]. Its infection, adaptation, and survival mechanisms are diverse, and the vacA gene induces apoptosis, increased permeability in gastric cell, and causes inhibition of the immune response of T cells [7]. This gene shows high genetic diversity, presenting two allele families, s1/s2 and m1/m2, which are associated with the development of gastric cancer [8]. The AlpA adhesin is constitutive of H. pylori this has as main function the adhesion to gastric epithelium [9]. Also, this gene could induce the IL-6 and IL-8 expression in human host [10].

The incidence of gastric cancer in Latin America has been characterized by being high in the mountains and low on the coasts [11]. In Colombia, in the Andean zone in the city of Tuquerres, the incidence rate is 150/100,000 inhabitants, while on the Pacific coast, in the city of Tumaco it is 6/100,000 inhabitants, in spite of a similar incidence of H. pylori (~90%) [12,13]. Likewise, human populations have different ancestry: in the Andean zone 67% is Native American, 31% is European, and 2% is African. On the other hand, the ancestry of the people in Tumaco is 58% African, 23% Native American, and 19% European [14]. This phenomenon is known as the “Colombian enigma” [12].

Multilocus Sequence Typing (MLST) is a technique that has been used to estimate the evolutionary relations among different strains and to study historical migrations of H. pylori and its host. Some studies that used the MLST technique on H. pylori based on seven housekeeping genes have been able to identify various population groups: hpEurope, hpNEAfrica, hpWAfrica, hpAfrica1, hpAfrica2, hpAsia2, hpSahul, and hpEastAsia. Bacteria subpopulations have been identified: hspAmerindian, hspEAsia, and hspMaori [15-18].

In previous studies it is being found that bacteria that infected human population from American Continent and Colombia in particular belong to European origin (hpEurope) being an important risk factor due to the human host-bacteria evolutive desynchronization [19,20]. However, H. pylori whole genome recent studies have shown the emergence of new independent lineages for several countries from Latin America [21-23]. One of the characteristics of H. pylori is its great genetic diversity, and it is known that different ancestry strains might interact differently with their human host clearly influencing pathogenesis [24]. Therefore, to describe the evolutive process of H. pylori could allow us to obtain information about the risk of cancer development. To develop our main aim of this study, we suggested carrying out an analysis of the evolutionary relations of H. pylori from the department of Narino, Colombia, who had a different risk of gastric cancer.

Materials and Methods

Subjects and initial bioinformatics data of Helicobacter pylori

DNA blood samples and gastric biopsies were taken from 10 patients with an age average of ~40 years, from the Andean zone of Tuquerres, and 9 samples from patients from the Pacific coast from the city of Tumaco, Colombia. The sequences were annotated using the prokaryote genome annotation from NCBI. The genomes correspond to sequenced data by our group in collaboration with Valderlbilt University and available on NCBI data base (Table 1) [25-30].

<table>
<thead>
<tr>
<th>Risk</th>
<th>Isolate ID</th>
<th>Access code</th>
<th>Diagnosis</th>
<th>Genome size</th>
<th>N° of contigs</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR*</td>
<td>SV328_2</td>
<td>MTW0000000000</td>
<td>Dysc</td>
<td>1,645,479</td>
<td>56</td>
</tr>
<tr>
<td>HR</td>
<td>SV340_2</td>
<td>MTW0000000000</td>
<td>NAGd</td>
<td>1,633,298</td>
<td>53</td>
</tr>
<tr>
<td>HR</td>
<td>SV355_2</td>
<td>MTW0000000000</td>
<td>IMFe</td>
<td>1,635,504</td>
<td>53</td>
</tr>
<tr>
<td>HR</td>
<td>SV376_1</td>
<td>MTW0000000000</td>
<td>IM</td>
<td>1,650,127</td>
<td>207</td>
</tr>
<tr>
<td>HR</td>
<td>SV380_1</td>
<td>MTW0000000000</td>
<td>IM</td>
<td>1,635,819</td>
<td>40</td>
</tr>
<tr>
<td>HR</td>
<td>SV397_2</td>
<td>MTW0000000000</td>
<td>NAG</td>
<td>1,668,205</td>
<td>47</td>
</tr>
<tr>
<td>HR</td>
<td>SV449_1</td>
<td>MTW0000000000</td>
<td>IM</td>
<td>1,654,884</td>
<td>41</td>
</tr>
<tr>
<td>HR</td>
<td>PZ5056*</td>
<td>ASU0000000000</td>
<td>NAG</td>
<td>1,578,164</td>
<td>335</td>
</tr>
<tr>
<td>HR</td>
<td>PZ5080*</td>
<td>ASY0000000000</td>
<td>IM</td>
<td>1,597,127</td>
<td>283</td>
</tr>
<tr>
<td>HR</td>
<td>PZ5086*</td>
<td>ASY0000000000</td>
<td>NAG</td>
<td>1,547,845</td>
<td>295</td>
</tr>
<tr>
<td>LR*</td>
<td>PZ5005_3A3</td>
<td>MTU0000000000</td>
<td>NAG</td>
<td>1,672,956</td>
<td>51</td>
</tr>
<tr>
<td>LR</td>
<td>PZ5006_3A3</td>
<td>MTU0000000000</td>
<td>NAG</td>
<td>1,642,170</td>
<td>53</td>
</tr>
<tr>
<td>LR</td>
<td>PZ5009_3A2</td>
<td>MSY0000000000</td>
<td>NAG</td>
<td>1,677,035</td>
<td>53</td>
</tr>
<tr>
<td>LR</td>
<td>PZ5016_3A3</td>
<td>MTW0000000000</td>
<td>MAGf</td>
<td>1,644,124</td>
<td>44</td>
</tr>
<tr>
<td>LR</td>
<td>PZ5019_3A3</td>
<td>MTW0000000000</td>
<td>IM</td>
<td>1,681,561</td>
<td>44</td>
</tr>
<tr>
<td>LR</td>
<td>PZ5033_3A2</td>
<td>MTW0000000000</td>
<td>IM</td>
<td>1,656,908</td>
<td>60</td>
</tr>
<tr>
<td>LR</td>
<td>PZ5004*</td>
<td>ASF0000000000</td>
<td>NAG</td>
<td>1,569,902</td>
<td>303</td>
</tr>
<tr>
<td>LR</td>
<td>PZ5024*</td>
<td>ASY0000000000</td>
<td>NAG</td>
<td>1,496,849</td>
<td>413</td>
</tr>
<tr>
<td>LR</td>
<td>PZ5026*</td>
<td>ASY0000000000</td>
<td>NAG</td>
<td>1,604,992</td>
<td>253</td>
</tr>
</tbody>
</table>

Note: *High Risk: Host is resident of Tuquerres where risk for gastric cancer is high; †Low Risk: Host is resident of Tumaco where risk for gastric cancer is low; Dys, dysplasia; NAG, nonatrophic gastritis; IM, intestinal metaplasia; Multifocal atrophic gastritis; *Data samples from previous work by Sheh et al. [29].

Table 1: Complete genome description of H. pylori strains in NCBI from two regions of the department of Nariño, Colombia.
The blood and gastric biopsies were coding as follow, for Tuquerres samples (SV328_2, SV340_2, SV355_2, SV376_1, SV380_1, SV397_2, SV449_1), for Tumaco (PZ5005_3A3, PZ5006_3A3, PZ5009_3A2, PZ5016_3A3, PZ5019_3A3, PZ5033_3A2) (29). All participants provided informed consent; the study was approved by the institutional and local hospitals review boards. The bioinformatics analysis were performed during January and February of 2021. The human samples were genotyped using an Immunochip previously reported [25], which identifies around 196 x 103 SNPs in genes involved in immune disorders. The Admixture model of STRUCTURE assuming correlated allele frequencies, (50,000 iterations after a burn-in of 50,000 iterations).

The reference populations used in this study were published previously in Human Genome Diversity Project that content European, Amerindian and African ancestries [26,27]. The number of tentative populations (K) was set from 1 to 3 and 10 runs were executed for each K. The STRUCTURE results (mixing model) showed that the model probability was maximized in k=3 [14]. CLUMPP was used to collate replicate runs and calculate means of individual ancestry [28].

To make the phylogenetic modelling we used African, Asian, European and Native American reference genomes [21]. We included genomes from populations of Managua (Nicaragua) and Mexico City (Mexico) belonging to previous study [23]. The genomes from Colombia were 7 isolated from Bogota (CG22366, CA22327, CA22331, CA22339, CA22312, CM22360, CM22351), 10 from Cundinamarca department (CC22402, CC26084, CC26093, CM23346, CA22337, CM23341, CG22389, CG22322, CA22393, CM22388), 14 from Boyaca department (CG22025, CG22087, CG22023, CG22046, CM22013, CG22367, CM22021, CG22370, CM22331, CM22368, CA22020, CM22315, CA22335, CM22347), one from Caldas department (CM22390), One from Caqueta department (CC26100), one from Meta (CA26024), four from Santander department (CG22385, CG22378, CA22019, CA22095) and two from Tolima department (CA22362, CA24004) [22].

**Multilocus Sequence Typing (MLST) analysis based on genomes**

The housekeeping genes *atpA, efp, mutY, ppa, trpC, ureI*, and *yphC* were annotated using PubMLST, and the sequences were selected, downloaded and concatenated. The concatenated sequences were aligned using Muscle software [31]. The phylogenetic analysis was constructed and calculated using a similarity analysis by means of Neighbor-joining [32] with the evolutionary model T92+G+I (Tamura with Gamma variation and invariable sites). The bootstrap analysis was done with 1000 replicates and the phylogenetic tree was edited in iTol v3.

**Helicobacter pylori phylogenomic analysis**

To the core genome analysis all the sequences were imported from bacteria isolated genome sequences database BIGSdb [33]. Then an alignment of gene by gene was done using *H. pylori* coding sequences CDS from African strain J99 as reference, and the alignment was exported from the database. The output matrix from the genome comparing obtained by BIGSdb was used to create the phylogenomic tree using MEGA V7 [34].

The phylogenomic analysis based on SNPs was carried out using CSI-phylogeny [35] with the default parameters. The genome assembly was analyzed with the following parameters: minimum depth at SNPs positions of 10; relative depth at SNPs positions of 10; minimum distance between SNPs (prune) of 10; minimum SNPs quality of 30; minimal read mapping quality of 25; minimum Z-score: 1.96 corresponding to a p<0.05 value. The reads were mapped to the reference genome J99 with BMW mem, and the SNPs were assigned with the mpileup tool from SAMTools [36]. The SNPs were filtered according to the assigned parameters to obtain a high-quality matrix. The SNPs matrix was created evaluating all the positions for each genome, which were concatenated creating a multiple FASTA file used in the Maximum-likelihood phylogenetic analysis, where we found 175,856 SNPs. The results were visualized and edited with FigTree v1.4.0 [37].

**VacA cytotoxin and AlpA adhesin phylogenetic analysis**

A phylogenetic analysis of virulence gene *vacA* and adhesine gene *alpA* were studied. The sequences were depurated and aligned using Muscle software [31]. We used the tool Gblocks [38] to determine the parsimony site due to the high diversity of the genes. The evolutive model that better adjusted to the alignment was the General Time Reversible GTR+G+I that shows a Bayesian information criterion BIC=125767.110, lnL=-60531.321. The variation rate between sites was modeled with Gamma distribution=0.57. The analysis involved 196 DNA sequences with long of 2948pb. The vacA gene the phylogenetic analysis was determined, created and calculated using Maximum-likelihood estimation and a bootstrap analysis of 1000 replicates for more statistical accuracy in PhyML v. 3.0 [39].

To the gene *alpA* the best evolutive model was determined by the General Time Reversible GTR+G+I with BIC=43727.101, lnL=-19161.291. The variation rate between sites was modeled with a Gamma distribution of 0.66. The analysis involved 215 DNA sequences with long of 1093pb. For the gene *alpA* we applied the Maximum-likelihood method along with a bootstrap of 1000 replicates in PhyML v. 3.0 [39]. The phylogenetic trees were visualized and edited with FigTree v.1.4.0 [33].

**Results**

**Human ancestry analysis**

In Colombia, during European colonization African and Spanish human population arrived then, they met Native American population in the American continent. The results from the human ancestry analysis showed that on in the Andean zone of Tuquerres, the population showed a marked Amerindian and European ancestry (Figure 1A). On the other hand, in the Pacific coast, patients presented a high ethnic combination, a great proportion of African together with an Amerindian and European mixture (Figure 1B), which corroborates the results obtained in a previous work [14].
Multilocus Sequence Typing (MLST) analysis based on *Helicobacter pylori* genomes from Narino, Colombia

In the MLST analysis we observed independent clades of hspWAfrica, hpAfrica2, hpEAsia, hpEurope, hspAmerindian and independent lineages of Latin American strains (hspColombia and hspNicaragua). In the clade of Native American we observed a close evolutive relation with Asian strains. In the hpEurope lineage, we found isolates from Asia continent that has been reported as ancestors in these populations [15,40]. Likewise, a group of bacteria from Colombia, Mexico and Nicaragua located in the Amerindian and European clades were also mixed with European isolates.

The strains from both study sites (SV449_1 and PZ5009_3A2) showed in the phylogenetic tree association with independent Colombian strains (hspColombia). Although, we observed four strains from the pacific coast (PZ5019_3A3, PZ5016_3A3, PZ5006_3A3, PZ5005_3A3) that along with others from Mexico and Nicaragua were associated to West Africa (hspWAfrica), and interestingly the phylogenetic tree we found a Colombian isolated group composed by an isolate from Pacific coast (PZ5004) and an isolated from Andean zone (SV340_2) that showed association with the most ancestral lineage of *H. pylori* (hpAFrica2) from the South African Continent. A group of four isolates from Tuquerres (SV376_1, SV397_2, SV328_2 and SV355_2) and two from Tumaco (PZ5026, PZ5033_3A2) clustered with Latin American ancestors (Figure 2).

The tree was created by mean of Neighbour-joining. The isolates from Tumaco, Colombia (low risk of gastric cancer) were marked with green point, and those from Tuquerres, Colombia (high risk of gastric cancer) were marked with yellow point. The IDs corresponding to each *H. pylori* isolate were discriminated by color, and they can be seen in the outward perimeter of the tree. The different *H. pylori* populations (hp) and subpopulations (hsp) are described on the right.

Phylogenomic analysis of *Helicobacter pylori* confirming the positioning of the isolates from Narino in the groups of Colombian (hspColombia) and hspWAfrica isolates

In the tree obtained from the core phylogenomic analysis we also observed the formation of independent clades of hpEurope, hspAmerindian, hpEAsia, hspWAfrica, hspColombia, hspNicaragua and hspMexico, with hpAfrica2 being the farthest group. These results were consistent with those from the MLST analysis, which showed that hspAmerindian is among the group of bacteria from hpEAsia. From Tuquerres site 8/10 (80%) *H. pylori* isolates (SV449_1, SV355_2, PZ5080, PZ5086, SV328_2, SV376_1, SV380_1 and SV397_2) formed an independent lineage with Colombian isolates (hspColombia) and only 2/10 (20%) isolates (SV380_1, SV397_2) grouped with African ancestors (hspWAfrica). Regarding the *H. pylori* isolates from Tumaco 5/9 (56%) (PZ5004, PZ5005_3A3, PZ5024, PZ5016_3A3 and PZ5006_3A3) clustered with African ancestors (hspWestAfrica). However, 4/9 (44%) isolates (PZ5019_3A3, PZ5033_3A2, PZ5009_3A2 and PZ5026) grouped with Colombian lineage (hspColombia) (Figure 3).

Those in yellow point are from the Andean zone, green point for the Pacific zone. Different *H. pylori* populations (hp) and subpopulations (hsp) are described on the right.

The phylogenetic tree created using SNPs we observed similarly to previous results formation of independent clades for *H. pylori* from each continent. In addition, we found a group of strains with Colombian and Nicaraguan origin inside the hpEurope. Also, the hpEurope lineage showed a close evolutive relation to Asian strains (hpEAsia) as same as the MLST analysis. In the West Africa lineage we observed isolated from Mexico, Nicaragua and Colombia. The *H. pylori* isolates from Tumaco 5/9 (56%) that were reported in the phylogenomic tree based on SNPs genome grouped with hspWAfrica (PZ5006_3A3, PZ5016_3A3, PZ5004, PZ5024, PZ5005_3A3) and 4/9 (44%) *H. pylori* isolates (PZ5019_3A3, PZ5033_3A2, PZ5009_3A2 y PZ5026) were observed forming an independent lineage (hspColombia) (Figure 4).
Figure 2: Phylogenetic tree of *H. pylori* done by MLST of samples from regions with high and low risk of gastric cancer of the department of Nariño, Colombia.

Note: ( ) hpAfrica2; ( ) hspWAfrica; ( ) hspAmerindian; ( ) hpEAsia; ( ) hpEurope; ( ) hspMexico; ( ) hspColombia; ( ) hspNicaragua.

Figure 3: The core phylogenomic tree of the isolates from the high mountain and the Pacific zone based on comparison with complete genomes.

Note: ( ) hpAfrica2; ( ) hspWAfrica; ( ) hspAmerindian; ( ) hpEAsia; ( ) hpEurope; ( ) hspMexico; ( ) hspColombia; ( ) hspNicaragua.

Figure 4: The two study sites isolates phylogenomic tree based on SNPs.

Note: ( ) hpAfrica2; ( ) hspWAfrica; ( ) hspAmerindian; ( ) hpEAsia; ( ) hpEurope; ( ) hspMexico; ( ) hspColombia; ( ) hspNicaragua.
Those in yellow point are from the Andean zone, green point for the Pacific zone. Different *H. pylori* populations (hp) and subpopulations (hsp) are described on the right.

From Tuquerres population 8/10 (80%) isolates formed in the phylogenomic tree based on SNPs a cluster with Colombian ancestors (hspColombia) and only 2/10 (20%) isolates (SV397_2 y SV380_1) were grouped with African ancestors (hspWAfrica), (Figure 4). We observed that both used methods showed identical proportions of *H. pylori* ancestry for both phylogenomic trees in both studied populations.

**Phylogenetic analysis of vacA and alpA revealing the origin and rapid evolution of Helicobacter pylori isolates from Narino, Colombia**

The analyzed nucleotide sequence of the vacA gene showed results where the six isolates from the Andean zone (SV326_2, SV449_1, SV340_2, PZ5080, SV355_2 and PZ5056) and five strains from Pacific coast (PZ5026, PZ5016_3A3, PZ5033_3A2, PZ5019_3A2, PZ5009_3A2) are grouped preferentially with hspColombia. Andean strain SV380_1 indicated a strong relation with Mexican strains. Pacific coast strains PZ5005_3A3 and PZ5006_3A3 were grouped with African strains hspWAfrica. Nevertheless, Amerindian strain PeCan4 was observed with the Mexican strains (Figure 5).

The isolates from Tumaco, Colombia (low risk of gastric cancer) were marked with green point, those from Tuquerres, Colombia (high risk of gastric cancer). Different *H. pylori* populations (hp) and subpopulations (hsp) are described on the right.

In the alpA gene phylogenetic tree we observed independent lineage described previously, it is clear that a group of Colombian and Mexican isolates are very close evolutive related to hspAfrica2 strains. We observed that a group of isolates from both study sites Tumaco (Three strains: PZ5005_3A3, PZ5016_3A3 and PZ5009_3A2) and Tuquerres (Three strains: SV340_2, SV328_2 and SV380_1) were differentiated in the hspColombia clade, although there were strains from the same regions among the hispanoamerican and European strains. Curiously, in this phylogenetic tree we observed the strains SV355_2 from Tumaco and PZ5019_3A3 and PZ5006 from Tumaco in the group of isolates provenient from Native Americans (hspAmerindian). Only one isolated from Tumaco (PZ5006_3A3) grouped with West African ancestors (hspWAfrica) (Figure 6).

The tree was created by means of Maximum-likelihood. The isolates from Tumaco, Colombia (low risk of gastric cancer) were marked in green point, those from Tuquerres, Colombia (high risk of gastric cancer) were marked in yellow point. Different *H. pylori* populations (hp) and subpopulations (hsp) are described on the right.
Discussion

Since the first human migration from East Africa approximately 60,000 years ago [5], *H. pylori* shows in its genome different migration routes and its human host geographic settlements. The first humans in America came from Asia crossing the Bering Strait 12,000 years ago, by the time and the geographic isolation generated a new *H. pylori* genetic heritage that infected the Native Americans population (hspAmerindian) [15,18,41-44]. However, the recent colonization by Europeans and Africans no more than 500 years ago involved the mixing of a new microorganism’s burden as *H. pylori*.

The MLST evolutive and phylogenomic analysis of *H. pylori* showed that the inhabitants from Andean zone (20%) and Pacific coast (56%) area had some *H. pylori* strains from West Africa (hspWAfrica) and South Africa continent (hpAfrica2). There is genetic evidence that from those areas came to America African slaves [45]. Curiously, in several isolates from both populations Tuquerres: (80%); Tumaco: (44%) we observed the development of a new lineage to Colombian strains (hspColombia). In contrast to our study, there is other study using MLST that reported *H. pylori* isolates from the same study sites were associated with isolates from European origin (hpEurope) [19].

The results of virulence gene *vacA* (isolates PZ5005_3A3 and PZ5006_3A3) and adhesine *alpA* (isolate PZ5006_3A3) from Tumaco indicated ancestral homology with African strains. Interestingly, the *alpA* gene showed that the strains SV355_2 and PZ5086 from Tuquerres and PZ5019_3A3 from Tumaco were associated with Amerindian origin strains (hspAmerindian). By the other side, we observed that most of the isolates from both study sites were clustered with small groups of hspColombia.

In Narino department, using MLST it is been identified that Native American strains hspAmerind from populations from Andean area were moved by European strains hpEurope in mestizo coming from Amerindian and Spanish ancestors and hspWAfrica in population from Africa, Europe and Amerindian ancestors [14,19,20]. However, the new findings about *alpA* gene from isolates with ancestral homology allow to understand that there is no had movement or substitution of native strains.

All this suggest multiple colonization in these populations by strains with different phylogenography origin as hpAfrica2, hpWAfrica, hpEurope and hspAmerindian. The multiple origins of those genes could be used as prerequisite to be able of fixation and colonization on gastric mucosa, in response to genetic, immune and physiological host characteristics [21]. In consequence, these genes will be present due to the variability of competitive *H. pylori* strains that inside a host cooperated through quorum sensing [46,47].

In the Andean region (Tuquerres) the presence of hspColombia cluster could be product of the adaptation to the new ecological niche from mestizo populations. However, the infection by a new subpopulations implies a new encounter between the host and bacteria, in case of a relation no synchronized that favor the genotypes and phenotypes co-evolved, involve the hosts and pathogen extinction [48]. This includes a new set of interactions between virulence factors that provide immune mechanisms evasion systems and colonization by *H. pylori*. The bacterium performs this two process by two genes, *vacA* that synthesize a protein that inhibits the host immune response and increase the signaling pathways of gastric mucous cells allowing the insertion of *CagA* proteins that improve the *H. pylori* growing [47,49].

The recent gene adaptation probably brought consequences on pathogenesis in Tuquerres, for example, the *vacA* alleles s1m1...
that are associated to major virulence are more frequent in this population and the cagA gene expression is higher in the Andean region than Pacific coast [50,51]. Possibly causing a high \( H. pylori \) proliferation and an ecological unbalanced of bacteria communities allowing the pre-cancerous lesion development in human population.

The incidence of gastric cancer rates changes around the world, for example, in several African countries the cancer gastric incidence is ubicuos [52-55]. This could be due to that during the hunter-gatherers time in the African savannah, the process of human evolution was like a bottle effect to very pathogenic microorganisms, selecting and transmitting to the next generations only the commensal bacteria. This could extinguish the very pathogenic bacteria with its host, like \( H. pylori \) [56]. The arriving of these African populations to Colombian pacific coast brought \( H. pylori \) commensal strains that have co-evolved with human host during thousands of years leading to less probability of disease development [14].

It is important to mention that the Th immune response plays an important role in gastric cancer pathogenesis [12]. The response is multifactorial against a specific strain [14]. This has been observed in the African continent where the Th2 response is predominant, while in Japan (high gastric cancer incidence) the immune response is Th1 type. These differences have been related with a minor gastric cancer incidence on Africa [56].

In both study regions the coinfection with other microorganisms like helminths has modulated the immune response [56]. In Tumaco population where the human ancestry is the African origin mainly, it is observed an immune response pattern Th2, where the co-infection with helminths is a factor that has modulated this mechanism [57]. The Th2 response is an anti-inflammatory type as immune mechanism to infection by \( H. pylori \) strains of African ancestry, product of a co-evolution period since the \( H. sapiens \) sapiens evolved from a common African branch in parallel to \( H. pylori \) since more than 110,000 years [14,19,20].

However, in Tuquerres population the infection by \( H. pylori \) of hspColombia lineage that causes an immune response Th1 pro-inflammatory [14,19,57], which interaction with human host has happened since about 500 years, short co-evolution and adaptation time after European and African colonization, that suggest evidence of high disease risk in Andean area from Narino department.

In addition, other studies have reported that the dietary differences between the mountain and coast regions influenced the incidence rates, for example, in the coast region there are high intake of fresh vegetables and fruits rich in antioxidants and seafood. While in Andean regions the dietary regimen consists on potatoes and broad beans [12,55,58].

We acknowledge that our study has a low isolates number that limits the correlation statistically the strain origin with gastric lesion type. Also, our isolates number limits us to corroborate with whole genome analysis if the strains with African origin in human population with high Amerindian ancestry produce a higher risk of carcinogenesis as found by MLST in \( H. pylori \) and human ancestry as in a previous study [14]. Also, the low number of \( H. pylori \) isolates in both study areas limits our study related with the mutation or variants analysis that give major capability of colonization and virulence of \( H. pylori \).

**Conclusion**

The perspectives to overcome our study limitations are related with the enrollment of new population in Narino with high gastric cancer risk. However, this study gives a preliminary knowledge about \( H. pylori \) evolution with mixed human host from Colombia. Also, this study could be a good example to understand the mixing process in other bacteria after the European and African colonization in America territory.

**Acknowledgements**

We would like to thank the reviewers for the constructive comments and guidelines, and also the Universidad de Narino for the availability of the server for the evolutionary analyses.

**Funding Source**

This work was funded by the Grupo de Salud Publica and the Vicerrectoría de Investigaciones e Interacción Social from the Universidad de Narino.

**Competing Interests**

The authors declare that they have no competing interests.

**References**


