AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

| Bacteriology | Announcement

Complete genome sequence of *Helicobacter pylori* isolated from residents in southwestern Colombia using Oxford Nanopore sequencing technology

Lizeth Mejia,[1,2] Julie Benavides-Melo,[3] Ernesto Argoty,[4] Liliana Montenegro,[5] Nelson Rivera-Franco,[1,6] Diana López-Alvarez,[6,7] Álvaro Pazos[2,8]

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT Genome sequences of *Helicobacter pylori* strains isolated from patients in Nariño, Colombia, with gastric lesions were assembled using Nanopore sequencing. Plasmids were detected in some strains and were predicted to be mobilizable, with relaxases of the MOBP type. *H. pylori*'s virulence genes may explain the link with gastric cancer.

KEYWORDS whole-genome sequencing, *Helicobacter pylori*, precancerous gastric lesions

Helicobacter pylori, a pathogen colonizing the gastric mucosa of over half the global population (1), induces gastric inflammation, though most infected individuals remain asymptomatic. However, a significant percentage develop gastric or duodenal ulcers (10%–15%) or gastric cancer (1%–3%) (2).

This study analyzed gastric biopsies from the residents of high- and low-risk gastric cancer regions in Nariño, Colombia. Histopathological diagnosis and culturing of antral biopsies were performed according to established protocols (3). Histopathology revealed non-atrophic gastritis (six patients), atrophic gastritis with intestinal metaplasia (one patient), and diffuse gastric cancer (one patient) (Table 1). *H. pylori* strains were identified by Gram staining and urease, oxidase, and catalase tests. The gastric biopsy sample was cultured on Columbia agar (Oxoid, UK) supplemented with 10% sheep defibrinated blood, selective supplement Dent (Oxoid, UK), and 1% enrichment supplement Isovitalex (Oxoid, UK), under 10% CO_2 at 37°C for 7–10 days, then cryopreserved in NUNC tubes with 80% sterile thioglycolate and 20% glycerol at −80°C. For DNA extraction, bacterial cultures were grown under the same conditions, but without Dent, using the UltraClean Blood DNA Isolation Kit (MOBIO) and quantified with a Qubit version 3.0 fluorometer (Thermo Fisher Scientific).

MinION sequencing libraries were prepared with ligation sequencing kit (SQK-LSK109) and native barcoding expansion kit (SQK-NBD104) (Oxford Nanopore Technologies, ONT). No DNA shearing or size selection was performed before library preparation. Sequencing was performed on a flow cell (R9.4.1) for 72 h on the MinION (ONT). Default parameters were used for all software unless otherwise specified. Fast5 reads were converted to fastq with Guppy version 6.5.7 in high-accuracy mode (4). Raws fastq were trimmed by cutadapt version 4.9 (5) according to the needs of each sample. The reads were assembled using Flye version 2.9.4 (6) and polished with Pilon version 1.24 (7) (--fix "bases," "gaps"), Medaka version 2.0.1 (ONT, 2018), and homopolish version 0.4.1 (8). Genome completeness was assessed with BUSCO version 5.5.0 (9) and quality with checkM version 1.2.2 (10). Circularity was determined using Flye version 2.9.4 (6), and coverage depth with Qualimap version 2.2.2 (11). Gene predictions and annotations were

Editor David Rasko, University of Maryland School of Medicine, Baltimore, Maryland, USA

Address correspondence to Diana López-Alvarez, dilopezal@unal.edu.co.

The authors declare no conflict of interest.

See the funding table on p. 4.

Received 5 February 2025
Accepted 14 April 2025
Published 20 May 2025

TABLE 1 Assembly genomic statistics and gene predictions and annotations of Colombian *H. pylori* strains using Nanopore sequencing[a]

Strain ID	GenBank accessions	Diagnosis	Age	Sex	Host residence	SRA accession	Raw reads	Reads N_{50}	Genome size (bp)	Mean coverage (×)	No. of contigs	Contig N_{50} (bp)	CDS (total)	rRNAs	tRNAs	ncRNAs	Lineage
CR12	CP175941, CP175942	NAG	33	F	Florida	SRR26717568	809,317	536	1,706,154	193.6	2	1,700,838	1,601	6	36	3	hspAfrica1LatinAmerica
CR41	JBLDXM000000000	NAG	30	F	Barbacoas	SRR26717567	225,971	477	1,612,862	49.9	31	108,384	1,596	4	36	3	hspAfrica1LatinAmerica
CR44	CP175940	AGIM	36	M	Pasto	SRR26717566	1,192,402	2,111	1,681,553	787.4	1	1,681,553	1,586	6	36	3	hspSWEuropeLatinAmerica
CR45	JBLDXL000000000	NAG	34	M	Pasto	SRR26717565	1,635,951	453	1,734,444	345.7	70	44,824	1,683	6	38	3	hspSWEuropeLatinAmerica
CR46	CP175939	NAG	45	M	Samaniego	SRR26717564	1,842,573	840	1,623,746	653.4	1	1,623,746	1,528	6	36	3	hspSWEurope
CR56	JBLDXK000000000	DGC	24	M	Pasto	SRR26717563	282,596	600	1,712,951	78.8	36	111,822	1,669	4	37	3	hspSWEuropeLatinAmerica
CR60	CP175938	NAG	47	F	Pasto	SRR26717562	3,332,538	833	1,672,799	1,129.5	1	1,672,799	1,561	6	36	3	hspSWEuropeLatinAmerica
CR71	JBLDXJ000000000	NAG	34	F	Pasto	SRR26717561	803,778	403	1,683,646	151.1	31	132,000	1.652	5	36	3	hspSWEuropeLatinAmerica

[a]F, female; M, masculine; NAG, non-atrophic gastritis; AGIM, atrophic gastritis with intestinal metaplasia; and DGC, diffuse gastric cancer.

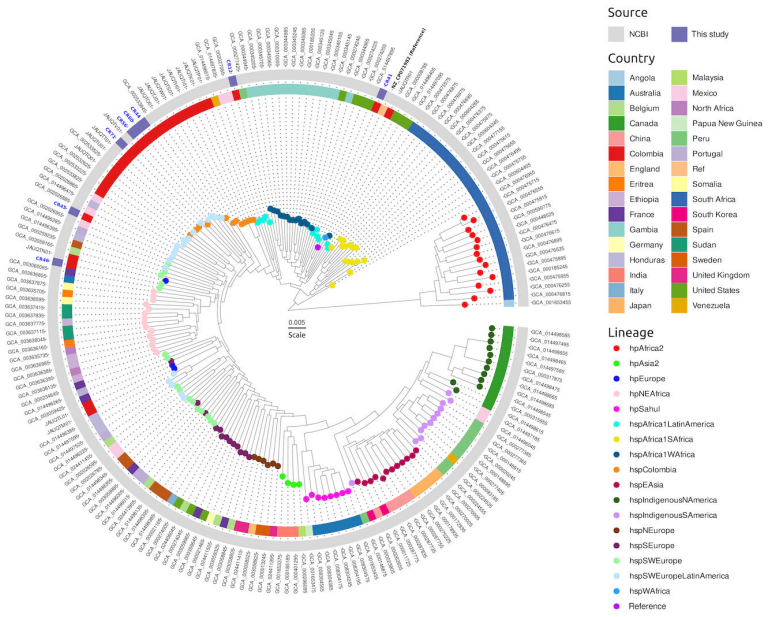

FIG 1 Maximum likelihood tree of the full database (192 sequences) of genomes of *H. pylori*. Color annotations are given in the circles around the terminal nodes, indicating the country of origin for the *H. pylori* isolates included in the tree. The color of the tips represents the lineage from which each genome was reported. The gray area indicates NCBI genomes, and the blue one the genomes obtained in this study.

performed using PGAP (12). The assembled genomes showed an average GC content of 38%, a size of 1.7 Mbp, and 1,593 protein-coding genes (Table 1).

Four plasmids were detected using MOB-Suite version 3.1.0 (13) in strains CR12 (CP175942), CR41 (JBLDXM010000005 and JBLDXM010000021), and CR71 (JBLDXJ010000026). Plasmids ranged in size from 4,260 to 9,178 bp, with three plasmids classified as mobilizable and containing MOBP-type relaxases (AB508, AE952, and novel plasmid from CR71).

Phylogenetic analysis, based on SNP alignments generated with RealPhy version 1.12, was performed with IQ-TREE version 2.2.6 (14) and visualized with ggtree version 3.12.0 (15) and ggtreeExtra version 1.14.0 (16) packages. A total of 184 *H. pylori* genomes from NCBI (PRJNA529500) were included, representing previously defined genetic populations and subpopulations (17, 18). The analysis identified four genomes related to hspSWEuropeLatinAmerica, two to hspAfrica1LatinAmerica, one to hspSWEurope, and one to hspColombia (Fig. 1). Raw FASTQ and assembled genomes are available in NCBI under BioProject PRJNA1037030 (Table 1).

ACKNOWLEDGMENTS

This research was supported under The Neuroinfections Emerging in the Americas Study funded by the U.S. National Institutes of Health (R01-NS110122) and Universidad del Valle (project "Neurovirus Emergentes y Enfermedades Neuroinflamatorias," C.I 1885).

The funders had no role in study design, data collection, analysis, publication decision, or manuscript preparation.

Thanks to Dr. Sandra Cifuentes from Centro Médico la Riviera, Pasto, Colombia, and Grupo Salud Pública, Universidad de Nariño, for obtaining gastric isolates.

AUTHOR AFFILIATIONS

[1]Laboratorio de Técnicas y Análisis Ómicos-TAOLab/CiBioFi, Facultad de Ciencias Naturales y Exactas, Universidad del Valle, Cali, Valle del Cauca, Colombia

[2]Departamento de Biología, Facultad de Ciencias Exactas y Naturales, Universidad de Nariño, Pasto, Nariño, Colombia
[3]Grupo GIISE, Facultad de Medicina, Universidad Cooperativa de Colombia, Pasto, Nariño, Colombia
[4]Alcaldía de Pasto, Secretaria de Salud–Salud Pública, Pasto, Nariño, Colombia
[5]Grupo HOSDERNAR, Hospital Universitario Departamental de Nariño, Pasto, Nariño, Colombia
[6]Grupo VIREM - Virus Emergentes y Enfermedad, Escuela de Ciencias Básicas, Facultad de Salud, Universidad del Valle, Cali, Valle del Cauca, Colombia
[7]Departamento de Ciencias Biológicas, Facultad de Ciencias Agropecuarias, Universidad Nacional de Colombia, Palmira, Valle del Cauca, Colombia
[8]Grupo de investigación Salud Pública, Centro de Estudios en Salud, Universidad de Nariño (CESUN), Pasto, Nariño, Colombia

AUTHOR ORCIDs

Diana López-Alvarez http://orcid.org/0000-0002-7734-8481

FUNDING

Funder	Grant(s)	Author(s)
National Institutes of Health	R01-NS110122	Diana López-Alvarez
Universidad del Valle	C.I 1885	Nelson Rivera-Franco

AUTHOR CONTRIBUTIONS

Lizeth Mejia, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Validation | Julie Benavides-Melo, Conceptualization, Investigation, Methodology, Writing – review and editing | Ernesto Argoty, Conceptualization, Investigation, Methodology, Writing – review and editing | Liliana Montenegro, Conceptualization, Investigation, Methodology, Writing – review and editing | Nelson Rivera-Franco, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – review and editing | Diana López-Alvarez, Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Writing – review and editing | Álvaro Pazos, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Writing – review and editing

ETHICS APPROVAL

This study was approved by the Ethics Committee of Universidad de Nariño (act 58, 10 April 2017). The study included eight participants (Table 1) who signed informed consent.

REFERENCES

1. Ramis IB. 2017. Molecular methods for detection of *Helicobacter pylori* infection: could they be the gold standard? J Bras Patol Med Lab 53:4. https://doi.org/10.5935/1676-2444.20170005
2. Thorell K, Bengtsson-Palme J, Liu OH-F, Palacios Gonzales RV, Nookaew I, Rabeneck L, Paszat L, Graham DY, Nielsen J, Lundin SB, Sjöling Å. 2017. *In vivo* analysis of the viable microbiota and *Helicobacter pylori* transcriptome in gastric infection and early stages of carcinogenesis. Infect Immun 85:e00031-17. https://doi.org/10.1128/IAI.00031-17
3. Kodaman N, Pazos A, Schneider BG, Piazuelo MB, Mera R, Sobota RS, Sicinschi LA, Shaffer CL, Romero-Gallo J, de Sablet T, Harder RH, Bravo LE, Peek RM Jr, Wilson KT, Cover TL, Williams SM, Correa P. 2014. Human and *Helicobacter pylori* coevolution shapes the risk of gastric disease. Proc Natl Acad Sci USA 111:1455–1460. https://doi.org/10.1073/pnas.1318093111
4. Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol 20:129. https://doi.org/10.1186/s13059-019-1727-y
5. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10. https://doi.org/10.14806/ej.17.1.200
6. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8
7. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963

8. Huang Y-T, Liu P-Y, Shih P-W. 2021. Homopolish: a method for the removal of systematic errors in nanopore sequencing by homologous polishing. Genome Biol 22:95. https://doi.org/10.1186/s13059-021-02282-6
9. Manni M, Berkeley MR, Seppey M, Zdobnov EM. 2021. BUSCO: assessing genomic data quality and beyond. Curr Protoc 1:e323. https://doi.org/10.1002/cpz1.323
10. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114
11. García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, Dopazo J, Meyer TF, Conesa A. 2012. Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics 28:2678–2679. https://doi.org/10.1093/bioinformatics/bts503
12. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569
13. Robertson J, Nash JHE. 2018. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. Microb Genom 4:e000206. https://doi.org/10.1099/mgen.0.000206
14. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300
15. Yu G, Smith DK, Zhu H, Guan Y, Lam T-Y. 2017. Ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol 8:28–36. https://doi.org/10.1111/2041-210X.12628
16. Xu S, Dai Z, Guo P, Fu X, Liu S, Zhou L, Tang W, Feng T, Chen M, Zhan L, Wu T, Hu E, Jiang Y, Bo X, Yu G. 2021. ggtreeExtra: compact visualization of richly annotated phylogenetic data. Mol Biol Evol. https://doi.org/10.1093/molbev/msab166
17. Guzman KA, Daza AP, Gomez RV, Montenegro LM, Pazos A. 2024. Whole-genome sequences of *Helicobacter pylori* isolated from patients with high risk of gastric cancer in the Andes of Nariño, Colombia. Microbiol Resour Announc 13:e0123223. https://doi.org/10.1128/mra.01232-23
18. Thorell K, Muñoz-Ramírez ZY, Wang D, Sandoval-Motta S, Boscolo Agostini R, Ghirotto S, Torres RC, Falush D, Camargo MC, Rabkin CS, HpGP Research Network. 2023. The *Helicobacter pylori* genome project: insights into *H. pylori* population structure from analysis of a worldwide collection of complete genomes. Nat Commun 14:8184. https://doi.org/10.1038/s41467-023-43562-y