The *Helicobacter pylori* genome evolution in different gastric cancer risk Colombian populations

Kevin Guzman (keving346@gmail.com)
Universidad de Nariño https://orcid.org/0000-0002-6483-4894
Lidia Montenegro
Universidad de Nariño
Alvaro Pazos
Universidad de Nariño https://orcid.org/0000-0001-5603-7898

Research Article

Keywords: Evolution, Colombian populations, gastric cancer, H. pylori, phylogenomics

Posted Date: October 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-960798/v1

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.

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 linages 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 Nariño, 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**) (29,30).

Table 1. Complete genome description of *H. pylori* strains in NCBI from two regions of the department of Nariño, Colombia					
Risk	**Isolate ID**	**Access code**	**Diagnosis**	**Genome size**	**N° of contigs**
HR[a]	SV328_2	MTWO00000000	Dys[c]	1,645,479	56
HR	SV340_2	MTWP00000000	NAG[d]	1,633,298	53
HR	SV355_2	MTWQ00000000	IM[e]	1,635,304	39
HR	SV376_1	MTWR00000000	IM	1,691,791	207
HR	SV380_1	MTWU00000000	IM	1,631,819	40
HR	SV397_2	MTWS00000000	NAG	1,668,205	47
HR	SV449_1	MTWT00000000	IM	1,654,884	41
HR	PZ5056*	ASYU00000000	NAG	1,578,164	335
HR	PZ5080*	ASYV00000000	IM	1,597,127	283
HR	PZ5086*	ASYW00000000	NAG	1,547,845	295
LR[b]	PZ5005_3A3	MTWJ00000000	NAG	1,672,956	51
LR	PZ5006_3A3	MTWK00000000	NAG	1,643,170	53
LR	PZ5009_3A2	MSYO00000000	NAG	1,677,035	53
LR	PZ5016_3A3	MTWL00000000	MAG[f]	1,644,424	44
LR	PZ5019_3A3	MTWM00000000	IM	1,681,561	44
LR	PZ5033_3A2	MTWN00000000	IM	1,656,908	60
LR	PZ5004*	ASZF00000000	NAG	1,569,902	303
LR	PZ5024*	ASYS00000000	NAG	1,496,849	413
LR	PZ5026*	ASYT00000000	NAG	1,604,992	253

[a]High Risk: Host is resident of Tuquerres where risk for gastric cancer is high.

[b]Low Risk: Host is resident of Tumaco where risk for gastric cancer is low.

[c]Dys, dysplasia

[d]NAG, nonatrophic gastritis

[e]IM, intestinal metaplasia

[f]Multifocal atrophic gastritis

*Data samples from previous work by Sheh et al., 2013 (29)

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 10^3 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, CA22311, CA22339, CA22312, CM22360, CM22351), 10 from Cundinamarca department (CC22402, CC26084, CC26093, CM22346, CA22337, CM22341, CG22389, CG22322, CA22393, CM22388), 14 from Boyaca department (CG22025, CG22087, CG22023, CM22046, 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, urel*, and *yphC* were annotated using PubMLST (https://pubmlst.org/helicobacter/), 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 (34). 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 (35).

The phylogenomic analysis based on SNPs was carried out using CSI-phylogeny (36) 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 (37). 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 (33).

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, *ln*L= -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, *ln*L=-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 (Fig. 1-A). 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 (Fig. 1-B), which corroborates the results obtained in a previous work (14).

Multilocus Sequence Typing (MLST) analysis based on Helicobacter pylori genomes from Nariño, Colombia

In the MLST analysis we observed independent clades of hspWAfrica, hpAfrica2, hpEAsia, hpEurope, hspAmerindian and independent linages 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 linage, we found isolates from Asia continent that has been reported as ancestors in this 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 linage 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 (Fig. 2).

Phylogenomic analysis of Helicobacter pylori confirming the positioning of the isolates from Nariño 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, SV397_2) formed an independent linage 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 linage (hspColombia) (Fig. 3).

The phylogenomic 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 linage showed a close

evolutive relation to Asian strains (hpEAsia) as same as the MLST analysis. In the West Africa linage 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 linage (hspColombia) **(**Fig. 4**)**.

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), **(**Fig. 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 Nariño, 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 (Fig. 5).

In the *alpA* gene phylogenetic tree we observed independent linage described previously, it is clear that a group of Colombian and Mexican isolates are very close evolutive related to hpAfrica2 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 Tuquerres and PZ5019_3A3 and PZ5086 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) (Fig. 6).

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, 43, 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 habitants 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 linage 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 Nariño 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 a multiple colonization in these populations by strains with different phylogeography origin as hpAfrica2, hspWAfrica, 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 include 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 synthetize 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 low (0,6 per 100,000 inhabitants), even although the *H. pylori* infection is ubicuos (52–55). This could be due to that during the hunter-gatherers time in the African savannah, the process of human evolution were like a bottle effect to very pathogenic microorganisms, selecting and transmitting to the next generations only the commensal bacteria. This could extinguished the very pathogenic bacteria with its host, like *H. pylori* (56). The arriving of this African populations to Colombian pacific coast brought *H. pylori* commensal strains that has 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. This 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 patron 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* 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 linage 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 Nariño 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*. The perspectives to overcome our study limitations are related with the enrollment of new population in Nariño 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.

Declarations

Acknowledgements

We would like to thank the reviewers for the constructive comments and guidelines, and also the Universidad de Nariño 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 Nariño.

Competing Interests

The authors declare that they have no competing interests

References

1. Peek RM, Blaser MJ. *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. Nat Rev Cancer. 2002;2:28–37. https://doi.org/10.1038/nrc703
2. Cover TL, Blaser MJ. *Helicobacter pylori* in health and disease. Gastroenterology. 2009;136:1863–1873. https://doi.org/10.1053/j.gastro.2009.01.073
3. **International Agency for Research on Cancer.** Schistosomes, liver flukes and *Helicobacter pylori*. IARC working group on the evaluation of carcinogenic Risks to Humans. Lyon, 7–14 June 1994. IARC Monogr. Eval. Carcinog. Risks Hum. 1994;61:1–241
4. Moodley Y, Linz B, Bond RP, Nieuwoudt M, Soodyall H, Schlebusch CM, **et al**. Age of the association between *Helicobacter pylori* and man. PLoS Pathogens. 2012;8:e1002693. https://doi.org/10.1371/journal.ppat.1002693
5. Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P, **et al**. An African origin for the intimate association between humans and *Helicobacter pylori*. Nature. 2007;445:915–918. https://doi.org/10.1038/nature05562
6. Atherton JC, Blaser MJ. Coadaptation of *Helicobacter pylori* and humans: ancient history, modern implications. The Journal of clinical investigation. 2009;119:2475–2487. https://doi.org/10.1172/JCI38605
7. Jones KR, Whitmire JM, Merrell DS. A tale of two toxins: *Helicobacter pylori* CagA and VacA modulate host pathways that impact disease. Front Microbiol. 2010;1:115. https://doi.org/10.3389/fmicb.2010.00115

8. Gangwer KA, Shaffer CL, Suerbaum S, Lacy DB, Cover TL, Bordenstein SR. Molecular evolution of the *Helicobacter pylori* vacuolating toxin gene *vacA*. Journal of bacteriology. 2010;192:6126–6135. https://doi.org/10.1128/JB.01081-10
9. Odenbreit S, Till M, Hofreuter D, Faller G, Haas R. Genetic and functional characterization of the *alpAB* gene locus essential for the adhesion of *Helicobacter pylori* to human gastric tissue. Molecular Microbiology. 1999;31:1537–1548. https://doi.org/10.1046/j.1365-2958.1999.01300.x
10. Lu H, Wu JY, Beswick EJ, Ohno T, Odenbreit S, Haas R, **et al**. Functional and intracellular signaling differences associated with the *Helicobacter pylori* AlpAB adhesin from western and East Asian strains. Journal of Biological Chemistry. 2007;282:6242–6254. https://doi.org/10.1074/jbc.M611178200
11. Torres J, Correa P, Ferreccio C, Hernandez-Suarez G, Herrero R, Cavazza-Porro M, **et al**. Gastric cancer incidence and mortality is associated with altitude in the mountainous regions of Pacific Latin America. Cancer Causes Control. 2013;24:249–256. https://doi.org/10.1007/s10552-012
12. Correa P, Piazuelo B. Gastric cancer: the colombian enigma. Revista Colombiana de Gastroenterología. 2010;25:334–337
13. Matta AJ, Pazos AJ, Bustamante-Rengifo JA, Bravo LE. Genomic variability of *Helicobacter pylori* isolates of gastric regions from two Colombian populations. World journal of gastroenterology. 2017;23:800–809. https://doi.org/10.3748/wjg.v23.i5.800
14. Kodaman N, Pazos A, Schneider BG, Piazuelo MB, Mera R, Sobota RS, **et al**. Human and *Helicobacter pylori* coevolution shapes the risk of gastric disease. Proc Natl Acad Sci USA. 2014;111:1455–1460. https://doi.org/10.1073/pnas.1318093111
15. Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, **et al**. Traces of human migrations in *Helicobacter pylori* populations. Science. 2003;299:1582–1585. https://doi.org/10.1126/science.1080857
16. Devi SM, Ahmed I, Francalacci P, Hussain MA, Akhter Y, Alvi A, **et al**. Ancestral European roots of *Helicobacter pylori* in India. BMC Genom. 2007;8:184. https://doi.org/10.1186/1471-2164-8-184
17. Moodley Y, Linz B, Yamaoka Y, Windsor HM, Breurec S, Wu JY, **et al**. The peopling of the Pacific from a bacterial perspective. Science. 2009;323:527–530. https://doi.org/10.1126/science.1166083
18. Yamaoka Y. *Helicobacter pylori* typing as a tool for tracking human migration. Clin Microbiol Infect. 2009;15:829–834. https://doi.org/10.1111/j.1469-0691.2009.02967.x
19. **de** Sablet T, Piazuelo MB, Shaffer CL, Schneider BG, Asim M, Chaturvedi R, **et al**. Phylogeographic origin of *Helicobacter pylori* is a determinant of gastric cancer risk. Gut. 2011;60:1189–1195. https://doi.org/10.1136/gut.2010.234468
20. Shiota S, Suzuki R, Matsuo Y, Miftahussurur M, Huyen T, Thanh T, **et al**. *Helicobacter pylori* from gastric cancer and duodenal ulcer show same phylogeographic origin in the Andean region in Colombia. PLoS ONE. 2014;9:e105392. https://doi.org/10.1371/journal.pone.0105392
21. Gutiérrez-Escobar AJ, Trujillo E, Orlando E. Acevedo S, Bravo MM. Phylogenomics of Colombian *Helicobacter pylori* isolates. Gut Pathogens. 2017;9:52–61. https://doi.org/10.1186/s13099-017-

0201-1

22. Muñoz-Ramírez ZY, Mendez-Tenorio A, Kato I, Bravo MM, Rizzato C, Thorell K, **et al**. Whole genome sequence and phylogenetic analysis show *Helicobacter pylori* strains from Latin America have followed a unique evolution pathway. Front Cell Infect Microbiol. 2017;7:50. https://doi.org/10.3389/fcimb.2017.00050
23. Thorell K, Yahara K, Berthenet E, Lawson DJ, Mikhail J, Kato I, **et al**. Rapid evolution of distinct *Helicobacter pylori* subpopulations in the Americas. PLoS Genet. 2017;13:e1006546. https://doi.org/10.1371/journal.pgen.1006546
24. Correa P, Piazuelo MB. Evolutionary History of the *Helicobacter pylori* Genome: Implications for Gastric Carcinogenesis. Gut and liver. 2012;6:21–28. https://doi.org/10.5009/gnl.2012.6.1.21
25. Cortes A, Brown MA. Promise and pitfalls of the Immunochip. Arthritis Res Ther. 2011;13:101. https://doi.org/10.1186/ar3204
26. **International HapMap Consortium**. A haplotype map of the human genome. Nature. 2005;437:1299–1320. https://doi.org/10.1038/nature04226
27. Abecasis GR, **et al**. 1000 Genomes Project Consortium An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491:56–65. https://doi.org/10.1038/nature11632
28. Jakobsson M, Rosenberg NA. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007;23:1801–1806. https://doi.org/10.1093/bioinformatics/btm233
29. Pazos A, Kodaman N, Piazuelo MB, Romero-Gallo J, Sobota RS, Israel DA, Bravo LE, Morgan DR, Wilson KT, Correa P, Peek RM **Jr**, Williams SM, Schneider BG. Draft Genome Sequences of 13 Colombian Helicobacter pylori Strains Isolated from Pacific Coast and Andean Residents. Genome Announc. 2017;13;5(15):e00113-17. doi: https://doi.org/10.1128/genomeA.00113-17
30. Sheh A, Piazuelo MB, Wilson KT, Correa P, Fox JG. Draft genome sequences of *Helicobacter pylori* strains isolated from regions of low and high gastric cancer risk in Colombia. Genome announcements. 2013;1:e00736-13. https://doi.org/10.1128/genomeA.00736-13
31. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–1797. https://doi.org/10.1093/nar/gkh340
32. Saitou N, Nei M. The neighbour-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 1987;4:406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454
33. Rambaut, A. FigTree. Tree figure drawing tool. 2009. http://tree.bio.ed.ac.uk/software/figtree/
34. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25(11):1451–1452 https://doi.org/10.1093/bioinformatics/btp187
35. Hudson RR, Kaplan NL. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics. 1985;111:147–164

36. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010;11: 595. https://doi.org/10.1186/1471-2105-11-595
37. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger data sets. Molecular Biology and Evolution. 2016;33:1870–1874. https://doi.org/10.1093/molbev/msw054
38. Kaas RS, Leekitcharoenphon P, Aarestrup FM, Lund O. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. PLoS One. 2014;9:e104984. https://doi.org/10.1371/journal.pone.0104984
39. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. Oxford University Press. 2011;27:2987–2993. https://doi.org/10.1093/bioinformatics/btr509
40. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 2007;56:564–577. https://doi.org/10.1080/10635150701472164
41. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology. 2010;59(3):307–321. https://doi.org/10.1093/sysbio/syq010
42. Maixner F, Krause-Kyora B, Turaev D, Herbig A, Hoopmann MR, Hallows JL, **et al**. The 5300-year-old *Helicobacter pylori* genome of the Iceman. Science. 2016;351:162–165. https://doi.org/10.1126/science.aad2545
43. Kersulyte D, Kalia A, Gilman RH, Mendez M, Herrera P, Cabrera L, **et al**. *Helicobacter pylori* from Peruvian Amerindians: traces of human migrations in strains from remote Amazon, and genome sequence of an Amerind strain. PLoS ONE. 2010;5:e15076. https://doi.org/10.1371/journal.pone.0015076
44. Kersulyte D, Bertoli MT, Tamma S, Keelan M, Munday R, Geary J, **et al**. Complete genome sequences of two *Helicobacter pylori* strains from a Canadian arctic aboriginal community. Genome Announc. 2015;3:e00209-15. https://doi.org/10.1128/genomeA.00209-15
45. Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, **et al**. The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet. 2004;74:454–465. https://doi.org/10.1086/382194
46. Lu H, Hsu PI, Graham DY, Yamaoka Y. Duodenal ulcer promoting gene of *Helicobacter pylori*. Gastroenterology. 2005;128:833–848. https://doi.org/10.1053/j.gastro.2005.01.009
47. Argent RH, Hale JL, El-Omar EM, Atherton JC. Differences in *Helicobacter pylori CagA* tyrosine phosphorylation motif patterns between western and East Asian strains, and influences on interleukin-8 secretion. J. Med. Microbiol. 2008;57:1062–1067. https://doi.org/10.1099/jmm.0.2008/001818-0
48. Linz B, Windsor HM, Gajewski JP, Hake CM, Drautz DI, Schuster SC, **et al**. *Helicobacter pylori* genomic microevolution during naturally occurring transmission between adults. PLoS ONE. 2013;8:e82187.

https://doi.org/10.1371/journal.pone.0082187

49. Aras RA, Kan J, Tschumi A, Harasaki Y, Blaser MJ. Extensive repetitive DNA facilitates prokaryotic genome plasticity. Proc. Natl Acad. Sci. USA. 2003;100:13579–13584. https://doi.org/10.1073/pnas.1735481100
50. Bravo LE, **van** Doom LJ, Realpe JL, Correa P. Virulence-associated genotypes of *Helicobacter pylori*: Do they explain the African enigma?. Am J Gastroenterol. 2002;97:2839–2842. https://doi.org/10.1111/j.1572-0241.2002.07031.x
51. Loh JT, Shaffer CL, Piazuelo MB, Bravo LE, McClain MS, Correa P, **et al**. Analysis of *cagA* in *Helicobacter pylori* strains from Colombian populations with contrasting gastric cancer risk reveals a biomarker for disease severity. Cancer Epidemiology and Prevention Biomarkers. 2011;2237-2249. https://doi.org/10.1158/1055-9965.EPI-11-0548
52. Mitchell HM, Ally R, Wadee A, Wiseman M, Segal I. Major differences in the IgG subclass response to *Helicobacter pylori* in the first and third worlds. Scand. J. Gastroenterol. 2002;37:517–522. https://doi.org/10.1080/00365520252903044
53. Holcombe C. *Helicobacter pylori*: The African enigma. Gut. 1992;33:429–431.
54. Campbell D, Warren B, Thomas J, Figura N, Telford J, Sullivan P. The African enigma: Low prevalence of gastric atrophy, high prevalence of chronic inflammation in West African adults and children. Helicobacter. 2001;6:263–267. https://doi.org/10.1046/j.1083-4389.2001.00047.x
55. Piazuelo MB, Correa P. Gastric cáncer: Overview. Colomb Med. 2013;44(3),192–201 https://doi.org/10.25100/cm.v44i3.1263
56. Blaser MJ, Kirschner D. "The equilibria that permit bacterial persistence in human hosts". Nature. 2007:449;843-49. https://doi.org/10.1038/nature06198
57. Piazuelo MB, Camargo MC, Mera RM, Delgado AG, Peek RM **Jr**, Correa H, **et al**. Eosinophils and mast cells in chronic gastritis: possible implications in carcinogenesis. Hum Pathol. 2008;39:1360–1369. https://doi.org/10.1016/j.humpath.2008.01.012
58. Camargo MC, Burk RF, Bravo LE, Piazuelo MB, Hill KE, Fontham ET, **et al**. Plasma selenium measurements in subjects from areas with contrasting gastric cancer risks in Colombia. Arch Med Res. 2008;39:443–51. https://doi.org/10.1016/j.arcmed.2007.12.004

Figures

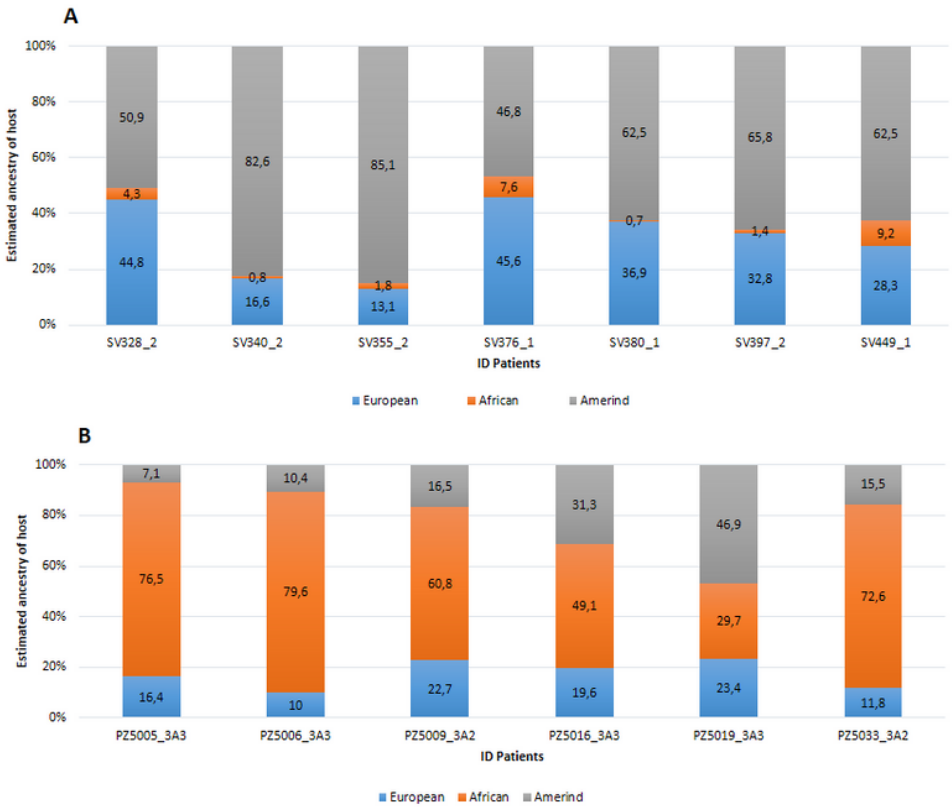

Figure 1

Estimated human ancestry of two populations with high and low risk of gastric cancer. (A) Human ancestry of the Túquerres population (High risk of gastric cancer). (B) Human ancestry of the Túmaco population (Low risk of gastric cancer). The Deep criteria were stablished to define the ancestry as previously reported (14).

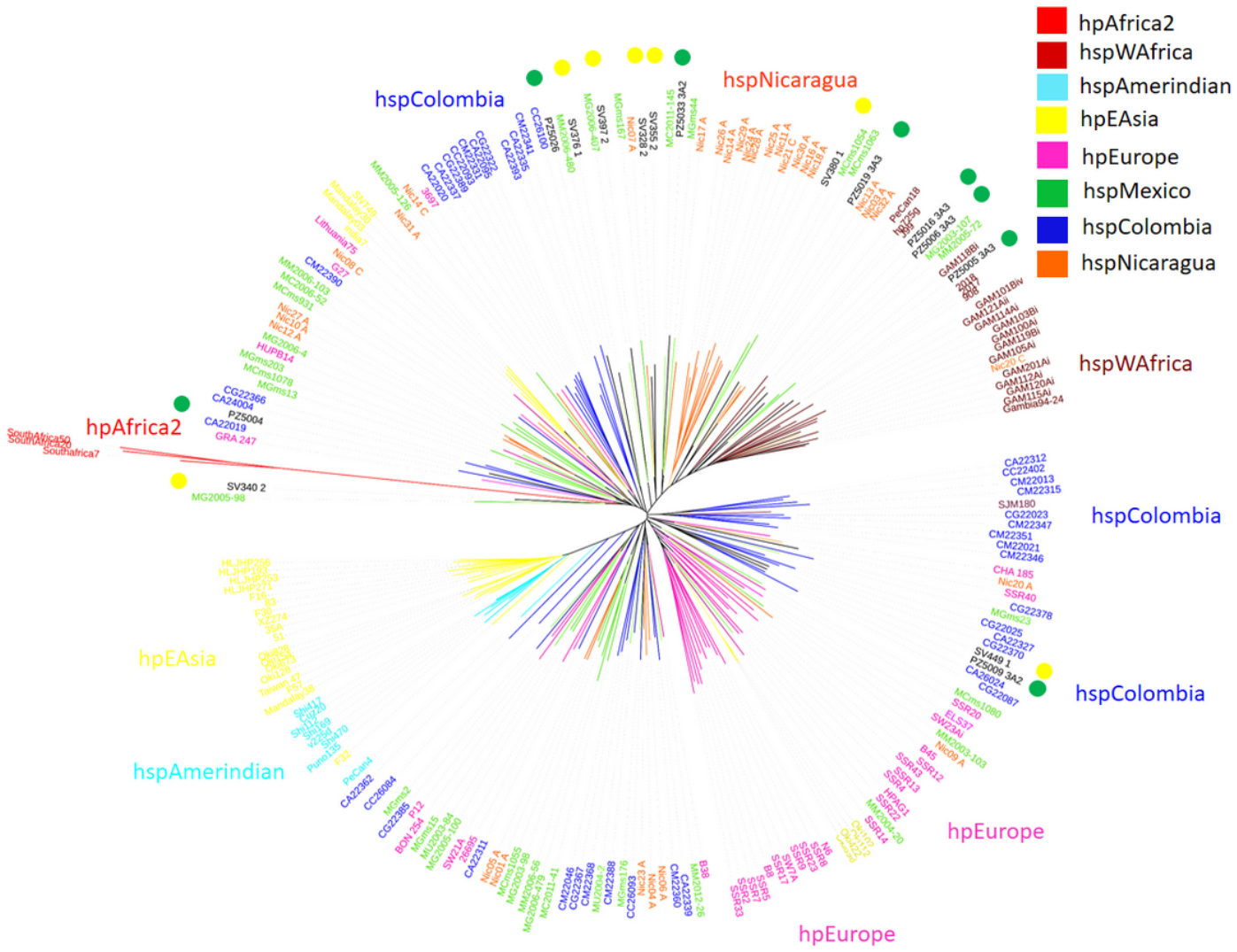

Figure 2

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. The tree was created by mean of Neighbour-joining. The isolates from Túmaco, Colombia (low risk of gastric cancer) were marked with green point, and those from Túquerres, 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.

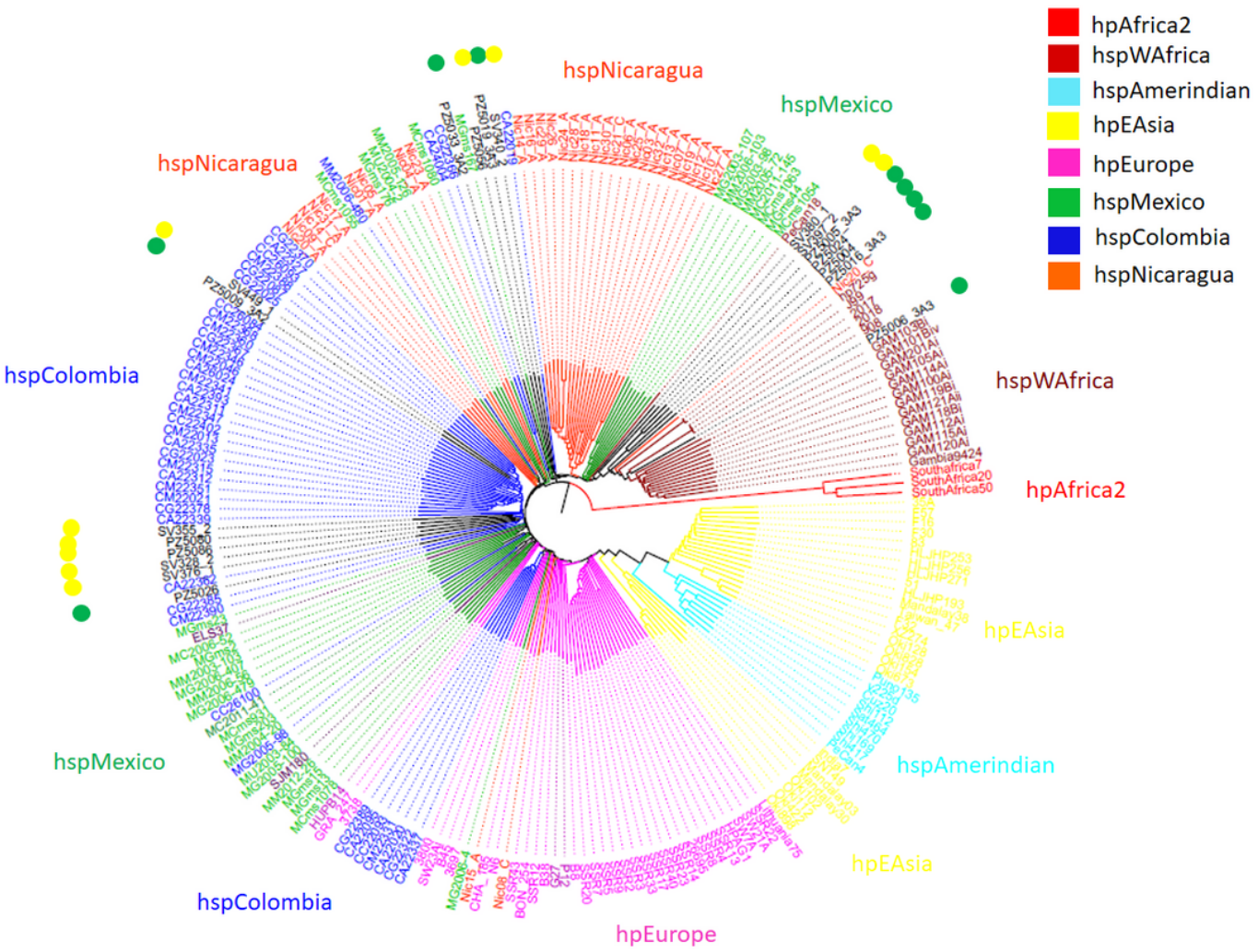

Figure 3

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

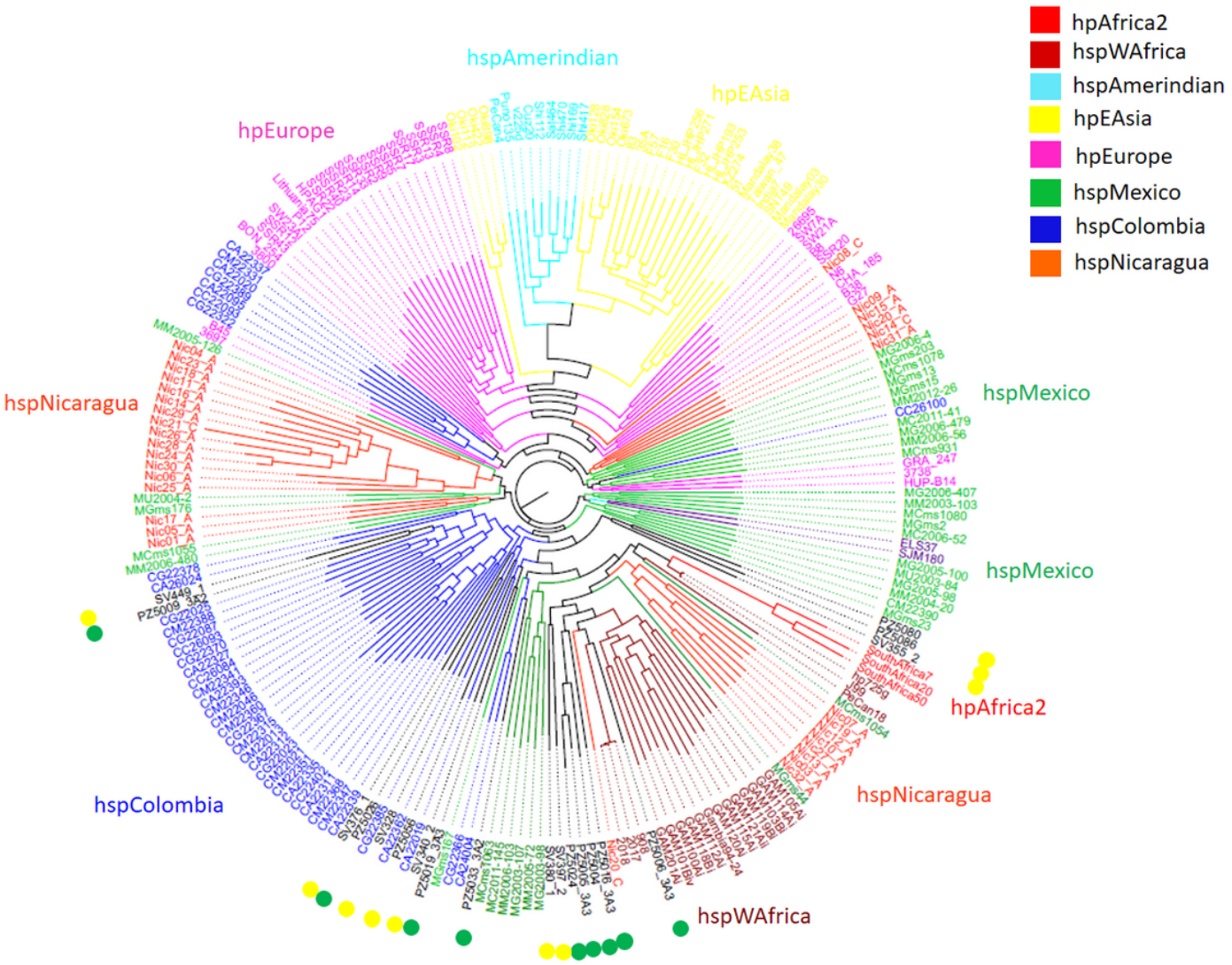

Figure 4

The two study sites isolates phylogenomic tree based on SNPs. 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.

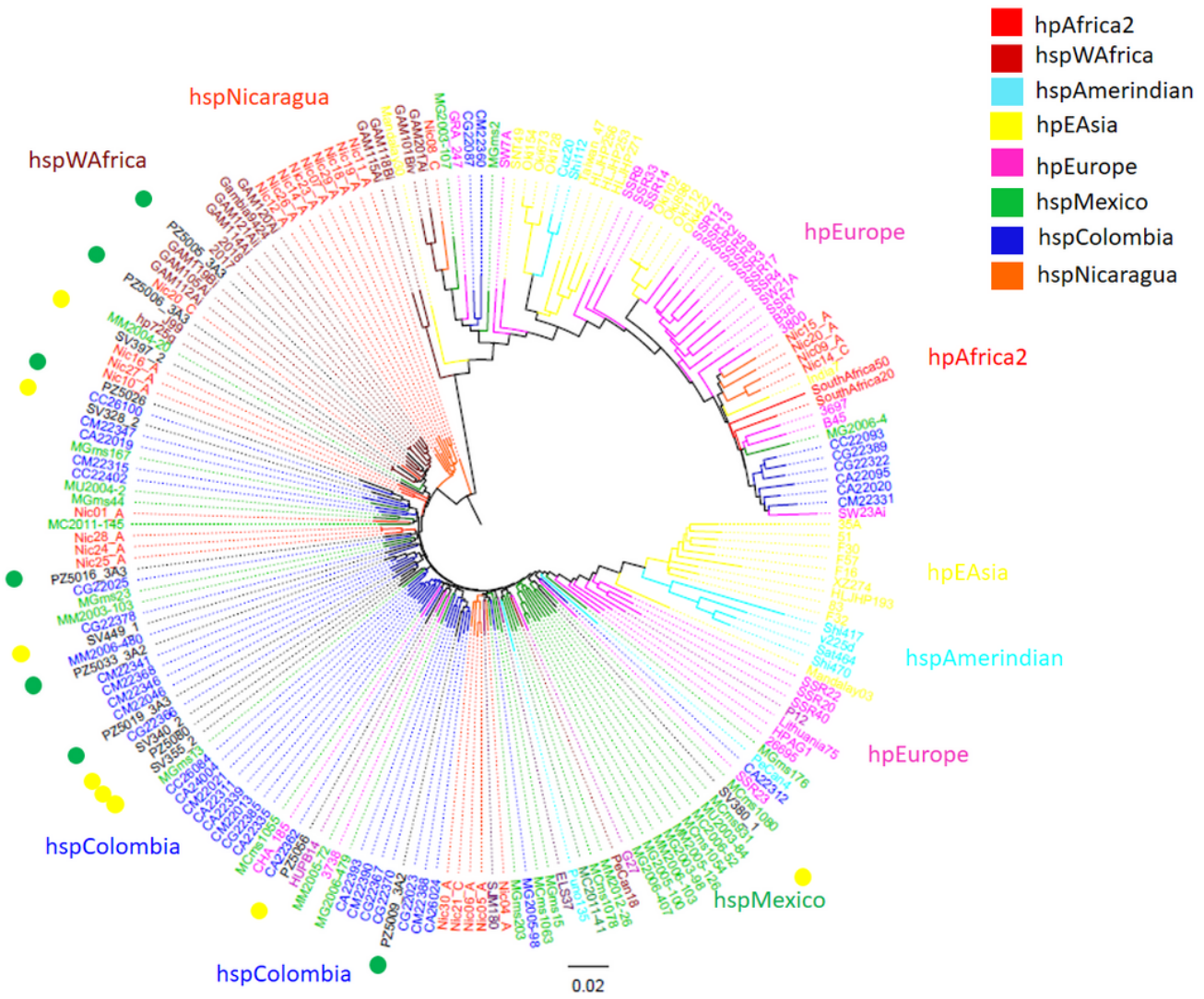

Figure 5

Phylogenetic analysis of DNA sequences of the vacA gene with Maximum-likelihood. 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.

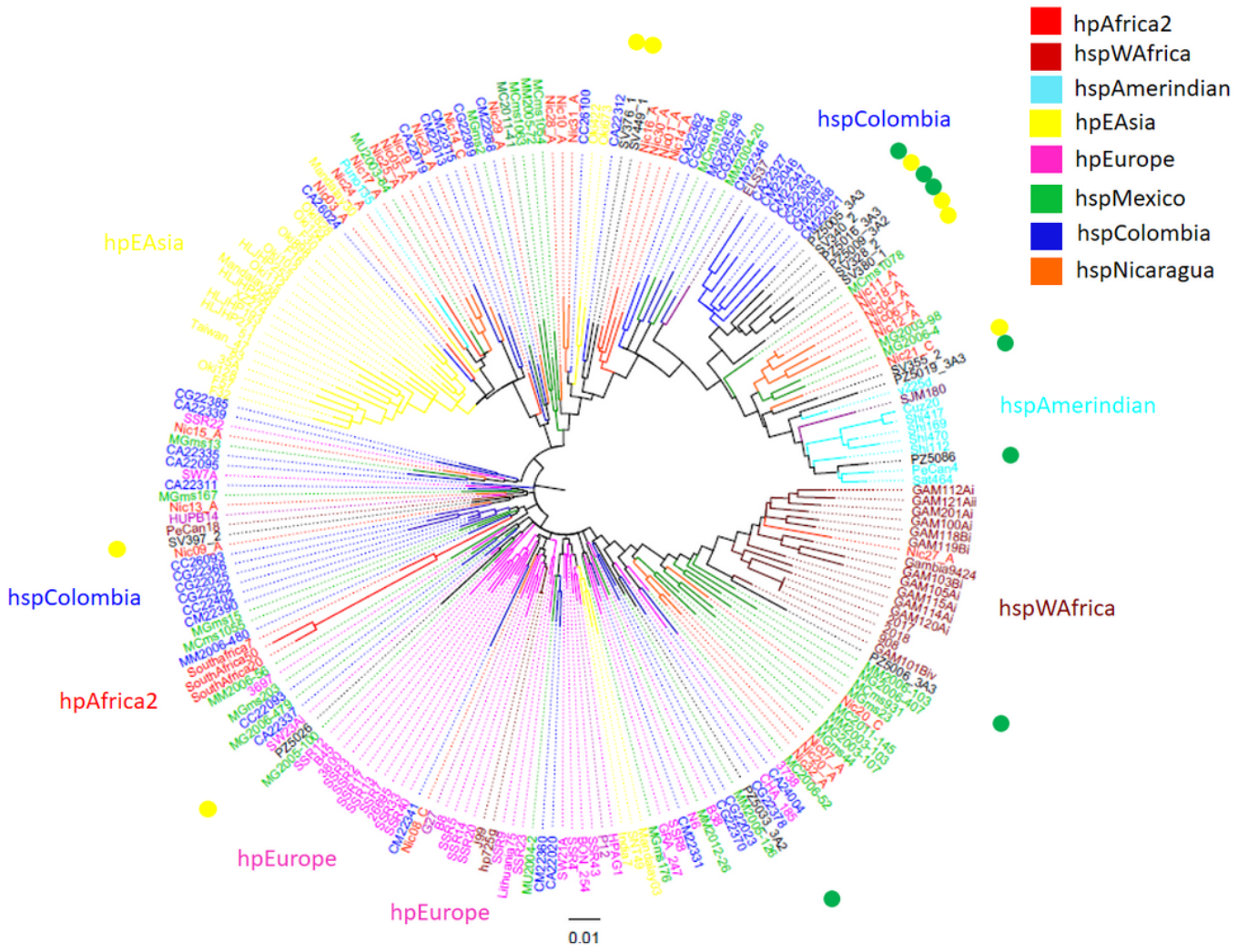

Figure 6

Phylogenetic analysis of the strains from the department of Nariño using sequences of the alpA gene. 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.