Received: 23 October 2020 | Revised: 30 November 2020 | Accepted: 10 December 2020

DOI: 10.1111/hel.12779

ORIGINAL ARTICLE

Helicobacter WILEY

An American lineage of *Helicobacter pylori* prophages found in Colombia

Angela B Muñoz[1,2] | Alba A Trespalacios-Rangel[1] | Filipa F Vale[2]

[1]Infectious Diseases Research Group, Microbiology Department, Sciences Faculty, Pontificia Universidad Javeriana, Bogotá, Colombia

[2]Host-Pathogen Interactions Unit, Faculdade de Farmácia, Research Institute for Medicines (iMed-ULisboa, Universidade de Lisboa, Lisbon, Portugal

Correspondence
Angela B Muñoz, Infectious Diseases Research Group, Microbiology Department, Sciences Faculty, Pontificia Universidad Javeriana, Bogotá, Colombia.
Emails: munozangela@javeriana.edu.co; angelamunoz7@hotmail.com

Filipa F Vale, Host-Pathogen Interactions Unit, Research Institute for Medicines (iMed-ULisboa), Faculdade de Farmácia, Universidade de Lisboa, Lisbon, Portugal.
Emails: f.vale@ff.ulisboa.pt; vale.filipa@gmail.com

Funding information
Departamento Administrativo de Ciencia, Tecnología e Innovación, Grant/Award Number: 120380763025/2018; Pontificia Universidad Javeriana, Grant/Award Number: PPTA_7676; Fundação para a Ciência e a Tecnologia, Grant/Award Number: CEECIND/03023/2017, PTDC/BTM-SAL/28978/2017, UIDB/04138/2020 and UIDP/04138/2020

Abstract

Background: *Helicobacter pylori* is a human gastric carcinogen that is highly prevalent in Latin American. The prophages of *H. pylori* show a structured population and contribute to the diversity of this bacterium. However, *H. pylori* prophages present in American strains have not been described to date. In this study, we identified, characterized, and present the phylogenetic analysis of the prophages present in Colombian *H. pylori* strains.

Methods: To characterize Colombian *H. pylori* strains and their prophages, a Multilocus Sequences Typing (MLST) and a Prophage Sequences Typing (PST), using the integrase and holin genes, were performed. Furthermore, five Colombian *H. pylori* had their full genome sequenced, and six Colombian *H.pylori* retrieved from databases, allowing to determine the prophage complete genome and insertion site.

Results: The integrase gene frequency was 12.6% (27/213), while both integrase and holin genes were present in 4.2% (9/213) of the samples analyzed. The PST analysis showed that Colombian prophages belong to different populations, including hpSWEurope, hpNEurope, hpAfrica1, and a new population, named hpColombia. The MLST analysis classified most of the Colombia strains in the hpEurope population.

Conclusions: The new *H. pylori* prophage population revealed that Colombian prophages follow a unique evolutionary trajectory, contributing to bacterial diversity. The global *H. pylori* prophage phylogeny highlighted five phylogenetic groups, one more than previously reported. After the arrival of Europeans, the Colombian *H. pylori* bacteria and their prophages formed an independent evolutionary line to adapt to the new environment and new human hosts.

KEYWORDS
Helicobacter pylori, phylogeography, prophage genetic diversity

1 | INTRODUCTION

Prophages, also known as temperate phages, are viral genomes that are integrated into the bacterial genome. Prophages can shape bacterial populations[1] and can influence bacterial evolution through horizontal gene transfer, promotion of gene transduction, or potential disruptive effect caused by integration. Prophages can also influence the bacterium's biological behaviors, including pathogenesis, environmental adaptation, virulence, and antibiotic resistance.[2,3] Recently, an increasing number of reports have revealed the presence of prophages in *Helicobacter pylori*.[4-13] *H. pylori* is a bacterium that colonizes the stomachs of approximately half of the world's population,[14] reaching prevalence rates up to 90% in developing countries,[15] reaching 80% in Colombia.[16] *H.*

pylori is recognized as the causative agent of chronic gastritis and peptic and duodenal ulcers and as the etiologic agent of gastric cancer.[17]

From the first *H. pylori* prophage report, in 2011 by Lehours et al.,[5] studies have demonstrated that *H. pylori* strains carry full or partial prophage sequences, and analyze integrase and holin gene sequences allows for geographical differentiation of strains.[5-7] These studies have shown that *H. pylori* display extensive intraspecies diversity[18] and that this diversity appears to be influenced by the presence of mobile genomic elements, including prophages.[9]

The Prophage Sequence Typing or PST method was reported by Vale *et al*. in 2015[7] and is based on determining the presence of integrase and holin genes to predict the presence of prophages in *H. pylori* strains. These genes have been used because they are found at both ends of the prophage. Therefore, they allow inferring about the presence of complete prophages. The integrase gene is responsible for the integration of the phage genome into the bacterial chromosome. It is usually placed at the 5' end, while the holin gene is involved in cell lysis when a lytic cycle occurs and is generally placed on the 3' end.[2] Some strains may have only one of these genes when they carry remnants or incomplete prophages.

The phylogenetic analysis of integrase and holin genes is a useful strategy to characterize the structure population prophage. To date, four *H. pylori* prophage populations have been described: hpAfrica1, hpEastAsia, hpNEurope, and hpSWEurope, which are comparable to those identified in bacteria using Multilocus Sequence Typing (MLST) analysis.[7]

However, these studies did not include strains from the American continent. Recent reviews of *H. pylori* genomes from Latin America reported the separation of American strains from those of European, African, Asian, and Amerindian origin. Moreover, frequent and recent recombination events have been detected in groups of exclusively Colombian, Mexican, and Nicaraguan strains.[19,20] *H. pylori* from American isolates results of genetic drift and admixture between *H. pylori* of European and African origin, rather than of Amerindian origin.[21] Currently, it is mostly unknown whether prophages occur in American strains, if prophage elements contribute to events that generate high levels of genetic diversity, to what prophage population they belong and if prophages share with the host bacteria the same ancestral origin.

In this study, we screened 213 *H. pylori* strains isolated from 175 Colombian patients to determine the prevalence of the integrase and holin prophage genes and identify the corresponding prophage population. Additionally, we characterized the complete prophage genome from Colombia *H. pylori* strains. We also searched for prophages in Colombian *H. pylori* genomes available in public databases to enrich our collection for phylogenetic analysis of the bacterial isolates and their prophages. These data are valuable to understand the prevalence of prophages in Colombian strains, their population structure, and ancestral origin.

2 | METHODS

2.1 | *H. pylori* strains

The presence of prophage holin and integrase genes was detected by PCR. DNA samples from 213 *H. pylori* strains isolated from 175 patients with functional dyspepsia and living in Bogotá, Colombia were used. The strains were obtained from previous studies performed in the Infectious Diseases research group at the Pontificia Universidad Javeriana in Bogotá, Colombia. Different strains were isolated from the same patient because they were obtained from biopsies taken on the same day but from different stomach locations.

Of the 175 patients from whom the strains were isolated, 72% (126) were women, and 28% (49) were men. They had an average age of 48 years (range: 18–79). Additionally, due to the histology result, 69.1% (121) of these patients were diagnosed with chronic non-atrophic gastritis, 20.6% (36) with chronic atrophic gastritis, and 10.3% had no information related.

From the 213 strains, total DNA was extracted following standard protocols,[22] using DNAzol® kits (Invitrogen, Carlsbad, CA, USA). The DNA samples were stored at −20 ° C until use for integrase and holin genes PCR.

For the characterization of complete prophages, the positive strains for the integrase and/or holin genes (Table 1) were cultured again for DNA extraction and sequencing using next-generation sequencing (NGS) technology. Following standard protocols,[23] we were able to recover by culture eight positive strains (Col-2PUJ, Col-5PUJ, Col-6PUJ, Col-7PUJ, Col-16PUJ, Col-17PUJ, Col-22PUJ, and Col 23-PUJ). Total DNA was extracted from these strains using a DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany) following the manufacturer's instructions. Fluorimetric assay DNA quantification was performed using a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, California, US). Each sample (1 µl) was examined using the Qubit dsDNA HS Assay Kit 0.2–100 ng/µl following the manufacturer's instructions. DNA samples were stored at −20 ° C until required for NGS. Genomic DNA from the eight selected strains was sequenced using Illumina MiSeq (Illumina, San Diego, CA) NGS; DNA libraries were prepared using a Nextera XT DNA Library Preparation Kit (Illumina), followed by 2 × 300 paired-end sequencing runs with ~80 times coverage. The reads were used for *de novo* genome assembly with SPAdes v13.3.[24]

Additionally, *H. pylori* strains carrying integrase and holin genes were searched in the National Center for Biotechnology Information (NCBI) public database. For this, a BLASTn analysis was performed on the *H. pylori* whole-genome shotgun (WGS) available at NCBI. A threshold limit of $<1e^{-6}$ was considered positive, and the genomes carrying these genes were selected (Table 2).

2.2 | *H. pylori* prophages search

All 213 genomic DNA samples were tested by PCR using GoTaq® Green Master Mix kit (Promega, Madison, WI) to

TABLE 1 Data from the present study Colombian strains presenting prophages.

N°	Strain	Integrase	Holin	Prophage determined by NGS	PST	MLST
1	Col 2- PUJ†	POS	NEG	Incomplete	hpColombia	hpEurope
2	Col 3- PUJ	POS	NEG	No NGS	hpSWEurope	ND
3	Col 4- PUJ	POS	NEG	No NGS	hpAfrica1	hpEurope
4	Col 5- PUJ†	POS	POS	Phage	hpAfrica1	hpEurope
5	Col 6- PUJ†	POS	POS	Phage	hpAfrica1	hpEurope
6	Col 7- PUJ†	POS	POS	Phage	hpAfrica1	hpEurope
7	Col 16- PUJ†	POS	NEG	Incomplete	hpColombia	hpEurope
8	Col 17- PUJ†	POS	NEG	Incomplete	hpColombia	hpEurope
9	Col 22- PUJ†	POS	POS	Phage	hpAfrica1	hpEurope
10	Col 23- PUJ†	POS	POS	Phage	hpAfrica1	hpEurope
11	Col 32- PUJ	POS	POS	No NGS	hpColombia	ND
12	Col 33- PUJ	POS	POS	No NGS	hpAfrica1	ND
13	Col 34- PUJ	POS	NEG	No NGS	hpColombia	ND
14	Col 35- PUJ	POS	NEG	No NGS	hpSWEurope	ND
15	Col 36- PUJ	POS	NEG	No NGS	hpSWEurope	ND
16	Col 37- PUJ	POS	NEG	No NGS	hpColombia	ND
17	Col 38- PUJ	POS	NEG	No NGS	hpSWEurope	ND
18	Col 39- PUJ	POS	POS	No NGS	hpAfrica1	ND
19	Col 40- PUJ	POS	NEG	No NGS	hpSWEurope	ND
20	Col 41- PUJ	POS	NEG	No NGS	hpSWEurope	ND
21	Col 42- PUJ	POS	POS	No NGS	hpAfrica1	ND
22	Col 43- PUJ	POS	NEG	No NGS	hpAfrica1	ND
23	Col 44- PUJ	POS	NEG	No NGS	hpColombia	ND
24	Col 45- PUJ	POS	NEG	No NGS	hpSWEurope	ND
25	Col 46- PUJ	POS	NEG	No NGS	hpAfrica1	ND
26	Col 47- PUJ	POS	NEG	No NGS	hpColombia	ND
27	Col 48- PUJ	POS	NEG	No NGS	hpColombia	ND

Note: Positive strains for Integrase and/or holin genes. Eight (8) isolates of these were NGS sequenced; these strains were highlighted with †. The table shows the result for PST and MLST analysis.

NGS=Next-Generation Sequences, POS=Positive, NEG=Negative, ND=No data

detect integrase and holin genes using the approach previously described by Vale et al.[7] Briefly, the degenerate primers used were for integrase, F1- AAGYTTTTTAGMGTTTTGYG, and R1- CGCCCTGGCTTAGCATC (Invitrogen, Carlsbad, CA), and for holin gene hol-F- CCATCCCGTATTTGTTGGTG and hol-R-ACCCAATGCCTCCACTAATC (Invitrogen, Carlsbad, CA). The expected PCR products sizes were 529 and 225 bp for integrase and holin, respectively. PCR was performed using the following conditions: initial denaturation at 95 °C for 4 minutes; followed by 35 cycles at 95 °C for 30 seconds, 58 °C for 30 seconds and 72 °C for 1 minute, and a final extension at 72 °C for 7 minutes. A DNA sample from the *H. pylori* B45 strain (first reported as carrying a prophage)[5] was donated by the Centre National de Référence des Campylobacters et Hélicobacters, Bordeaux, France and was used as a positive control.

The PCR products of strains positive for integrase and holin genes were purified and bidirectionally sequenced using an ABI PRISM 3500® analyzer DNA sequencer (Applied Biosystems, Foster City, CA).

To identify homology with other previously identified *H. pylori* phages, the genomes were subjected to analysis using the MEGABLAST strict search program (http://blast.ncbi.nlm.nih.gov/Blast.cgi), using KHP30 (accession number: NC_019928.1) or phiHP33 (NC_016568.1) phages as a reference. Prophage genomes were manually put together by mapping contigs with either KHP30 or phiHP33 genomes.

The complete phiHP33 phage sequence and some (fifteen) phages previously described by Vale *et al.*[8] were used as the reference for alignment and comparison with the phages found in the Colombian strains. phiHP33 was selected as a reference because it was the first *H. pylori* phage completely characterized. Additionally, by PST is classified in the hpAfrica1 population, making it closer to the Colombian prophages. The other phages that can be used (KHP30, KHP40, and 1961P) belong to the hpEastAsia

TABLE 2 Data from NCBI strains carried prophages.

			Integrase			Holin					
Accession number Genbank	Strain	Blast Database	Query cov.	E-value	% Id.	Query cov.	E-value	% Id.	Country	PST	MLST
MBGM01000000	HP3076	WGS†	94%	5,00E−145	0,8756	96%	3E−62	0,8612	Colombia	hpColombia	hpEurope
ASYV01000000	PZ5080	WGS	100%	2,00E−172	0,905	96%	1E−39	0,7696	Colombia	hpSWEurope	hpEurope
NZ_MTWM00000000.1	PZ5019_3A3	WGS	99%	2,00E−173	90.00%	96%	3.00E−45	78.57%	Colombia	hpAfrica1	hspWAfrica
NZ_MUOC00000000.1	CA22019	WGS	98%	2,00E−114	80.69%	-	-	-	Colombia	hpAfrica1	hspWAfrica
MUOL00000000.1	CA22362	WGS	100%	0	91.40%	-	-	-	Colombia	hpSWEurope	Hybrid
MUPB00000000.1	CG22371	WGS	100%	0	80.69%	-	-	-	Colombia	hpAfrica1	Hybrid
MUPQ00000000.1	CM22388	WGS	97%	1,00E−151	86.98%	100%	1.00E−50	79.72%	Colombia	hpAfrica1	hpEurope
MBGL00000000.1	3096	WGS	97%	9,00E−152	87.62%	-	-	-	Colombia	hpColombia	Hybrid
MBGL00000000.2	22386	WGS	100%	1,00E−46	69.96%	-	-	-	Colombia	hpSWEurope	Hybrid
MBGL00000000.3	NQ4044	WGS	99%	2,00E−126	82.76%	96%	2.00E−61	84.29%	Colombia	hpAfrica1	hpEurope
MBGL00000000.4	NQ1712	WGS	97%	5,00E−141	85.85%	-	-	-	Colombia	hpNEurope	hpEurope
MBGL00000000.5	NQ315	WGS	97%	5,00E−141	85.85%	-	-	-	Colombia	hpNEurope	hpEurope
MBGL00000000.6	2036	WGS	98%	2,00E−114	80.69%	-	-	-	Colombia	hpAfrica1	hpEurope
MBGL00000000.7	22371	WGS	98%	2,00E−114	80.69%	-	-	-	Colombia	hpAfrica1	Hybrid
MBGL00000000.8	22341	WGS	98%	2,00E−114	80.69%	-	-	-	Colombia	hpAfrica1	hspWAfrica
MBGL00000000.9	3133	WGS	95%	5,00E−148	87.38%	100%	3.00E−64	84.33%	Colombia	hpColombia	hpEurope
MVVX01000000	HP13024	WGS	95%	0.00E+00	96,19%	100%	1.00E−66	86.64%	Australia	hpSWEurope	hpEurope
MVUO010000000	HP15012	WGS	84%	2.00E−88	82.90%	100%	3.00E−67	86,64%	Australia	hpNEurope	hpEurope
MVYB010000000	HP04042	WGS	100%	0.00E+00	95.93%	100%	3.00E−67	86,64%	Australia	hpSWEurope	hpEurope
MVVZ010000000	HP13021	WGS	99%	0.00E+00	96.36%	98%	2.00E−64	89.12%	Australia	hpSWEurope	hpEurope
MVXE01000000	HP11032	WGS	97%	2.00E−175	92.07%	99%	1.00E−72	88.84%	Australia	hpNEurope	hpEurope
NYJJ010000000	HP16056	WGS	98%	6.00E−157	88.51%	100%	1.00E−66	86.64%	Australia	hpNEurope	hpEurope
MVTM01000000	HP16008	WGS	88%	2.00E−149	90.08%	100%	5.00E−48	78.80%	Australia	hpSWEurope	hpEurope
NYJI01000000	HP16062	WGS	88%	2.00E−149	90.08%	98%	5.00E−66	88.08%	Australia	hpSWEurope	hpEurope
MVTY01000000	HP15039	WGS	100%	0	96.83%	100%	4.00E−103	98.62%	Australia	hpSWEurope	hpEurope
QBQH01000000	HP30950	WGS	100%	7.00E−179	91.63%	96%	1.00E−45	80.48%	Belgium	hpSWEurope	hpEurope
PHLZ01000000	KH26	WGS	83%	2.00E−146	90.79%	92%	1.00E−65	88.50%	India	hpSWEurope	Hybrid
PHME01000000	KH20	WGS	88%	3.00E−130	87.53%	92%	8.00E−63	87.50%	India	hpSWEurope	hpEurope
PHMD01000000	KH21	WGS	87%	3.00E−131	87.80%	92%	8.00E−63	87.50%	India	hpSWEurope	hpEurope
PHLY01000000	KH27	WGS	97%	3.00E−141	86.25%	100%	4.00E−54	83.03%	India	hpNEurope	hpEurope

(Continues)

TABLE 2 (Continued)

Accession number Genbank	Strain	Blast Database	Integrase			Holin			Country	PST	MLST
			Query cov.	E-value	% Id.	Query cov.	E-value	% Id.			
PHMO01000000	KH10	WGS	97%	1.00E−152	87.96%	100%	1.00E−66	86.64%	India	hpNEurope	hpEurope
PHLK010000000	KH41	WGS	100%	1.00E−173	90.27%	92%	1.00E−65	88.50%	India	hpSWEurope	hpEurope
LIXF01000000	HP22	WGS	99%	2.00E−144	85.75%	100%	5.00E−58	83,41%	Kuwait	hpNEurope	hpEurope
MIKS010000000	MC2006-52	WGS	100%	0	93.21%	100%	3.00E−54	82.03%	Mexico	hpSWEurope	hpEurope
MILY01000000	MU2004-2	WGS	100%	0	92.99%	91%	2.00E−48	81.82%	Mexico	hpSWEurope	hpEurope
MILJ01000000	MGms15	WGS	100%	9.00E−178	91.40%	100%	4.00E−46	79.82%	Mexico	hpSWEurope	hpEurope
QDJR010000000	B373	WGS	100%	4.00E−176	91.18%	100%	4.00E−46	79.82%	Spain	hpSWEurope	hpEurope
QDJT010000000	B335	WGS	100%	9.00E−178	91.40%	100%	7.00E−49	80.18%	Spain	hpSWEurope	hpEurope
QEGU01000000	B497A	WGS	100%	2.00E−180	91.86%	96%	1.00E−45	80.48%	Spain	hpSWEurope	hpEurope
QEGM010000000	CRM21	WGS	100%	0	94.12%	92%	2.00E−50	82.50%	Spain	hpSWEurope	hspWAfrica
QDJO01000000	B491	WGS	100%	0	94.80%	92%	4.00E−52	83.00%	Spain	hpSWEurope	hpEurope
RJFW010000000	ZH50	WGS	97%	8.00E−124	83.22%	100%	1.00E−65	86.18%	Switzerland	hpNEurope	hpEurope
RJIN01000000	ZH128	WGS	97%	1.00E−134	84.95%	100%	6.00E−64	85.71%	Switzerland	hpNEurope	hpEurope
RJFJ010000000	ZH36	WGS	100%	5.00E−175	90.95%	96%	4.00E−39	77.51%	Switzerland	hpSWEurope	hpEurope
RJIH010000000	ZH122	WGS	100%	5.00E−175	90.95%	96%	1.00E−39	78.20%	Switzerland	hpSWEurope	hpEurope

Note: List of complete *Helicobacter pylori* genomes carrying prophage integrase and/or holin genes retrieved from public databases. The data include the MLST and PST.

NCBI=National Center for Biotechnology Information, WGS=Whole-Genome Shotgun, Query cov. = query coverage, % Id. = Identity percentage, MLST=multilocus sequence typing, PST=prophage sequence typing

population, and previous studies been described that this population has been replaced in Colombia.[25] According to PST analysis and genome alignment, the other fifteen phages were selected because they have negligible signs of introgression from other phage populations (Table S1). The MAUVE software was used for multiple alignment of the prophage genomes. Additionally, phiHP33 prophage and one (1) another prophage for each population (hpNEurope, hpSWEurope, hpEastAsia, and hpAfrica1) were selected to construct a comparison chart using BLASTn and Easyfig software.

2.3 | Data analysis of Prophage Sequence Typing (PST)

Integrase and holin gene sequences of Colombian strains, NCBI strains selected, and 95 previously characterized *H. pylori* sequences[6,8] (totaling 167 prophage sequences) were used to determine the population structure of prophages using PST.

Briefly, integrase and holin gene sequences were aligned using MAFFT (version 7).[26] The multi-fasta alignment file of these genes sequences was converted to the Structure 2.3.4[27-29] program input file using xmfa2struct by X. Didelot and D. Falush (http://www.xavierdidelot.xtreemhost.com/clonalframe.htm). The Structure program was used to study the number of K populations using the admixture model, performing duplicate runs. In each run, a Markov Chain Monte Carlo (MCMC) of 10,000 iterations and a burn-in period of 10,000 iterations were chosen. The highest mean value of ln likelihood was compared for multiple runs of $2 \leq K \leq 6$. To evaluate the PST analysis results generated and visualize likelihood values across multiple values of K, the Evanno method of the Structure Harvester[30] tool was used, namely to detect the number of populations that best fit the data.

Using concatenated integrase and holin gene nucleotide alignments, together or individually, a neighbor-joining statistical phylogenetic tree was constructed using the Kimura two-parameter[31] substitution model in the Molecular Evolutionary Genetic Analysis Software, Version 7.0 (MEGA 7).[32] The number of bootstrap replications was set at 1000. Another neighbor-joining statistical phylogenetic tree was constructed, following the same guidelines but using the genome of the prophages found in Colombian strains (from this study and databases), as well as the reference phages, PhiHp33, KHP30, India7 (accession number CP002331), Cuz20 (CP002076), 1961P (NC_019512.1), KHP40 (NC_019931.1), and 20 previously described prophages,[8] totaling 40 prophage genomes.

2.4 | *H. pylori* MLST data analysis

Seven *H. pylori* housekeeping genes (*atp* A, *efp*, *trp* C, *ppa*, *mut* Y, *yph* C, and *urel)* were analyzed in the eight (8) carried prophages strains obtained in this study, 45 database strains (16 Colombian and 29 from others countries) (Table 2), and 741 strains available at PubMLST (http://pubmlst.org/helicobacter/) initially described by Falush *et al.*[33] and Linz *et al.*[34] These strains were used to perform phylogenetic analysis using software Structure 2.3.4 and MEGA 7.0 software, as described above.

3 | RESULTS

3.1 | Identification of *H. pylori* prophages.

The presence of prophages in Colombian *H. pylori* strains was initially determined by detecting integrase and holin genes. Among the 213 *H. pylori* DNAs, 27 strains (12.6%) were positive for integrase, and 9 (4.2%) were positive for both genes. The prophage sequences obtained in the present study from Colombian strains are available at GenBank (No: MW160241 to MW160271). The 27 positive strains for the integrase gene came from 24 patients, and the nine positive strains for the holin gene came from 7 patients. The frequency of these genes, not in strains but patients, was 13.7% for integrase and 4% for holin. All prophage carriers' strains came from patients diagnosed with non-atrophic chronic gastritis.

For complete genome analysis, a prophage was considered intact if its size was at least 20 kb, as described by others.[5] We sequenced by NGS eight strains and found complete prophages in five strains (Col-5PUJ, Col-6PUJ, Col-7PUJ, Col-22PUJ, and Col-23PUJ) and incomplete prophages in three strains (Col-2PUJ, Col-16PUJ, and Col-17PUJ). (Table 1). Some of the analyzed strains (Col-6PUJ and Col-7PUJ, Col-22PUJ and Col-23PUJ, Col-16PUJ, and Col-17PUJ) had been isolated from the same patient and were identical (identity percentages greater than 99%). For these strains, it was found that the prophages were present in all the isolates from the same patient. The percentage of identity for the prophage sequences present in each patient was greater than 99%. Thus, one strain from each identical pair was withdrawn from subsequent analyzes. Genomes of strains included in the study were submitted to NCBI under the BioProject number PRJNA656306 and the Colombian *H. pylori* prophages described here are available under the access number MW247143 to MW247147.

Intact prophages had an average of 33 predicted genes, 28.6 kb in length, and 37% GC. BLASTn and Easyfig results showed synteny between previously described phages and those found in Colombian strains. Indeed, the multiple alignment results showed that Colombian phages have lengths and gene organization like those previously described (Figure 1). In the incomplete prophages, specific open reading frames were likely missed in the genome assembly, leading to less than 20 kb genomes. In all prophages, the missing part was the final part of the prophage.

The annotation and alignment procedures used allowed the identification of an insertion sequence (IS) in Col 23-PUJ phage. This IS encodes a transposase enzyme, was identified as ISCco1, corresponds to an IS606 (tnpA), and is a member of the IS200/IS605 family transposases.[35,36] BLASTp analysis revealed that this IS usually is found in different *H. pylori* strains and *H. pylori* prophages.[8]

For the Colombian *H. pylori* prophages, the insertion site (Table S4) could be determined only at the 5' end. In all prophages, at the 3' end, the contig ended with prophage sequences. For the Col 2-PUJ incomplete prophage, it was not possible to determine the insertion site at either end. The complete phages, Col 5-PUJ and Col 6-PUJ, were inserted contiguous to the gene encoding a Type II

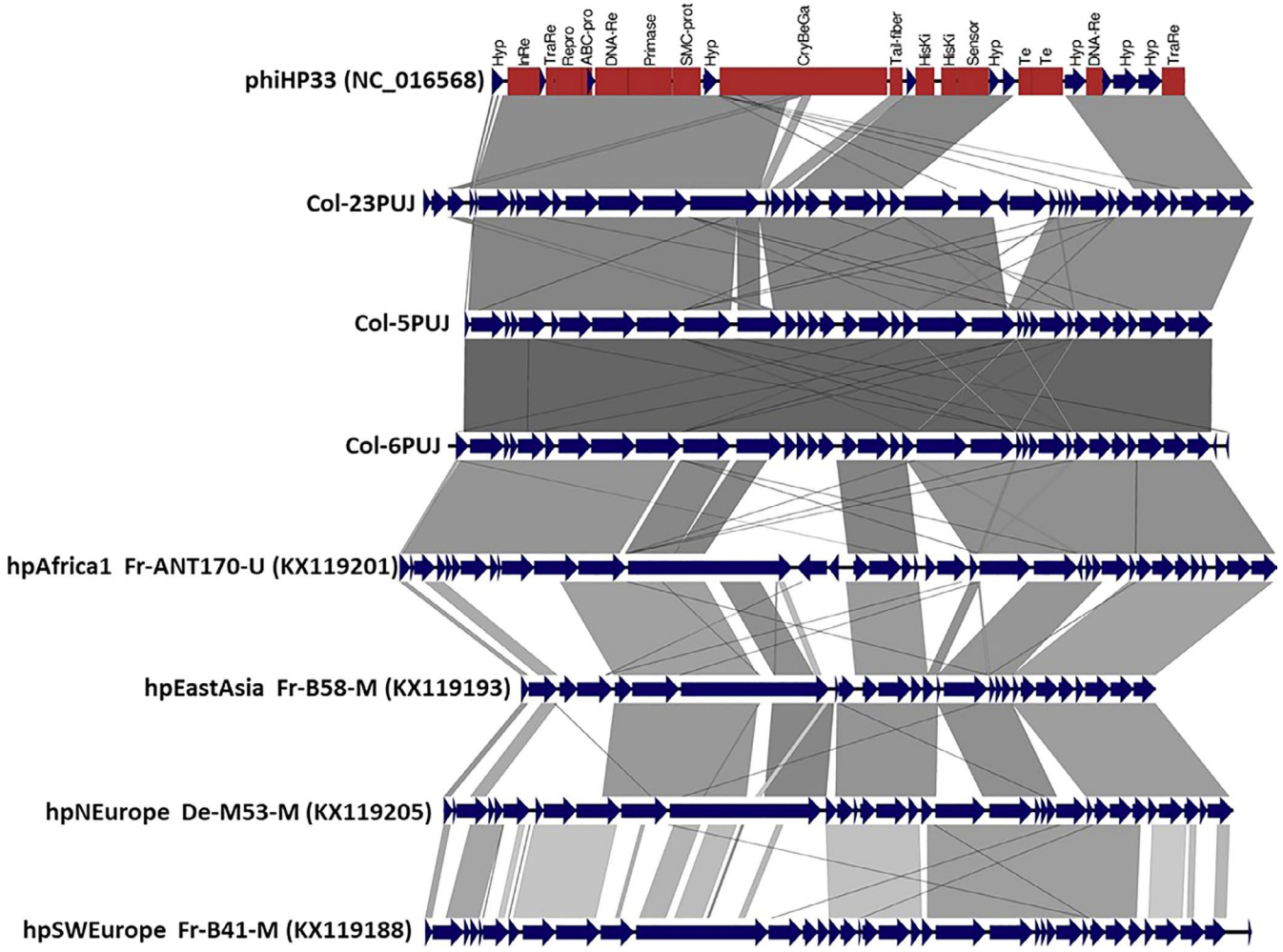

FIGURE 1 Synteny in genome organization of prophages from different populations. Using the BLASTn and Easyfig software, a phage comparison chart was constructed. The three complete Colombian prophages were used, and as a reference, phiHP33 and some phages (one for each population group) previously described by Vale et al. In 2017 were used. The chart shows synteny between the different phages and that the phages of Colombian origin conserve most of the soft-core genes.

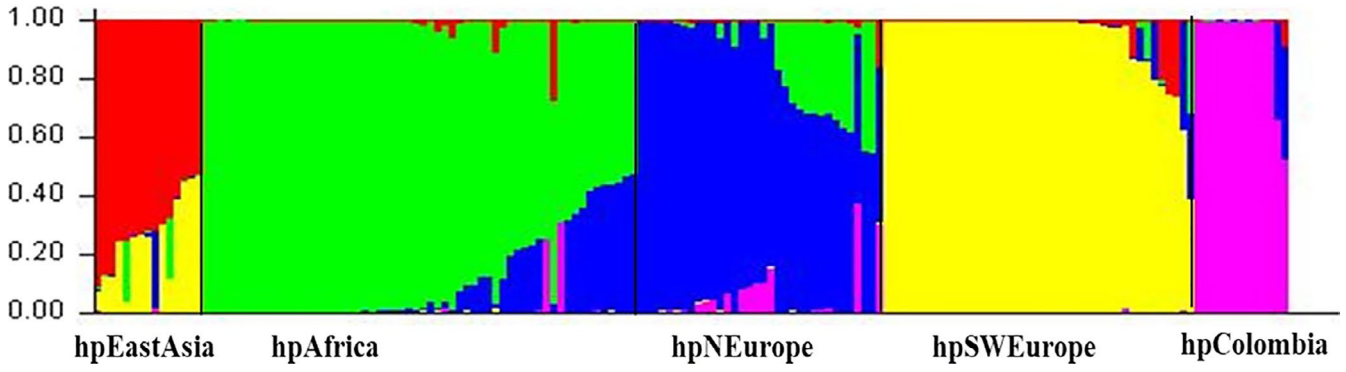

FIGURE 2 H. pylori prophage populations. Bayesian population assignments using Structure 2.3.4 software and the admixture model to analyze integrase and holin genes sequences. The plot shows five colored segments (K) representing the membership coefficients in each cluster. The fifth population was named hpColombia.

DNA modification enzyme, the Col 23-PUJ phage was inserted contiguous to the Holliday junction ATP-dependent DNA helicase RuvA gene, and the incomplete Col 16-PUJ phage was inserted adjacent to the membrane-associated phospholipid phosphatase gene. We determined the insertion site for the Colombian prophages obtained from the NCBI strains at the 5' end in five prophages. Four of them (HP3076, CM22388, NQ4044, and HP3133) were inserted next to the S-adenosylmethionine synthetase gene; and PZ5019-3A3 prophage was located next to the gene coding for Trigger factor. The insertion sites at the 3' end only were determined in two prophages, Hp3076, and NQ4044; in both prophages, the insertion site was next to the gene coding for UDP-3-O- [3-hydroxymyristoyl] glucosamine N-acyltransferase.

3.2 | *H. pylori* and prophage population structure

STRUCTURE 2.3.4 analysis for PST, surprisingly, revealed that K = 5 had the best probability value, indicating the presence of five populations. The five populations identified by PST are hpEastAsia, hpAfrica1, hpSWEurope, hpNEurope, and the fifth new population, containing only Colombian strains, was named hpColombia (Figure 2, Table S1). Evaluation of these results using the Structure Harvester tool and the Evanno method[30] confirmed that K = 5 was the best value (Table S2).

The neighbor-joining phylogenetic trees for concatenated integrase and holin genes and the prophage's complete genome also showed a new clustering with prophage sequences from Colombia strains, supporting the assignment of a new population (Figure 3A and 3B). In the tree with the complete genomes (Figure 3B), we observed that most of the prophages gathered by the phylogeographic group clustering accordingly to their population assigned by the PST analysis. However, some exceptions were found, namely for Pt-4472-G, Fr-G12-G, Fr-CG43-A, Pt-21299R-U, and Cuz20 (Figure 3B) that cluster with another prophage population.

MLST analysis, performed using Structure (Figure S1, Table S3), showed that K = 6 was the best value. The populations defined were consistent with the prior assignment of major populations, and these were used to classify the new strains included in this study. For the Colombian strains, MLST analysis revealed that most of

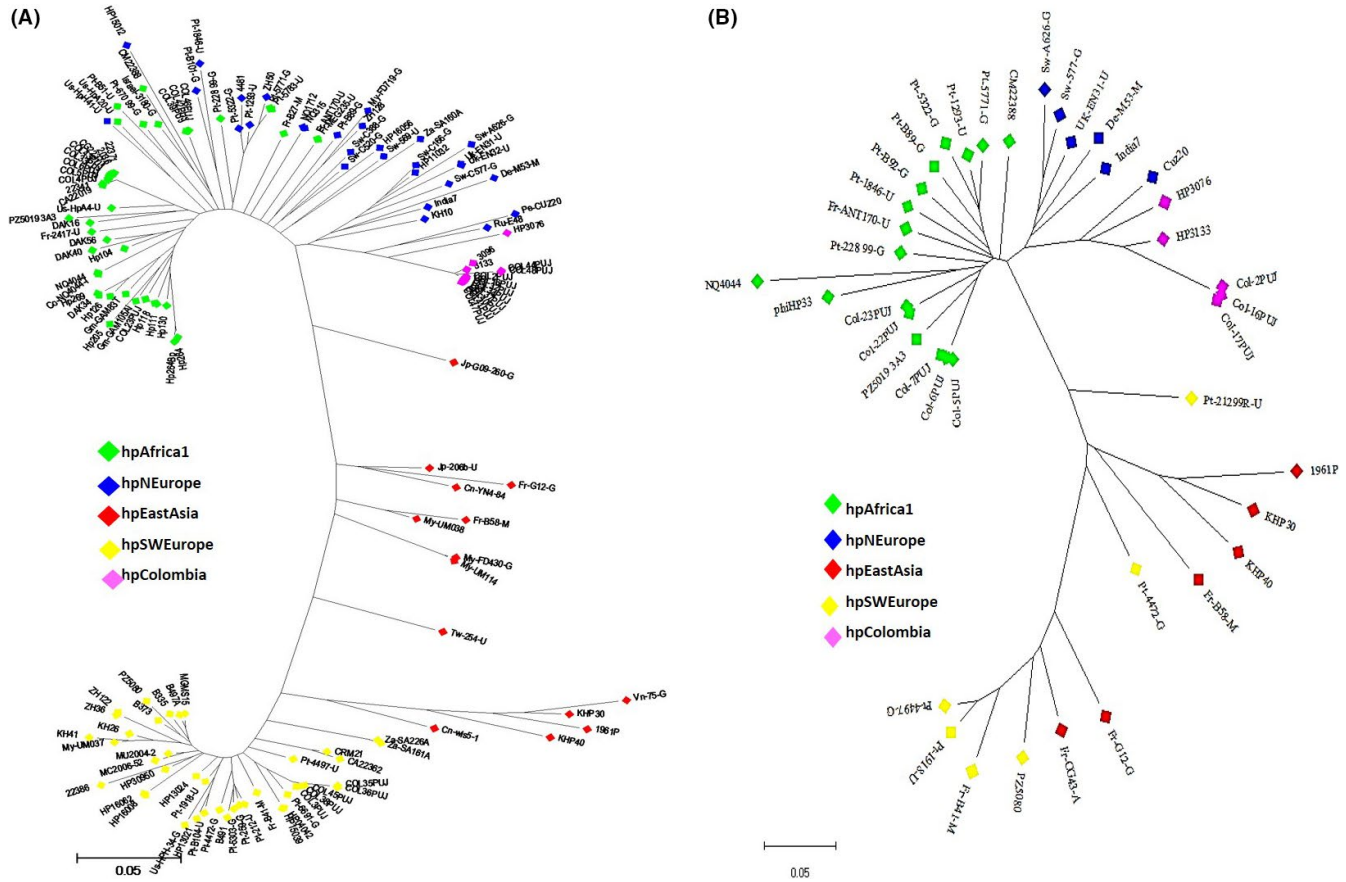

FIGURE 3 Phylogenetic distribution of the H. pylori prophages. A. Neighbor-joining phylogenetic tree topology (using the Kimura two-parameter model, running 1000 bootstrapping replications) of nucleotide alignments of the concatenated integrase and holin genes from 167 prophage sequences (27 Colombian sequences obtained in this study, 45 sequences from NCBI selected strains, and 95 sequences from previously characterized prophages H. pylori). B. A neighbor-joining phylogenetic tree topology (using the Kimura two-parameter model, running 1000 bootstrapping replications) of 40 prophages complete genomes. Eight Colombian prophages found in this study, six Colombian prophages recovered from databases, 20 previously described prophages, and six reference prophages (PhiHp33, KHP30, India7, Cuz20, 1961P, KHP40) were used to build this tree.

the Colombia strains belonged to the hpEurope population, while a few belonged to the HpAfrica1 population. The phylogenetic tree, constructed using the seven housekeeping genes (Figure S2), was consistent with the Structure findings. Taken together, these results support the presence of the six major *H. pylori* populations based on the assigned population available at PubMLST.

4 | DISCUSSION

Here, we have described the identification of prophages in Colombian *H. pylori* strains. While increasing numbers of studies report the presence of these prophages, this is the first description of these elements in isolates of American origin.

The frequency for integrase and holin genes was calculated on the strains and the patients in which these strains were obtained. However, even considering both scenarios, both genes' observed prevalence was not in line with those previously determined. Previous reports indicate that prophages are present in about 20% of *H. pylori* strains.[37] We found that only 12.6% of Colombian strains were positive for integrase, and only 4.2% were positive for both genes. This is similar to what we observed for *H. pylori* Colombian genomes present in public databases. Thus, of the 186 Patric database[38] Colombian strains analyzed, only 17 (9.1%) were positive for the integrase gene, and 6 (3.2%) for both genes. These results suggest a decreased prevalence of prophage sequences in American isolates, which may have been due to a bottleneck of *H. pylori* genomes poor in prophage sequences that arrived in America from Europe and Africa after the continent discovery (discussed below). Alternatively, the *H. pylori* strains that adapt better to the Colombian population may have been those without prophages.

The prophages identified in Colombian *H. pylori* strains are consistent with the sizes and characteristics estimated for the previously reported *H. pylori* phages used in this study for comparison, pointing to a common origin. Indeed, the Colombian prophages showed genetic synteny among themselves and with phiHP33 phage.[5] As described for other *H. pylori* phages, the prophages ORFs described here were mostly in the same direction, many ORFs were annotated as hypothetical proteins or with no known function, and no virulence genes were identified. All prophage soft-core genes[37] were present in Colombian prophages, though the holin gene presented a smaller percent identity (72%). Taken together, these results support that the Colombia prophages have the same origin as the rest of the *H. pylori* prophages.

In this study, all *H. pylori* strains carriers of prophages came from patients diagnosed with non-atrophic chronic gastritis. Therefore, we cannot infer that prophages could be related to specific disease patterns. However, previous reports[5,7] suggest that there is no direct association between the presence of prophages and the generation of disease. Despite this, prophages can play critical roles in bacterial populations. Likewise, many prophages contain genes that contribute to bacterial fitness and colonization capacity in a particular niche.[2] Genes present in prophages increase bacterial fitness through a wide range of mechanisms and play an essential role in bacterial diseases.[39] In the case of *H. pylori*, phage orthologous genes have been found in the genomes of strains that also carry *cagA* and/or *vacA* virulence genes.[40] Studies have suggested that *H. pylori* prophages could represent an essential element for *H. pylori* adaptation to a hostile environment, such as the human stomach.[8,9] However, additional studies are required to elucidate the specific functions of prophages found in *H. pylori* and these prophages role in establishing *H. pylori*-related disease.[8]

A previous report about *H. pylori* prophages revealed that IS is present in around 39% of prophages.[8] In the Colombian isolates, we identified an IS in only one prophage (1/11, 9% of the Colombian prophages) (prophage Col 23-PUJ). Previously, IS606 was also identified at the 3' end of a prophage remnant (Sw-C388-G), belonging to the hpNEurope prophage population.[8]

A relationship between the prophage's insertion site and the phylogenetic classification by PST had previously been described.[8] In this study, we found that the Colombian prophages' insertion sites were mostly related to those that had been described for the hpNEurope and hpAfrica1 populations, which is consistent with the populations described here. However, in three prophages, novel insertion sites at the 5' end are described. In fact, the Col 5-PUJ and Col 6-PUJ prophages (from hpAfrica1 population) were inserted next to the Type II DNA modification enzyme gene. Curiously, the enzyme product corresponds to a restriction endonuclease, a component of the restriction-modification system that protects the bacteria against the invasion of foreign DNA.[41] The Col 23-PUJ prophage (from hpAfrica1 population) was inserted next to the Holliday junction ATP-dependent DNA helicase RuvA gene, which codes for a hexameric protein that is part of the homologous recombination machinery in bacteria[42]; finally, the prophage present in the strain PZ5019 (available from NCBI) (from hpAfrica1 population) was inserted next to the gene coding for a Trigger factor for which there is not much information. In these last three prophages, the insertion site could correspond to random locations. Gene expression studies would be needed to evaluate the impact of the prophage insertion site on the bacteria.

There are seven well established different *H. pylori* populations: hpEurope, hpNEAfrica, hpAfrica1, hpAfrica2, hpAsia2, hpSahul, and hpEastAsia, reflecting *H. pylori* and man have co-evolved together.[33,34,43,44] Particularly, the hpEastAsia population is divided into three subpopulations, hspEAsia, hspMaori, and hspAmerind.[33,34,43-46] The hspAmerind subpopulation reflects the migration from Asia to the Americas across the Bering Strait that began 12,000 years ago.[47] However, this population group was replaced in the American mestizo population due to the entry of new pathogens, including *H. pylori*, 500 years ago, during the Spanish colonization migratory process and the arrival of African slaves.[48] Accordingly, of the strains analyzed in this study by MLST, none belonged to the hpAmerind population. Indeed, our MLST results showed that the analyzed strains mainly belong to hpEurope and, to a lesser extent, hpAfrica populations (Table S3). These results are similar to those of the first studies of the Colombian strains and agree with the expected results based on migration involved in Spanish colonization.[48,49] It was recently

demonstrated by studies that used the whole-genome sequences, the existence of a specific population component in strains isolated from the Colombian mestizo population,[19-21,25] suggesting that the *H. pylori* Colombian strains have followed unique evolutionary pathways[19] and have formed new subpopulations from a European source,[21] localized at the Iberian Peninsula. Thus, this Colombian subpopulation named hspSWEuropeColombia (previously reported as hspSEuropeColombia) showed hspSWEurope predominance.[20] Interestingly and opposing, for a new cluster of Colombia prophages (presented below), the closer population has hpNEurope predominance and not hpSWEurope, supporting that hpColombia diverges from hpNEurope population, whose prevalence is residual. Our findings agree with the observation that a new subpopulation of European origin is evolving in Colombia and confirm that most Colombian strains are definitely of European origin.

Using the technique known as PST, Vale et al. identified four prophage populations, differentiating two European populations, mainly present in north Europe (hpNEurope) and the southwest of Europe (hpSWEurope).[7] Recently, using the entire *H. pylori* genome, two subpopulations of hpEurope were also observed.[21] Our PST analysis revealed five phylogenetic groups, one more than previously reported, confirming the high-resolution level of PST analysis in discriminating *H. pylori* prophage populations. Considering the Colombian prophages, 25.0% (10 prophages) of were assigned to hpSWEurope population, 5.0% (2 prophages) to hpNEurope, 42.5% (17 prophages) to hpAfrica1, and 27.5% (11 prophages) to a new population. The fifth new population was composed only of prophages with Colombian origin for which it was named hpColombia. None of the Colombian prophages were classified in the hpEastAsia population. The diversity of the populations of the Colombian prophages reflects the complex colonization history of Colombia. Possible recombination events and introgression between strains may have led to the generation of a novel lineage. The hybrid nature of the Colombian prophage populations is also evident in another strain from Latin America, Cuz20, which corresponds to an isolate from Peru (discussed below). These findings should be related to the recent history of human migration waves to Colombia, by Europeans and Africans.

The phylogenetic trees place hpColombia population closer to the hpNEurope population. As referred above, this new population may have diverged from hpNEurope prophages arriving in the Americas due to the close phylogenetic clustering (Figure 3A and 3B) between the Colombian prophages and those belonging to the hpNEurope population. The identification of a new population of prophages from the Americas is in agreement with other studies pointing to the existence of *H. pylori* subpopulations in Colombia.[20,21] We cannot affirm that the new subpopulation of prophages (hpColombia) described here is associated in the new subpopulation of Colombian strains, previously described. The same strains were not analyzed. In previous studies, no prophage carriers' strains were included. However, the findings demonstrate that in Colombia, there has been genomic variation in the *H. pylori* subpopulations and their prophages, which show the adaptation and admixture between strains originating from the Americas and Europe.

The prophage insertion sites of the hpColombia population also match the ones most frequently found for the hpNEurope populations, reinforcing that these two prophages populations are closely related. The genome phylogenetic tree (Figure 3B) showed that few prophages displayed discrepant phylogeographic segregation from their PST classification, suggesting the existence of putative recombination events, as previously reported.[8] This finding suggests that these prophages have a genomic mosaic composed of both populations. For instance, the Cuz20 prophage (genome available at the NCBI) was assigned to hpNEurope population, but the PST result indicated some introgression from hpColombia and hpEastAsia. (Table S1). In agreement, the Cuz20 prophage clusters together with hpColombia prophages in the phylogenetic tree (Figure 3B), suggesting an introgression from the novel hpColombia population. This strain was isolated from Peru, which supports the theory about the American strains' hybrid nature.

Our results highlight the relationship between *H. pylori* genetic diversity and tracking of human migrations and verify the *H. pylori* prophage coevolution, leading to the development of a novel prophage lineage in the new world.

5 | CONCLUSIONS

Our results showed the existence of a new population of *H. pylori* prophages, hpColombia, corresponding to a population composed exclusively of strains of Colombian origin. Our results verify that PST can differentiate individual populations and that prophages affect *H. pylori* genetic diversity, which is strongly related to geography. Performing additional studies with a higher number of Colombian strains and other strains from the Americas using the entire genome could allow for in-depth characterization of the prophage population structure at a more exceptional level.

ACKNOWLEDGMENTS

We thank the entities that financially supported the development of this work: AM is a recipient of a scholarship from the Centro de Estudios Interdisciplinarios Básicos y Aplicados –CEIBA Foundation, Colombia– AM and AT are recipients of a project grant (120380763025/2018) from the Departamento Administrativo de Ciencia, Tecnología e Innovación de Colombia. Colciencias. The work is partially supported by Research Vice-rectory. Pontificia Universidad Javeriana (PPTA_7676) and F.P.I.T. BanRepCultural (Project 3.956). FFV is financed by National funds from the Fundação para a Ciência e a Tecnologia (FCT) through an Assistant Researcher grant CEECIND/03023/2017, a project grant (PTDC/BTM-SAL/28978/2017), and projects UIDB/04138/2020 and UIDP/04138/2020; this founds partially supported this work.

DATA AVAILABILITY STATEMENT

The genomes of Colombian strains included in the study are available in the NCBI (https://www.ncbi.nlm.nih.gov/) with the BioProject number PRJNA656306. The Colombian *H. pylori*

prophages described here are available under the access number MW247143 to MW247147. The access numbers for the strains collected from NCBI included in this study are available in the Table 2 of this article.

ORCID

Angela B Muñoz https://orcid.org/0000-0002-7451-3189
Alba A Trespalacios-Rangel https://orcid.org/0000-0002-7540-5868
Filipa F Vale https://orcid.org/0000-0003-4635-0105

REFERENCES

1. Harrison E, Brockhurst MA. Ecological and evolutionary benefits of temperate phage: what does or doesn't kill you makes you stronger. *Bio Essays*. 2017;39(12).
2. Brüssow H, Canchaya C, Hardt WD. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev*. 2004;68(3):560-602.
3. Paterson S, Vogwill T, Buckling A, et al. Antagonistic coevolution accelerates molecular evolution. *Nature*. 2010;464(7286):275-278.
4. Luo CH, Chiou PY, Yang CY, Lin NT. Genome, integration, and transduction of a novel temperate phage of *Helicobacter pylori*. *J Virol*. 2012;86(16):8781-8792.
5. Lehours P, Vale FF, Bjursell MK, et al. Genome sequencing reveals a phage in *Helicobacter pylori*. *MBio*. 2011;2(6).
6. Secka O, Vale FF, Buissonnière A, Thomas JE, Mégraud F, Lehours P. Phylogeographic agreement between prophage and bacterial housekeeping genes in *Helicobacter pylori* strains from The Gambia. *Helicobacter*. 2017;22(5).
7. Vale FF, Vadivelu J, Oleastro M, et al. Dormant phages of *Helicobacter pylori* reveal distinct populations in Europe. *Sci Rep*. 2015;5:14333.
8. Vale FF, Nunes A, Oleastro M, et al. Genomic structure and insertion sites of *Helicobacter pylori* prophages from various geographical origins. *Sci Rep*. 2017;7:42471.
9. Vale FF, Lehours P. Relating Phage Genomes to *Helicobacter pylori* Population Structure: General Steps Using Whole-Genome Sequencing Data. *Int J Mol Sci*. 2018;19(7).
10. Uchiyama J, Takeuchi H, Kato S-I, et al. Complete genome sequences of two *Helicobacter pylori* bacteriophages isolated from Japanese patients. *J Virol*. 2012;86(20):11400-11401.
11. Uchiyama J, Takeuchi H, Kato S-I, et al. Characterization of *Helicobacter pylori* bacteriophage KHP30. *Appl Environ Microbiol*. 2013;79(10):3176-3184.
12. Uchiyama J, Takemura-Uchiyama I, Kato S, et al. Screening of KHP30-like prophages among Japanese *Helicobacter pylori* strains, and genetic analysis of a defective KHP30-like prophage sequence integrated in the genome of the H. pylori strain NY40. *FEMS Microbiol Lett*. 2016;363(16).
13. Muñoz AB, Stepanian J, Trespalacios AA, Vale FF. Bacteriophages of *Helicobacter pylori*. *Front. Microbiol*. 2020;11:549084.
14. Mégraud F, Lehours P, Vale FF. The history of *Helicobacter pylori*: from phylogeography to paleomicrobiology. *Clin Microbiol Infect*. 2016;22(11):922-927.
15. Hooi JKY, Lai WY, Ng WK, et al. Global prevalence of helicobacter pylori infection. Systematic review and meta-analysis. *Gastroenterology*. 2017;153(2):420-429.
16. Bravo LE, Cortés A, Carrascal E, et al. *Helicobacter pylori*: patología y prevalencia en biopsias gástricas en Colombia. *Colombia Médica*. 2003;34(3).
17. Camilo V, Sugiyama T, Touati E. Pathogenesis of *Helicobacter pylori* infection. *Helicobacter*. 2017;22(Suppl):1.
18. Thorell K, Lehours P, Vale FF. Genomics of *Helicobacter pylori*. *Helicobacter*. 2017;22(Suppl):1.
19. Muñoz-Ramírez ZY, Mendez-Tenorio A, Kato I, 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).
20. Muñoz-Ramirez ZY, Pascoe B, Mendez-Tenorio A, et al. A 500-year tale of co-evolution, adaptation, and virulence: *Helicobacter pylori* in the Americas. *ISME J*. 2020.
21. Thorell K, Yahara K, Berthenet E, et al. Rapid evolution of distinct *Helicobacter pylori* subpopulations in the Americas. *PLoS Genet*. 2017;13(2):e1006546.
22. Trespalacios AA, Rimbara E, Otero W, Reddy R, Graham DY. Improved allele-specific PCR assays for detection of clarithromycin and fluoroquinolone resistant of *Helicobacter pylori* in gastric biopsies: identification of N87I mutation in GyrA. *Diagn Microbiol Infect Dis*. 2015;81(4):251-255.
23. Trespalacios AA, Otero W, Caminos JE, et al. Phenotypic and genotypic analysis of clarithromycin-resistant *Helicobacter pylori* from Bogotá D.C., Colombia. *J Microbiol*. 2013;51(4):448-452.
24. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455-477.
25. Gutiérrez-Escobar AJ, Trujillo E, Acevedo O, Bravo MM. Phylogenomics of Colombian. *Gut Pathog*. 2017;9:52.
26. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772-780.
27. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*. 2003;164(4):1567-1587.
28. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes*. 2007;7(4):574-578.
29. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155(2):945-959.
30. Earl DA, vonHoldt BM.STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method.
31. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*. 1980;16(2):111-120.
32. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33(7):1870-1874.
33. Falush D, Wirth T, Linz B, et al. Traces of human migrations in *Helicobacter pylori* populations. *Science*. 2003;299(5612):1582-1585.
34. Linz B, Balloux F, Moodley Y, et al. An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature*. 2007;445(7130):915-918.
35. He S, Corneloup A, Guynet C, et al. The IS200/IS605 family and "peel and paste". Single-strand transposition mechanism. *Microbiol Spectr*. 2015;3(4).
36. Kersulyte D, Akopyants NS, Clifton SW, Roe BA, Berg DE. Novel sequence organization and insertion specificity of IS605 and IS606: chimaeric transposable elements of *Helicobacter pylori*. *Gene*. 1998;223(1–2):175-186.
37. Yahara K, Lehours P, Vale FF. Analysis of genetic recombination and the pan-genome of a highly recombinogenic bacteriophage species. *Microb Genom*. 2019;5(8):e000282.
38. Wattam AR, Abraham D, Dalay O, et al. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res*. 2014;42(Database issue):D581-D591.

39. Taylor VL, Fitzpatrick AD, Islam Z, Maxwell KL. The diverse impacts of phage morons on bacterial fitness and virulence. *Adv Virus Res*. 2019;103:1-31.
40. Kyrillos A, Arora G, Murray B, Rosenwald AG. The presence of phage orthologous genes in *Helicobacter pylori* correlates with the presence of the virulence factors CagA and VacA. *Helicobacter*. 2016;21(3):226-233.
41. Pingoud A, Fuxreiter M, Pingoud V, Wende W. Type II restriction endonucleases: structure and mechanism. *Cell Mol Life Sci*. 2005;62(6):685-707.
42. Chen YJ, Yu X, Egelman EH. The hexameric ring structure of the Escherichia coli RuvB branch migration protein. *J Mol Biol*. 2002;319(3):587-591.
43. Moodley Y, Linz B, Yamaoka Y, et al. The peopling of the Pacific from a bacterial perspective. *Science*. 2009;323(5913):527-530.
44. Achtman M, Azuma T, Berg DE, et al. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol Microbiol*. 1999;32(3):459-470.
45. Yamaoka Y. *Helicobacter pylori* typing as a tool for tracking human migration. *Clin Microbiol Infect*. 2009;15(9):829-834.
46. Devi SM, Ahmed I, Francalacci P, et al. Ancestral European roots of *Helicobacter pylori* in India. *BMC Genom*. 2007;8:184.
47. Kersulyte D, Kalia A, Gilman RH, 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(11):e15076.
48. de Sablet T, Piazuelo MB, Shaffer CL, et al. Phylogeographic origin of *Helicobacter pylori* is a determinant of gastric cancer risk. *Gut*. 2011;60(9):1189-1195.
49. Shiota S, Suzuki R, Matsuo Y, et al. *Helicobacter pylori* from gastric cancer and duodenal ulcer show same phylogeographic origin in the Andean region in Colombia. *PLoS One*. 2014;9(8):e105392.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Muñoz AB, Trespalacios-Rangel AA, Vale FF. An American lineage of *Helicobacter pylori* prophages found in Colombia. *Helicobacter*. 2021;00:e12779. https://doi.org/10.1111/hel.12779