








RESEARCH

Open Access



Genomic insights into the antimicrobial resistance and virulence of *Helicobacter pylori* isolates from gastritis patients in Pereira, Colombia

Adalucy Alvarez-Aldana^{1,2†}, Odion O. Ikhimiukor^{3†}, Yina Marcela Guaca-González^{1,2},
Manuela Montoya-Giraldo³, Stephanie S. R. Souza³, Ana Beatriz Garcez Buiatte^{3,4} and
Cheryl P. Andam^{3*}

Abstract

Background *Helicobacter pylori* infects the stomach and/or small intestines in more than half of the human population. Infection with *H. pylori* is the most common cause of chronic gastritis, which can lead to more severe gastroduodenal pathologies such as peptic ulcer, mucosa-associated lymphoid tissue lymphoma, and gastric cancer. *H. pylori* infection is particularly concerning in Colombia in South America, where > 80% of the population is estimated to be infected with *H. pylori* and the rate of stomach cancer is one of the highest in the continent.

Results We compared the antimicrobial susceptibility profiles and short-read genome sequences of five *H. pylori* isolates obtained from patients diagnosed with gastritis of varying severity (chronic gastritis, antral erosive gastritis, superficial gastritis) in Pereira, Colombia sampled in 2015. Antimicrobial susceptibility tests revealed the isolates to be resistant to at least one of the five antimicrobials tested: four isolates were resistant to metronidazole, two to clarithromycin, two to levofloxacin, and one to rifampin. All isolates were susceptible to tetracycline and amoxicillin. Comparative genome analyses revealed the presence of genes associated with efflux pump, restriction modification systems, phages and insertion sequences, and virulence genes including the cytotoxin genes *cagA* and *vacA*. The five genomes represent three novel sequence types. In the context of the Colombian and global populations, the five *H. pylori* isolates from Pereira were phylogenetically distant to each other but were closely related to other lineages circulating in the country.

Conclusions *H. pylori* from gastritis of different severity varied in their antimicrobial susceptibility profiles and genome content. This knowledge will be useful in implementing appropriate eradication treatment regimens for specific types of gastritis. Understanding the genetic and phenotypic heterogeneity in *H. pylori* across the geographical landscape is critical in informing health policies for effective disease prevention and management that is most effective at local and country-wide scales. This is especially important in Colombia and other South American countries that are poorly represented in global genomic surveillance studies of bacterial pathogens.

Keywords *Helicobacter pylori*, Gastritis, Genomics, Mobile genetic elements, Virulence, Antimicrobial resistance

[†]Adalucy Alvarez Aldana and Odion O. Ikhimiukor contributed equally to this work.

*Correspondence:
Cheryl P. Andam
candam@albany.edu

Full list of author information is available at the end of the article



Background

The Gram-positive *Helicobacter pylori* is an obligate, microaerophilic and fastidious bacterium that frequently colonizes the gastrointestinal mucosa of humans [1]. This opportunistic pathogen is known to colonize a significant portion of the world's population, ranging from 24.4% in Oceania to 79.1% in Africa and 63.4% in Latin America and the Caribbean [2, 3]. A recent systematic meta-analysis shows a global decline in the prevalence of *H. pylori* from 58.2% (1980 – 1990) to 43.1% (2011–2022); however, more than 40% of the world population is estimated to still be infected with *H. pylori* [4]. The wide variation that exists in *H. pylori* prevalence between regions is in part attributed to prevailing social and economic conditions [3, 5, 6]. *H. pylori* can colonize the stomach for decades without causing any detectable symptoms [1]. Nonetheless, colonization can progress to adverse complications. Infection with *H. pylori* is the most common cause of chronic gastritis, which can lead to more severe gastroduodenal pathologies such as peptic ulcer, mucosa-associated lymphoid tissue lymphoma, and gastric cancer [1]. *H. pylori* is the only bacterium currently classified as a Group 1 carcinogen by the International Agency for Research on Cancer (IARC) of the World Health Organization (WHO) due to it being the causative agent of 80% of all stomach cancers [7]. The US Centers for Disease Control and Prevention estimates that people with chronic *H. pylori* infection are at greater risk of developing stomach cancers when compared to uninfected people [8].

Treatment regimens for *H. pylori* infections often combine a strong acid suppressant (proton pump inhibitor [9]) and antibiotics such as clarithromycin and amoxicillin [10, 11]. Metronidazole has also been used as an alternative [12]. However, antimicrobial resistance is the most important factor in treatment failure of *H. pylori* infection. Worldwide, rates of resistance against these three primary antimicrobial agents have dramatically increased up to 20 – 30% [13, 14], threatening the first-line defense against eradicating *H. pylori*. As such, the WHO recommends greater efforts for research and development on novel therapeutics, screening and prevention strategies against *H. pylori* infections [7]. Effective eradication of *H. pylori* requires vigilance in implementing antimicrobial susceptibility testing, resistance surveillance, and antibiotic stewardship.

H. pylori displays extensive genetic diversity [15, 16] and a characteristic population structure that is shaped by their intimate co-evolution with their human hosts in the past ~ 100,000 years [17–19]. This has resulted to distinct geographically partitioned subpopulations that parallel the evolutionary history of human ethnic groups [15, 16]. Its highly adaptive nature lies in part to its natural

competence for DNA uptake and recombination [20, 21]. Understanding the genetic heterogeneity in *H. pylori* between geographical regions is critical to inform future health policy on prevention and management that is most effective at local and country-wide scales.

In this study, we used in vitro antimicrobial susceptibility testing and whole genome sequencing to characterize the virulence and antimicrobial resistance features and mobile genetic elements of five *H. pylori* isolates sampled from gastritis patients in a specialized medical center in Pereira, Colombia. Gastritis or the inflammation of the stomach lining can be classified into different types depending on the severity [22]. The isolates in our study came from chronic gastritis, antral erosive gastritis, and superficial gastritis. To place the Pereira isolates in the broader geographical context, we also carried out comparative genomics with publicly available genomes derived from other parts of Colombia and with the global population. This work contributes to the growing body of knowledge of *H. pylori* pathogenesis that will help inform current efforts of devising effective treatment and prevention strategies, especially in countries that are poorly represented in global genomic surveillance studies of bacterial pathogens.

Methods

Bacterial sampling and isolation

The study included 20 *H. pylori* isolates randomly selected from a previously published surveillance study carried out in 2015 [23]. Isolates were derived from dyspeptic patients in a specialized medical center in the city of Pereira, Colombia. One isolate was obtained from each patient. The surveillance project was approved by the Bioethics Committee of the Universidad Tecnológica de Pereira (UTP), which approved the informed consent before the start of the surveillance. No inclusion or exclusion criteria of patients were used. All individuals accepted voluntarily and signed the informed consent. *H. pylori* was cultured from gastric biopsy samples of the antrum and body. Samples were stored in brain–heart infusion (BHI) broth supplemented with 20% glycerol and antibiotics (vancomycin 10 mg/l, polymyxin B 0.33 mg/l, bacitracin 1.07 mg/l, and amphotericin B 5 mg/l). The biopsy specimens were mixed with sterile saline solution and macerated with a homogenizer (Deltaware Pellet Pestle). A total of 100 µl of each mashed solution was plated onto tryptic soy agar (TSA; Oxoid or Merck), supplemented with sheep blood (7%), isovitalax (0.5%), and the same antibiotics/concentrations were used for the BHI broth medium during transport and were incubated under microaerophilic conditions (5% O₂, 10% CO₂ and 85% N₂) at 37 °C

for 5 – 14 days. Colonies were confirmed with Gram staining, biochemical tests (positive urease, catalase and oxidase test), and *ureC*-PCR.

Antimicrobial susceptibility testing

H. pylori isolates obtained from the primary culture were subcultured on non-selective TSA (Oxoid or Merck) with 7% sheep blood, 0.5% isovitalax (BBL) under microaerophilic conditions at 37 °C for 48 h. An E-test (AB BIO-DISK North American Inc., Piscataway, New Jersey, USA) for metronidazole, clarithromycin, amoxicillin, levofloxacin, rifampicin, and tetracycline was carried out. Suspensions from pure 48-h subcultures were prepared in *Brucella* broth supplemented with 0.5% isovitalax and inoculum turbidity was adjusted to McFarland 3.0 standard. Thereafter, they were inoculated onto TSA plates supplemented with sheep blood (7%), isovitalax (0.5%) and without antibiotics. E-test strips were placed and incubated under microaerophilic conditions at 37 °C for 72 h. We measured the minimum inhibitory concentration (MICs) and used *H. pylori* strain ATCC 43504 as a control. MIC for clarithromycin was interpreted based on Clinical & Laboratory Standards Institute (CLSI) breakpoint (> 1.0 mg/l resistant) [24]. We used EUCAST breakpoints for the antibiotics amoxicillin (> 0.125 mg/l), levofloxacin (> 1 mg/l), tetracycline (> 1 mg/l), rifampin (> 1 mg/l), and metronidazole (> 8 mg/l) [25].

DNA extraction and whole genome sequencing

Genomic DNA was extracted from 48-h-old confluent cells obtained from pure cultures using QIAamp DNA Mini and Blood Mini kit following the manufacturer's guidelines (QIAGEN, Hilden, Germany). A Qubit fluorometer (Invitrogen, Grand Island, New York, USA) was used to measure DNA concentration. Sequencing was carried out at SeqCoast Genomics (Portsmouth, New Hampshire, USA). Briefly, DNA libraries were prepared for whole genome sequencing using the Illumina DNA Prep tagmentation kit and unique dual indexes. DNA samples were sequenced as multiplexed libraries on the Illumina NextSeq 2000 platform operated per the manufacturer's instructions. Sequencing was done using a 300-cycle flow cell kit to produce 2 × 150 bp paired reads. A 1–2% PhiX control was spiked into the run to support optimal base calling. Read demultiplexing, read trimming, and run analytics were performed using DRAGEN v.3.10.12, an on-board analysis software on the NextSeq2000. Quality of raw sequence data was assessed using FastQC to ascertain that > 85% of bases were higher than the Q30 Phred quality score.

Assembly, quality check, and annotation of draft genomes

Short reads generated by Illumina sequencing were assembled de novo into contiguous sequences using shovill v1.1.0 (<https://github.com/tseemann/shovill>). Adapter sequences were trimmed by enabling the -trim flag on shovill. Shovill also implements methods for subsampling read depth down to 150X, trimming adapters, correcting sequencing errors, and assembling using SPAdes v.3.14.1 [26]. Sequence quality of draft assemblies was assessed using QUAST v.5.0.2 [27] and CheckM v.1.1.3 [28] (Supplementary Table S1). Genomes with < 90% completeness and > 5% contamination, as recommended by CheckM [28], were excluded from downstream analyses. In all, five genomes from Pereira passed the sequence quality criteria (Supplementary Table S1).

The mash sketch function (<https://mash.readthedocs.io/en/latest/>) on mash v.1.1 [29] was used on all available reference *H. pylori* genomes in the National Center for Biotechnology Information (NCBI) database (https://ftp.ncbi.nlm.nih.gov/genomes/refseq/bacteria/Helicobacter_pylori/) to retrieve the closest matching reference genome. The results of the mash distance calculation showed that reference genome strain MT5135 (NZ_CP071982.1) was the most closely related to our five Pereira genomes. Using fastANI v.1.32 [30], we compared our genomes to reference MT5135 (NZ_CP071982.1) by calculating the genome-wide Average Nucleotide Identity (ANI) values. An ANI threshold of ≥ 95% was used to confirm that the five genomes belong to the same species [30]. Genomes were also confirmed on the type strain genome server of the Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures GmbH (<https://tygs.dsmz.de>). The draft assemblies were annotated using Prokka v.1.14.6 [31] and Bakta v.1.4.0 [32].

In silico sequence typing and identification of mobile genetic elements and genes encoding antimicrobial resistance, virulence, restriction enzymes, transposases, and phages

Sequence types (ST) were determined using allelic variation in seven housekeeping genes (*atpA*, *efp*, *mutY*, *ppa*, *trpC*, *ureI*, *yphC*) [33] and compared to the multi-locus sequence typing (MLST) database for *H. pylori* in pubMLST (<https://pubmlst.org/organisms/helicobacter-pylori/>) [34]. This was done by scanning the contigs of draft assemblies using mlst v.2.19.0 (<https://github.com/tseemann/mlst>). Genomes with novel combination of the seven multi-locus alleles were submitted to the *H. pylori* pubMLST database for curation and ST assignment [34].

We determined the presence, types, and abundance of specific genetic elements in the *H. pylori* genomes. Sequences of genes involved in restriction-modification

(RM) were retrieved from the REBASE database (<http://rebase.neb.com/rebase/>) [35] and used as a database in ABRicate v.1.0.1 (<https://github.com/tseemann/abricate>) to screen our draft assemblies for the presence of RM genes. To determine the presence of antimicrobial resistance genes in our genomes, two tools were used – AMRFinder-Plus v.3.10.23 implemented in NCBI and its accompanying NCBI-compiled antimicrobial resistance database [36] and ABRicate v.1.0.1 tool together with the Comprehensive Antimicrobial Resistance Database [37]. The Virulence Factor Database [38] implemented within ABRicate v.1.0.1 was used to screen the genome assemblies for virulence-associated genes. Next, the draft genomes were screened for the presence of EPIYA (Glu-Pro-Ile-Tyr-Ala) motifs in the Bakta [32]-translated sequences of the CagA toxin. The sequences were extracted using samtools v.1.15.1 [39]. We aligned the amino acid sequences using CLUSTALW [40] implemented in MEGA11 [41]. The alignments were inspected for the presence of the EPIYA motifs EPIYA-A (EPIYAKVNKKK(A/T)GQ), EPIYA-B (EPIY(A/T)QVAKKVNNAKI), EPIYA-C (EPIYATIDDLGGP), and EPIYA-D (EPIYATIDFDEANQAG) [42]. The presence of transposases in our genomes was determined using ISEScan v.1.7.2.3 [43]. ISEScan contains a built-in HMMER v.3.3.2 [44], which implements a probabilistic profile hidden Markov models (profile HMMs) to search a database of transposase profiles against our translated draft genomes. We implemented the ISEScan default settings of $<1e-5$ and $\geq 60\%$ for e-value and sequence coverage, respectively. This threshold is often utilized in IS detection (for examples, [45, 46]). Lastly, our genomes were examined for the presence of phage elements using VirSorter2 [47]. The length of individual phage DNA detected in each genome was summed to give the total length of all phage DNA segments in each genome.

All publicly available *H. pylori* genomes derived from human sources in Colombia were retrieved from the EnteroBase *H. pylori* database (available as of June 2024). Recent genomes submitted to the pubMLST database (<https://pubmlst.org>) from 2019 onwards (available as of June 2024) were also included. Altogether, the Colombian dataset included the five Pereira genomes and 259 genomes from six other regions sampled between 2000 – 2022. We compared the five Pereira genomes that we sequenced with genomes from other parts of Colombia for the presence of the genetic elements described above.

Phylogenetic tree reconstruction

To provide a broader phylogenetic and geographical context for the Pereira genomes, all the publicly available *H. pylori* genome assemblies derived from human sources worldwide and associated metadata from the Enterobase

H. pylori database ($n=3,285$ genomes; downloaded in June 2024) [48] and pubMLST as described earlier, were used to create a phylogeny. These genomes were assembled and annotated using the same methods as those used for the five Pereira genomes sequenced in this study. Using the combined 3,285 downloaded genomes and the five Pereira genomes, split k-mer distances was calculated using Split-Kmer Analysis (SKA) (<https://github.com/simonrharris/SKA>). Split k-mers refers to any two k-mers in a sequence separated by one or more bases. The use of SKA in epidemiological purposes have been previously validated and benefits from speed and less computational requirements [49, 50]. The reference-free alignment generated from the split k-mer files was used as input in IQ-TREE v.2.1.4 [51] to build a maximum likelihood phylogeny, implementing the general time reversible (GTR; [52])+Gamma model of nucleotide substitution model. Branch support was assessed using 1,000 bootstrap replicates implemented using the built-in ultrabootstrap -UFBoot software [53]. The phylogenetic tree was visualized and rooted at the midpoint using Figtree v.1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) and annotated using the Interactive Tree of Life (iTOL) v6.8.1 [54]. The SKA alignment and phylogenetic tree were used as input to partition the genomes into sequence clusters of genetically similar individuals using the Bayesian hierarchical clustering algorithm fastBAPS v.1.0.8 [55]. FastBAPS implements a rapid initial agglomerative clustering of the aligned sequences based on pairwise a single nucleotide polymorphism distance matrix, and followed by Bayesian model selection to decide which clusters should be merged at each step of hierarchical clustering [55].

Statistical analysis

All statistical tests was carried out using the `stat_compare_means` option in `ggpubr` v.0.4.0 package in RStudio v.2022.02.1 + 461 [56]. We used the Wilcoxon signed rank test to compare the number of virulence genes, genes encoding restriction-modification enzymes, and relative phage genome sizes in *H. pylori* isolated from different regions in Colombia. Results were considered significant when $p < 0.05$.

Results

Antimicrobial susceptibility profiles, genome characteristics, and sequence types of Colombian *H. pylori*

Five *H. pylori* isolates were obtained from unique patients diagnosed with varying levels of gastritis in a specialized center in Pereira, Colombia in 2015. Three of the isolates were recovered from patients with chronic gastritis (RPe_4, RPe_33, RPe_46), one isolate from a patient with antral erosive gastritis (RPe_43), and one isolate from a patient with superficial gastritis (RPe_23)

(Supplementary Table S1). CLSI and EuCAST breakpoint interpretation of MIC from antimicrobial susceptibility tests revealed the isolates to be resistant to at least one of the five antimicrobials tested (Fig. 1a and Supplementary Table S2). Four isolates were resistant to metronidazole, whereas two isolates each were resistant to clarithromycin and levofloxacin, respectively. Rifampin resistance was observed in only one isolate, RPe_43. In contrast, all isolates were susceptible to tetracycline and amoxicillin.

Genome assemblies of the five isolates yielded a range of 30 – 232 contigs per genome, largest contig per genome ranging from 113,302 – 250,354 bp, total genome length of 1,552,597 – 1,789,322 bp, N50 ranging

from 27,500 – 162,261 bp, and number of protein-coding genes per genome ranging from 1,397 – 1,496 (Supplementary Table S1). We sought to determine if these five isolates were related to known *H. pylori* clonal lineages based on the allele configuration of their seven house-keeping genes [33]. MLST results showed that all five genomes represented novel STs and were subsequently assigned ST designations on the PubMLST *H. pylori* database [34]. These novel STs were ST4353 (RPe_4), ST4354 (RPe_23), ST4355 (RPe_33 and RPe_46), and ST4356 (RPe_43) (Fig. 1a and Supplementary Table S2).

To place the five Pereira genomes in the broader Colombian population, the Colombian dataset was

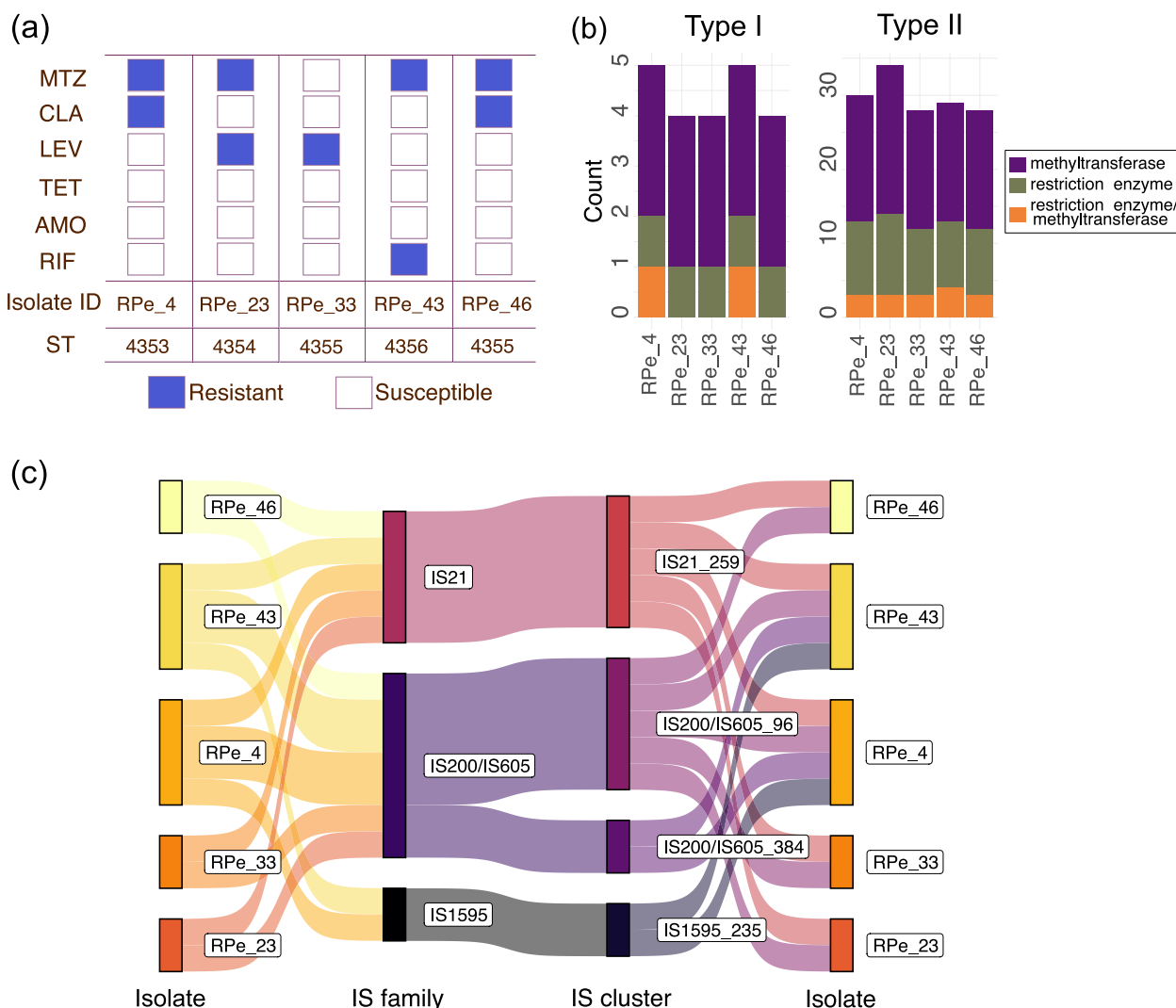


Fig. 1 Antimicrobial susceptibility profile and genomic features of the five gastritis isolates in this study. **a** Results of the antimicrobial susceptibility tests against six antimicrobial agents. Acronyms of antimicrobial agents: MTZ – Metronidazole; CLA – Clarithromycin; LEV – Levofloxacin; TET – Tetracycline; AMO – Amoxicillin; RIF – Rifampin. **b** Number and types of genes associated with Types I and II Restriction Modification systems. **c** Types of Insertion Sequence (IS) families and IS clusters

broadened to include all publicly available *H. pylori* genome sequences from Colombia derived from human samples ($n=259$ genomes) for a total of 264 genomes. A total of 163 known STs was detected in this larger Colombian population (Supplementary Table S3). The most prevalent STs were ST3809 (10 genomes), ST415 (8 genomes), and ST3806, ST3807, ST2811, ST3816, and ST3945 having four genomes each. Two STs were represented by three genomes each, 47 STs were represented by two genomes each, and 106 STs were represented by one genome each. Twenty genomes had unknown ST information.

A search for known genetic mechanisms of antimicrobial resistance present in the five Pereira genomes revealed the presence of the chromosomally located gene *hp1181* that encodes the efflux pump, in all five genomes. However, its presence does not interpret to an antibiotic resistant phenotype in *H. pylori* [57]. This gene was also detected in all other genomes from Colombia ($n=259$ genomes; Supplementary Table S4). A small fraction of genomes carried a quinolone resistance gene *qnrB5* in four genomes (two from Bogota and two from unknown location) and beta-lactam resistance gene *bla*_{TEM-181} in seven genomes from Nariño.

Distribution of virulence determinants and restriction modification systems

We examined the five genomes we sequenced for the presence of virulence-related genetic determinants. A total of 120 virulence-related genetic factors were detected in our genomes (range: 88 – 117 genes), of which 85 are present in all five genomes (Supplementary Table S5). Among the major virulence factors detected were those associated with adherence, effector delivery system (type IV secretion system [T4SS]), motility (flagella), exotoxin (*vacA*), immune modulation (lipopolysaccharide [LPS], outer inflammatory protein, Lewis antigen), and stress survival (urease). Eleven LPS-associated genes, which functions in innate immune response and evasion [58], were detected in all five genomes. These include HP_RS07005, *gluE*, *gluP*, *kdtB*, *lpxB*, *wbcJ*, *wbpB*, *rfaC*, *rfaI*, *rfaD*, and *rfaM*.

Two isolates RPe_4 (ST4353) and RPe_43 (ST4356) carried similar virulence-related genes. Both harbored the cytotoxin associated gene A (*cagA*) which is implicated in gastric carcinoma in humans [59]. The gene product CagA is introduced into the gastric epithelial cells through the T4SS carried on a pathogenicity island [60]. The T4SS genes were present in only these two genomes. The activity of the CagA toxin is regulated by tyrosine phosphorylation of the C-terminal EPIYA (Glu-Prp-Ile-Tyr-Ala) motifs inside host cells [61, 62]. There is considerable variation in the EPIYA motifs and such variation is

known to determine the clinical outcome of the disease [61–63]. The two genomes RPe_4 (ST4353) and RPe_43 (ST4356) harbored the variants EPIYA-A, EPIYA-B and EPIYA-C (Supplementary Figure S1), which is consistent with EPIYA patterns reported in *H. pylori* in Europe, Americas, and Australia [61]. Importantly, there were no repeats of the EPIYA-C in genomes from this study. The presence of fewer EPIYA-C repeats is associated with reduced risk of gastric cancer [62, 63]. Both genomes also carried the vacuolating cytotoxin gene (*vacA*), which is responsible for the vacuolation and induction of apoptosis and disruption of cellular pathways [64]. When we compared the Pereira genomes with genomes of *H. pylori* from other regions in Colombia (only those represented by ≥ 5 genomes), no significant difference was observed ($p > 0.05$, Wilcoxon signed rank test; Supplementary Figure S2) in the number of virulence genes per genome, except for genomes between Pereira and Tumaco (range = 117 – 119 virulence genes, $p = 0.0032$).

We examined the five genomes for the presence of genes involved in RM systems. RM enzymes play a critical role in bacterial defense against invasion by foreign DNA such as bacteriophages and conjugative plasmids [65]. Type I and II RMs were detected in all five genomes (Fig. 1b, Supplementary Table S6 and Table S7). For the type I RMs, the presence of either four (RPe_23, RPe_33 and RPe_46) or five (RPe_4 and RPe_43) genes encoding RM enzymes per genome were identified. Type I methyltransferases (M.Hpy99XX and M.Hpy99XV) and restriction endonuclease (Hpy99XVI) were present in all five genomes (Supplementary Table S6). Type II RMs system were also abundant in our five genomes. We identified genes encoding a total of 61 RM enzymes (Supplementary Table S7). The enzymes comprised of 37, 17 and seven unique methyltransferase, restriction enzymes, and restriction enzyme/methyltransferase, respectively. The number of genes encoding RM enzymes present per genome ranged from 28 (RPe_46) to 34 (RPe_23). Among the Type II enzymes, four methyltransferases (M.HpyUM037VIII, M.Hpy99III, M.Hpy99XII, M.Hpy99XI) and six restriction endonucleases (Hpy299IX, Hpy8I, HpyCH4V, HpyGI, HpyC1I, HpyCH4I) were detected in all five genomes. The only Type III RM enzyme detected was a methyltransferase (M.Hpy99XXI) carried in RPe_43 genome (Supplementary Table S8). No Type IV RM gene was detected in any of the five genomes. When genomes from Pereira were compared with genomes of isolates from other regions in Colombia (only those represented by ≥ 5 genomes), no significant difference was observed in the number of type I and type II RM genes per genome ($p > 0.05$, Wilcoxon signed rank test; Supplementary Figure S3, Tables S7 and S8). Type III RMs were detected in only a small fraction

(16.21%; 42/259 genomes) of genomes from other regions in Colombia (Supplementary Table S8).

Mobile genetic elements in Colombian *H. pylori* genomes

We sought to characterize the presence of different mobile genetic elements (MGE) that may contribute to DNA mobility and genome plasticity in *H. pylori*. A total of eight (range = 1 – 3 per genome) full and linear phage DNA were detected across the five genomes (Supplementary Table S9). The highest number of phage DNA detected was three and two in RPe_43 and RPe_23, respectively, whereas one phage DNA was detected in each of the other three genomes. All phage DNA were single stranded and linear, with the exception of two double stranded DNA segments detected in the RPe_43 genome. The total sizes of phage DNA as a proportion of the genome ranged from 0.1% (1,585 bp in RPe_33) to 0.6% (9,884 bp in RPe_43). The total length of phage DNA per genome from Pereira versus other regions in Colombia (represented by ≥ 5 genomes harboring phage DNA) showed no significant difference ($p > 0.05$, Wilcoxon signed rank test; Supplementary Figure S4).

Next, we search for the presence of insertion sequences (IS), which are transposable elements that encode transposases that catalyze intra- or inter-genome mobility [43]. The five genomes from this study harbored IS elements belonging to IS21 (IS21_259 IS cluster) and IS200/IS604 (IS200/IS604_96 and IS200/IS604_384 cluster) families (Fig. 1c and Supplementary Table S10). IS elements belonging to IS1595 (IS1595 cluster) family were present only in genomes RPe_4 and RPe_43. No plasmid replicon was detected in any of the five *H. pylori* genomes sequenced in this study.

Phylogenetic relationship of Colombian *H. pylori* with the global population

To place our five Pereira genomes in the broader geographical and phylogenetic context, we determined their evolutionary relationship with 3,285 publicly available genomes sampled from 60 countries (Fig. 2a,b and Supplementary Table S3). Majority of the genomes were collected from China ($n = 629$), Germany ($n = 500$), Australia ($n = 328$) and Colombia ($n = 264$, including the

five genomes from this study), representing 52.3% of the entire dataset (Fig. 2a). Genomes from this global dataset were derived from a sampling period between 1983 to 2022.

Analysis of MLST profiles showed 643 unique STs in 1,116 genomes, whereas no ST information was available for 66% ($n = 2,174$) of the genomes. We used Bayesian clustering algorithm implemented in fastBAPS [55] to delineate genomes into genetically similar phylogenetic sequence clusters. A total of 25 sequence clusters were observed in the global dataset. The most dominant were sequence clusters 5 ($n = 1,156$ genomes), 18 ($n = 1,134$ genomes), 2 ($n = 252$ genomes), 11 ($n = 142$ genomes), and 6 ($n = 113$ genomes), which altogether represented 85% of the global dataset (Fig. 2a). The five Pereira genomes sequenced in this study were found in different locations on the tree, but all are part of sequence cluster 18. The Pereira genomes were closely related to other genomes from Europe, North America, and South America, including the genomes from other parts of Colombia. The different sequence clusters appear to be differentially distributed across continents (Fig. 2b). For example, sequence cluster 18 was common in Africa ($n = 111/150$ genomes, 74%), Europe ($n = 226/880$ genomes, 25.68%), North America ($n = 362/442$ genomes, 81.90%), and South America ($n = 310/345$ genomes, 89.85%). In contrast, sequence cluster 5 was common in Asia ($n = 963/1092$ genomes, 88.18%) and sequence cluster 11 in Oceania ($n = 123/337$ genomes, 36.49%). Overall, the five Pereira genomes sequenced in this study exhibited close genetic similarity with other genomes from Colombia as well as other members of sequence cluster 18.

Discussion

H. pylori infection is particularly concerning in Colombia, where $> 80\%$ of the population is estimated to be infected with *H. pylori* [66] and virulent genotypes reach proportions of $> 90\%$ [67]. High levels of *H. pylori* contamination have also been reported in water from the Bogotá River and domestic wastewater treatment plants, which also pose grave risk to communities who re-use the water [68]. Age-standardized incidence and mortality rates of stomach cancer in Colombia is among the

(See figure on next page.)

Fig. 2 Population structure and phylogenetic relationships of the global *H. pylori* population. **a** Maximum likelihood phylogenetic tree built using split k-mer distances. The dataset includes the five Pereira genomes sequenced in this study (marked by red stars) and 3,285 publicly available genomes. The tree is rooted at the midpoint. The scale bar represents the number of nucleotide substitutions per site. The branches are shaded to delineate the five dominant sequence clusters determined by fastBAPS. Outer rings (innermost to outermost) show the continent, country of origin, and year of sampling. **b** Continent and country-wide distribution of *H. pylori* genomes belonging to fastBAPS sequence clusters. Bar charts show the number of genomes of the five dominant sequence clusters. For clarity, only countries with ≥ 10 genomes are shown. Sequence clusters and countries with few genomes were grouped together in the category "Others"

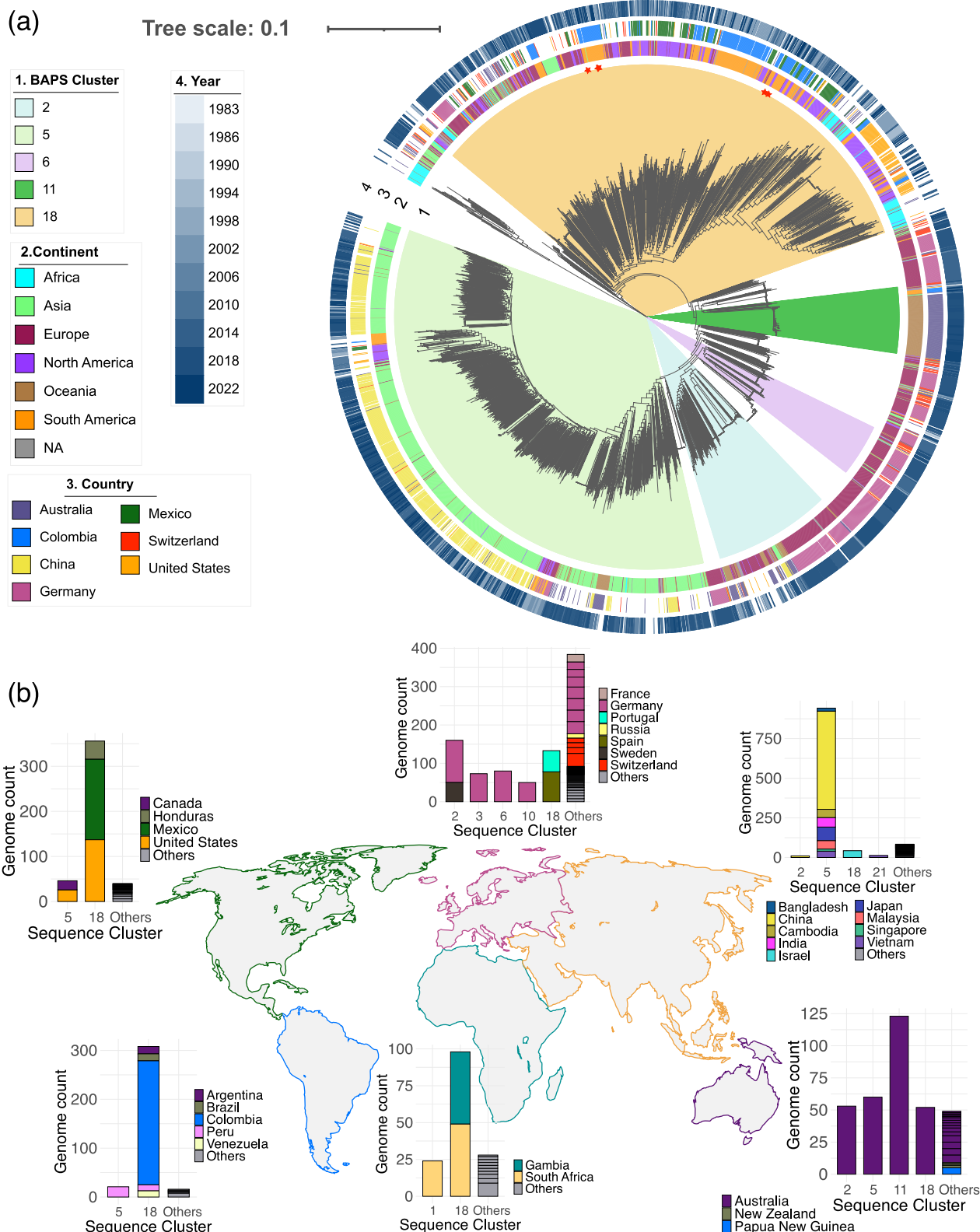


Fig. 2 (See legend on previous page.)

highest in Central and South American countries (25.3 and 17.8 per 100,000 people, respectively) [69]. *H. pylori* has been implicated in the progression of gastric precancerous lesions [70, 71]. It is also associated with gastric cancer and dysplasia in Colombian patients [72]. Continuous and vigilant monitoring of *H. pylori* genotypes is therefore critical to reduce the public health and clinical burden of *H. pylori* infection in the country.

Here, we report the antimicrobial susceptibility profiles and genome content of five isolates from patients diagnosed with varying severity of gastritis in Pereira, Colombia. The five isolates represent four novel STs. Our analysis augments current knowledge about the circulating genotypes in the country. Although our small Pereira dataset does not reflect country-wide prevalence of antimicrobial resistance phenotype, it presents valuable data that can be incorporated into current resistance surveillance and drug therapy efforts within the city of Pereira. Comparison of genomic features such as the frequency of virulence genes, RM enzymes, and phage DNA between our Pereira dataset and other regions in Colombia revealed no significant differences, indicating a lack of geographical structure in the gene pool of human-derived *H. pylori* across the country. These results show that only a few clonal lineages and specific clinically relevant genes are circulating in different parts of Colombia, which may reflect uniformity in selective pressures throughout the country. A more in-depth investigation of social attributes and antimicrobial usage within Colombia will clarify the factors contributing to the population genomic structure of *H. pylori* in the country.

A more extensive sampling and comparative genomic analysis of *H. pylori* isolates from different gastritis types will be particularly informative in elucidating the bacterial genetic basis of gastritis. Variants of the cytotoxin genes *cagA* and *vacA* have been previously proposed as risk markers in patients with premalignant gastric lesions [64, 73, 74]. Clinical manifestations of gastritis can range from mild cases to severe presentations associated with significant morbidity [22], and sequence variation in cytotoxin and other virulence genes may potentially be used to differentiate and predict gastritis types. Our findings of the presence of the Western type *cagA* EPIYA-ABC are consistent with previous reports [42, 61]. However, fewer EPIYA-C repeats were observed, and which has been previously reported to indicate reduced oncogenic risks [63]. Accurate diagnosis of oncogenic risks and implementation of appropriate treatment regimens that will be most effective for distinct clinical presentations is therefore necessary. Our results also provide an impetus for population-level genome surveillance efforts of gastritis-associated *H. pylori* in Colombia.

H. pylori utilizes a large arsenal of colonization and virulence factors to mediate the interaction with the human gastric niche and facilitate the onset of disease. Our results are consistent with previous reports on the high genetic diversity and virulence determinants of clinical *H. pylori* from other South American countries [75–78]. Genes encoding antimicrobial resistance genes, RM systems, and virulence, including those that encode cytotoxin genes were identified. The observed prevalence of the efflux pump gene *hp1181* in Colombian *H. pylori* is consistent with the results of a recent study of *H. pylori* in other locations [79]. The gene expression of *hp1181*, and possibly post-transcriptional regulation, has been reported to be associated with the resistant phenotype in multidrug resistant *H. pylori* [57]. Carriage of key virulence factors can affect the severity of gastroduodenal diseases and damage to the gastric epithelium by *H. pylori* [64]. The bacterial adoption of one or more mechanisms associated with adhesion, invasion, and toxin production are required to produce disease in the host [1]. In our study, *cagA* and *vacA* were detected in two strains resistant to metronidazole and clarithromycin or rifampin. The convergence of antimicrobial resistance phenotypes with the toxin determinants related to gastric lesions is troubling and these strains need to be closely monitored over the long term.

In addition to their pathogenicity attributes, *H. pylori* possess mechanisms to prevent adulteration of their DNA through the presence of RM systems. These unique defense mechanisms consist of restriction endonucleases, which functions to recognize and cleave DNA, and methyltransferase, which methylates DNA recognized by restriction endonucleases [80, 81]. Depending on co-factor requirements, the structure of recognition sequences, organization, and sequence specificity, RM systems may be classified into types I – IV [80, 81]. Our findings that the five genomes carry few mobile genetic elements was not surprising. This is because the presence of RMs acts as a barrier to the introduction of extrachromosomal DNA in *H. pylori* [65, 82]. Our results are consistent with previous studies of *H. pylori* showing that the majority of RMs in *H. pylori* are type II [83–85]. Furthermore, our results agree with previous reports that show strain-specific patterns in the number and variety of RMs in *H. pylori* [84, 85], as no two strains in our study possessed the exact complement of RM genes in their genomes. This strain-specific reservoir of RM genes phenomenon is suspected to play a role in the virulence of *H. pylori* [65, 80]. Future work is needed to further explore the under-appreciated contributions of RMs and whether the strain-specific RM distribution may be associated with pathogenicity and clinical presentations, which will

be beneficial in developing novel ways to treat specific *H. pylori* infections.

Whole genome sequencing of clinical isolates from poorly represented countries such as Colombia is integral to understanding the global population structure, mechanisms of infection and pathogenicity, disease burden, and outbreak detection of the opportunistic *H. pylori*. Our results on the phylogeny of Colombian and global *H. pylori* genomes are consistent with results from previous studies that show a close relationship with other countries as well as the evolution of independent clades within Colombia [86–88]. Our study offers valuable insights on the similarities of Pereira genomes and those from other regions in Colombia, which to our knowledge has not been reported before. Hence, our study should be considered as an important baseline census of the standing genomic and lineage diversity of *H. pylori* in different parts of the country. Our findings will inform current epidemiological efforts and spur countrywide efforts for a more systematic sampling scheme of *H. pylori* from a variety of clinical presentations, carriage, therapeutic interventions, hospitals, and environments in Colombia.

Conclusions

We present the draft genome sequences of five *H. pylori* isolates from patients diagnosed with gastritis in Pereira, Colombia. These genomes represent four novel STs. Their genome sequences reveal a variety of determinants of antimicrobial resistance and virulence, which can be mobilized by phages and IS elements. In the context of the Colombian and global *H. pylori* populations, the five isolates from Pereira were closely related to other lineages circulating in other parts of Colombia.

Abbreviations

IS	Insertion sequence
MLST	Multi-locus sequence typing
RM	Restriction-modification
ST	Sequence type

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10749-6>.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements

The authors thank the SUNY University at Albany Information Technology Services where all bioinformatics analyses were carried out. We thank Nicole Zac Soligno for her assistance in shipping the DNA samples.

Authors' contributions

C.P.A., O.O.I. and A.A.A. designed and guided the work. Y.M.G.G. and A.A.A. oversaw bacterial sampling, antimicrobial susceptibility testing, carried out metadata collection. O.O.I. performed all bioinformatics analyses. M.M.G., S.S.R.S. and A.B.G.B. assisted with the bioinformatics analyses. C.P.A. and O.O.I. wrote the initial manuscript. All authors have read, edited, and approved the final manuscript.

Funding

C.P.A. is supported by the National Institutes of Health (Award no. R35GM142924). The funders had no role in study design, data collection and analysis, decision to publish, and preparation of the manuscript. The findings in this study do not necessarily reflect views and policies of the authors' institutions and funders.

Availability of data and materials

The dataset supporting the conclusions of this article is included within the article and its supplementary files. Short reads data of the five *H. pylori* isolates in this study have been deposited in the Sequence Read Archive (SRA) of NCBI under BioProject accession number PRJNA1086786. The BioSample accession numbers are SAMN40411367, SAMN40411368, SAMN40411369, SAMN40411370, and SAMN40411371, which can be accessed in <https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA1086786>.

Declarations

Ethics approval and consent to participate

The five *H. pylori* isolates are part of the surveillance project "Characterization of the phenotype and genotype of antimicrobial profiles in *H. pylori* isolates obtained from dyspeptic patients from the Eje Cafetero". The study was approved by the Bioethics Committee of the Universidad Tecnológica de Pereira (Pereira, Colombia). All individuals in the study voluntarily accepted and signed the informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Grupo de Investigación en Microbiología y Biotecnología (MICROBIOTEC), Universidad Libre Seccional Pereira, Programa de Microbiología, Pereira, Colombia. ²Present Address: Grupo de Investigación en Enfermedades Infecciosas (GRIENI), Universidad Tecnológica de Pereira, Programa de Medicina, Pereira, Colombia. ³Present Address: Department of Biological Sciences, University at Albany, State University of New York, Albany, NY, USA. ⁴Molecular Epidemiology Laboratory, Federal University of Uberlândia, Minas Gerais, Brazil.

Received: 31 March 2024 Accepted: 30 August 2024

Published online: 09 September 2024

References

- Malfertheiner P, Camargo MC, El-Omar E, Liou J-M, Peek R, Schulz C, et al. Helicobacter pylori infection. *Nat Rev Dis Primers*. 2023;9:19.
- Sjomina O, Pavlova J, Niv Y, Leja M. Epidemiology of Helicobacter pylori infection. *Helicobacter*. 2018;23(Suppl 1): e12514.
- Hooi JKY, Lai WY, Ng WK, Suen MMY, Underwood FE, Tanyingoh D, et al. Global prevalence of Helicobacter pylori infection: systematic review and meta-analysis. *Gastroenterology*. 2017;153:420–9.
- Li Y, Choi H, Leung K, Jiang F, Graham DY, Leung WK. Global prevalence of Helicobacter pylori infection between 1980 and 2022: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol*. 2023;8:553–64.
- Zamani M, Ebrahmtabar F, Zamani V, Miller WH, Alizadeh-Navaei R, Shokri-Shirvani J, et al. Systematic review with meta-analysis: the worldwide prevalence of Helicobacter pylori infection. *Aliment Pharmacol Ther*. 2018;47:868–76.
- Shirani M, Pakzad R, Haddadi MH, Akrami S, Asadi A, Kazemian H, et al. The global prevalence of gastric cancer in Helicobacter pylori-infected individuals: a systematic review and meta-analysis. *BMC Infect Dis*. 2023;23:543.
- IARC Helicobacter pylori Working Group. Helicobacter pylori Eradication as a Strategy for Preventing Gastric Cancer. Lyon: World Health Organization. International Agency for Research on Cancer; 2014.
- Moss SF. The clinical evidence linking Helicobacter pylori to gastric cancer. *Cell Mol Gastroenterol Hepatol*. 2017;3:183–91.

9. Lerardi E, Losurdo G, Fortezza RFL, Principi M, Barone M, Leo AD. Optimizing proton pump inhibitors in *Helicobacter pylori* treatment: Old and new tricks to improve effectiveness. *World J Gastroenterol.* 2019;25:5097–104.
10. Vieira RR, Fontes LES, Pacheco RL, Fernandes MA, Malta PP, Riera R. Proton pump inhibitor- and clarithromycin-based triple therapies for *Helicobacter pylori* eradication. *Cochrane Database Syst Rev.* 2020;2020:CD013734.
11. Shih C-A, Shie C-B, Hsu P-I. Update on the first-line treatment of *Helicobacter pylori* infection in areas with high and low clarithromycin resistances. *Therap Adv Gastroenterol.* 2022;15:17562848221138168.
12. Puig I, Baylina M, Sánchez-Delgado J, López-Gongora S, Suarez D, García-Iglesias P, et al. Systematic review and meta-analysis: triple therapy combining a proton-pump inhibitor, amoxicillin and metronidazole for *Helicobacter pylori* first-line treatment. *J Antimicrob Chemother.* 2016;71:2740–53.
13. Mégraud F, Graham DY, Howden CW, Trevino E, Weissfeld A, Hunt B, et al. Rates of antimicrobial resistance in *Helicobacter pylori* isolates from clinical trial patients across the US and Europe. *Am J Gastroenterol.* 2023;118:269–75.
14. Savoldi A, Carrara E, Graham DY, Conti M, Tacconelli E. Prevalence of antibiotic resistance in *Helicobacter pylori*: a systematic review and meta-analysis in World Health Organization regions. *Gastroenterology.* 2018;155:1372–1382.e17.
15. Montano V, Didelot X, Foll M, Linz B, Reinhardt R, Suerbaum S, et al. World-wide population structure, long-term demography, and local adaptation of *Helicobacter pylori*. *Genetics.* 2015;200:947.
16. Thorell K, Muñoz-Ramírez ZY, Wang D, Sandoval-Motta S, Boscolo Agostini R, Ghirrotto S, et al. The *Helicobacter pylori* Genome Project: insights into *H. pylori* population structure from analysis of a worldwide collection of complete genomes. *Nat Commun.* 2023;14:8184.
17. Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, et al. Traces of human migrations in *Helicobacter pylori* populations. *Science.* 2003;299:1582–5.
18. Kodaman N, Pazos A, Schneider BG, Piazuolo MB, Mera R, Sobota RS, et al. Human and *Helicobacter pylori* coevolution shapes the risk of gastric disease. *Proc Natl Acad Sci U S A.* 2014;111:1455–60.
19. Thorpe HA, Tourrette E, Yahara K, Vale FF, Liu S, Oleastro M, et al. Repeated out-of-Africa expansions of *Helicobacter pylori* driven by replacement of deleterious mutations. *Nat Commun.* 2022;13:6842.
20. Baltzus DA, Guillemin K, Phillips PC. Natural transformation increases the rate of adaptation in the human pathogen *Helicobacter pylori*. *Evolution.* 2008;62:39–49.
21. Dorer MS, Sessler TH, Salama NR. Recombination and DNA repair in *Helicobacter pylori*. *Annu Rev Microbiol.* 2011;65:329–48.
22. Azer SA, Awosika AO, Akhondi H. Gastritis. In: *StatPearls*. Treasure Island: StatPearls Publishing; 2024.
23. Bedoya-Gómez IJ, Alvarez-Aldana A, Moncayo-Ortiz JI, Guaca-González YM, Santacruz-Ibarra JJ, Arturo-Arias BL, et al. Surveillance of the antimicrobial resistance rates of *Helicobacter pylori* ten years later in the Western Central Region. *Colombia Dig Dis.* 2020;38:196–203.
24. CLSI. Performance standards for antimicrobial susceptibility testing. CLSI supplement M100. 32nd ed. 32nd edition. Berwyn: Clinical Laboratory Standards Institute; 2022.
25. The European Committee on Antimicrobial Susceptibility Testing (EUCAST). Breakpoint tables for interpretation of MICs and zone diameters. Version 13.1. 2023.
26. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19:455–77.
27. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics.* 2013;29:1072–5.
28. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 2015;25:1043–55.
29. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 2016;17:132.
30. Jain C, Rodríguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun.* 2018;9:5114.
31. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30:2068–9.
32. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb Genom.* 2021;7: 000685.
33. Achtman M, Azuma T, Berg DE, Ito Y, Morelli G, Pan ZJ, et al. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol Microbiol.* 1999;32:459–70.
34. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* 2018;3:124.
35. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE: a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* 2023;51:D629–30.
36. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci Rep.* 2021;11:12728.
37. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res.* 2023;51:D690–9.
38. Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res.* 2022;50:D912–7.
39. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience.* 2021;10:giab008.
40. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;22:4673–80.
41. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol.* 2021;38:3022–7.
42. Rodríguez Gómez ER, Otero Regino W, Monterrey PA, Trespalcacios Rangel AA. *cagA* gene EPIYA motif genetic characterization from Colombian *Helicobacter pylori* isolates: Standardization of a molecular test for rapid clinical laboratory detection. *PLoS ONE.* 2020;15:e0227275.
43. Xie Z, Tang H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics.* 2017;33:3340–7.
44. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol.* 2011;7:e1002195.
45. Moura de Sousa JA, Pfeifer E, Touchon M, Rocha EPC. Causes and consequences of bacteriophage diversification via genetic exchanges across lifestyles and bacterial taxa. *Mol Biol Evol.* 2021;38:2497–512.
46. Xu S, Wei M, Li G, Li Z, Che Y, Han L, et al. Comprehensive analysis of the *Nocardia cyriacigeorgica* complex reveals five species-level clades with different evolutionary and pathogenicity characteristics. *mSystems.* 2022;7:e0140621.
47. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome.* 2021;9:37.
48. Zhou Z, Alikhan NF, Mohamed K, Fan Y, Agama Study Group, Achtman M. The Enterobase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity. *Genome Res.* 2020;30:138–52.
49. Maechler F, Weber A, Schwengers O, Schwab F, Denkel L, Behnke M, et al. Split k-mer analysis compared to cgMLST and SNP-based core genome analysis for detecting transmission of vancomycin-resistant enterococci: results from routine outbreak analyses across different hospitals and hospital networks in Berlin. *Germany Microb Genom.* 2023;9:mgen000937.
50. Ikhimiukor OO, Souza SSR, Akintayo IJ, Marcovici MM, Workman A, Martin IW, et al. Phylogenetic lineages and antimicrobial resistance determinants of clinical *Klebsiella oxytoca* spanning local to global scales. *Microbiol Spectr.* 2023;11:e0054923.
51. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020;37:1530–4.
52. Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. In: *American mathematical society: lectures on mathematics in the life sciences.* Amer Mathematical Society; 1986. p. 57–86.

53. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018;35:518–22.
54. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021;49:W293–6.
55. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Res*. 2019;47:5539–49.
56. RStudio Team. RStudio: Integrated Development for R. <http://www.rstudio.com/>. 2020.
57. Falsafi T, Ehsani A, Attaran B, Niknam V. Association of hp1181 and hp1184 genes with the active efflux phenotype in multidrug-resistant isolates of *Helicobacter pylori*. *Jundishapur J Microbiol*. 2016;9:e30726.
58. Sijmons D, Guy AJ, Walduck AK, Ramsland PA. *Helicobacter pylori* and the role of lipopolysaccharide variation in innate immune evasion. *Front Immunol*. 2022;13:868225.
59. Hatakeyama M. Oncogenic mechanisms of the *Helicobacter pylori* CagA protein. *Nat Rev Cancer*. 2004;4:688–94.
60. Backert S, Haas R, Gerhard M, Naumann M. The *Helicobacter pylori* Type IV Secretion System encoded by the cag pathogenicity island: architecture, function, and signaling. *Curr Top Microbiol Immunol*. 2017;413:187–220.
61. Higashi H, Tsutsumi R, Fujita A, Yamazaki S, Asaka M, Azuma T, et al. Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. *Proc Natl Acad Sci U S A*. 2002;99:14428–33.
62. Ferreira RM, Machado JC, Leite M, Carneiro F, Figueiredo C. The number of *Helicobacter pylori* CagA EPIYA C tyrosine phosphorylation motifs influences the pattern of gastritis and the development of gastric carcinoma. *Histopathology*. 2012;60:992–8.
63. Li Q, Liu J, Gong Y, Yuan Y. Association of CagA EPIYA-D or EPIYA-C phosphorylation sites with peptic ulcer and gastric cancer risks: A meta-analysis. *Medicine (Baltimore)*. 2017;96:e6620.
64. 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.
65. Ando T, Xu Q, Torres M, Kusugami K, Israel DA, Blaser MJ. Restriction-modification system differences in *Helicobacter pylori* are a barrier to interstrain plasmid transfer. *Mol Microbiol*. 2000;37:1052–65.
66. Campuzano-Maya G, Hoyos-Castaño D, Calvo-Betancur VD, Suárez-Ramírez OA, Lizcano-Cardona D, Rojas-Arbeláez CA [Prevalence of *Helicobacter pylori* infection in physicians in Medellín, Colombia]. *Acta Gastroenterol Latinoam*. 2007;37:99–103.
67. 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;23:800–9.
68. Vesga F-J, Beltrán-Benavides AR, Márquez-Duque AM, Venegas C, Trespalacios A-A. *Helicobacter pylori* virulence genotypes in Bogotá River and wastewater treatment plants in Colombia. *Helicobacter*. 2023;28: e13023.
69. Sierra MS, Cueva P, Bravo LE, Forman D. Stomach cancer burden in Central and South America. *Cancer Epidemiol*. 2016;44(Suppl 1):S62–73.
70. Mera RM, Bravo LE, Camargo MC, Bravo JC, Delgado AG, Romero-Gallo J, et al. Dynamics of *Helicobacter pylori* infection as a determinant of progression of gastric precancerous lesions: 16-year follow-up of an eradication trial. *Gut*. 2018;67:1239–46.
71. Carlosama-Rosero Y, Acosta-Astaiza C, Sierra-Torres CH, Bolaños-Bravo H, Quiroga-Quiroga A, Bonilla-Chaves J. Virulence genes of *Helicobacter pylori* increase the risk of premalignant gastric lesions in a Colombian population. *Can J Gastroenterol Hepatol*. 2022;2022:7058945.
72. Carlosama-Rosero YH, Acosta-Astaiza CP, Sierra-Torres CH, Bolaños-Bravo HJ. *Helicobacter pylori* genotypes associated with gastric cancer and dysplasia in Colombian patients. *Rev Gastroenterol Mex (Engl Ed)*. 2022;87:181–7.
73. Bridge DR, Merrell DS. Polymorphism in the *Helicobacter pylori* CagA and VacA toxins and disease. *Gut Microbes*. 2013;4:101–17.
74. Panayotopoulou EG, Sgouras DN, Papadakis KS, Petraki K, Breurec S, Michopoulos S, et al. CagA and VacA polymorphisms are associated with distinct pathological features in *Helicobacter pylori*-infected adults with peptic ulcer and non-peptic ulcer disease. *J Clin Microbiol*. 2010;48:2237–9.
75. Benigno TGDS, Ribeiro Junior HL, Azevedo OGR, Pinheiro RF, Oliveira RTG, Maciel FS, et al. Clarithromycin-resistant *H. pylori* primary strains and virulence genotypes in the Northeastern region of Brazil. *Rev Inst Med Trop Sao Paulo*. 2022;64:e47.
76. Bustos-Fraga S, Salinas-Pinta M, Vicuña-Almeida Y, de Oliveira RB, Baldeón-Rojas L. Prevalence of *Helicobacter pylori* genotypes: cagA, vacA (m1), vacA (s1), babA2, dupA, iceA1, oipA and their association with gastrointestinal diseases. A cross-sectional study in Quito-Ecuador. *BMC Gastroenterol*. 2023;23:197.
77. Gutiérrez-Escobar AJ, Velapatiño B, Borda V, Rabkin CS, Tarazona-Santos E, Cabrera L, et al. Identification of New *Helicobacter pylori* Subpopulations in Native Americans and Mestizos From Peru. *Front Microbiol*. 2020;11:601839.
78. González-Stegmaier R, Aguila-Torres P, Villarroel-Espíndola F. Historical and molecular perspectives on the presence of *Helicobacter pylori* in Latin America: a niche to improve gastric cancer risk assessment. *Int J Mol Sci*. 2024;25:1761.
79. Alfaray RI, Saruuljavkhan B, Fauzia KA, Torres RC, Thorell K, Dewi SR, et al. Global antimicrobial resistance gene study of *Helicobacter pylori*: comparison of detection tools, ARG and efflux pump gene analysis, worldwide epidemiological distribution, and information related to the antimicrobial-resistant phenotype. *Antibiotics (Basel)*. 2023;12:1118.
80. Vasu K, Nagaraja V. Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol Mol Biol Rev*. 2013;77:53–72.
81. Dimitriu T, Szczelkun MD, Westra ER. Evolutionary ecology and interplay of prokaryotic innate and adaptive immune systems. *Curr Biol*. 2020;30:R1189–202.
82. Humbert O, Salama NR. The *Helicobacter pylori* HpyAXII restriction-modification system limits exogenous DNA uptake by targeting GTAC sites but shows asymmetric conservation of the DNA methyltransferase and restriction endonuclease components. *Nucleic Acids Res*. 2008;36:6893–906.
83. Xu Q, Morgan RD, Roberts RJ, Blaser MJ. Identification of type II restriction and modification systems in *Helicobacter pylori* reveals their substantial diversity among strains. *Proc Natl Acad Sci U S A*. 2000;97:9671–6.
84. Lin LF, Posfai J, Roberts RJ, Kong H. Comparative genomics of the restriction-modification systems in *Helicobacter pylori*. *Proc Natl Acad Sci U S A*. 2001;98:2740–5.
85. Ailloud F, Gottschall W, Suerbaum S. Methylome evolution suggests lineage-dependent selection in the gastric pathogen *Helicobacter pylori*. *Commun Biol*. 2023;6:839.
86. Gutiérrez-Escobar AJ, Trujillo E, Acevedo O, Bravo MM. Phylogenomics of Colombian *Helicobacter pylori* isolates. *Gut Pathog*. 2017;9:52.
87. 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.
88. Guevara-Tique AA, Torres RC, Bravo NM, Carvajal Carmona LG, Echeverry de Polanco MM, Bohórquez ME, et al. Recombination events drives the emergence of Colombian *Helicobacter pylori* subpopulations with self-identity ancestry. *Virulence*. 2022;13:1146–60.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.