Draft Genome Sequences of 29 *Helicobacter pylori* Strains Isolated from Colombia

Angela B. Muñoz, Johanna Stepanian, Carmen Acosta, Juan S. Solano-Gutierrez, Filipa F. Vale, Alba A. Trespalacios-Rangel

**ABSTRACT**

Here, we present the draft genome sequences of 29 Colombian *Helicobacter pylori* strains. These strains were isolated in Bogotá, Colombia, from patients diagnosed with chronic gastritis. The genomic characterization of these strains will provide more information on the genetic composition of *H. pylori* strains from Colombia.

*Helicobacter pylori* is a Gram-negative, pathogenic bacterium capable of colonizing and persisting in the human stomach. The infection is considered the most frequent chronic bacterial infection worldwide (1–3), reaching prevalence rates of up to 80% in Colombia (4).

This report announces the genome sequences of 29 *H. pylori* strains isolated between 2009 and 2010 from patients residing in Bogotá, Colombia. The patients who signed informed consent were 48 years old on average (range, 18 to 79 years); from the histology results, 65.5% were diagnosed with chronic nonatrophic gastritis and 34.5% with chronic atrophic gastritis. The strains were recovered from gastric biopsy samples, and those were cultivated on BBL *Brucella* agar (Becton, Dickinson) supplemented with 7% horse blood, 0.4% IsoVitalex (BD, USA), and 0.2% Dent selective supplement (Oxoid, UK) under microaerophilic conditions (11% CO₂) at 37°C for 4 to 7 days. The strains were preserved in 20% glycerol and stored until required for DNA extraction. They were recovered by culture every time. After that, total DNA was extracted using a DNeasy blood and tissue kit (Qiagen, Hilden, Germany) following the manufacturer’s instructions. Fluorometric assay DNA quantification was performed using a Qubit 2.0 fluorometer and the Qubit double-stranded DNA (dsDNA) high-sensitivity (HS) assay kit (Life Technologies, Carlsbad, CA, USA). To verify that the DNA obtained was from *H. pylori*, a conventional PCR technique for the vacA gene was carried out. The primers and protocols previously described by Atherton et al. (5) were used.

Genomic DNA was sequenced using the MiSeq platform (Illumina, San Diego, CA); DNA libraries were prepared using a Nextera XT DNA library preparation kit (Illumina), followed by 2 × 300-bp paired-end sequencing resulting in 80 × coverage. The low-quality sequences were removed with the software package Trimmmomatic v0.39 (6). The reads were used for de novo genome assembly with SPAdes v13.3 (7). Assembly statistics for analyzed strains are provided in Table 1. The sequences were annotated using the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) (8). Default parameters were used for all software tools unless otherwise specified.

A multilocus sequence typing (MLST) analysis was performed based on seven *H. pylori* housekeeping genes (*atpA*, *efp*, *trpC*, *ppa*, *mutY*, *yphC*, and *ureI*). The sequences of these genes from 741 strains available at PubMLST (http://pubmlst.org/helicobacter/) (9) and previously described by Falush et al. (10) and Linz et al. (11), plus the 29 strains included in this study, were aligned using MAFFT v7 (12). Then, the aligned sequences...
| Strain name | GenBank accession no. | SRA accession no. | CDS | Genome size (bp) | GC content (%) | MLST | No. of contigs | N_50 value (bp) | Genome coverage (x) | No. of raw reads |
|-------------|----------------------|------------------|-----|-----------------|---------------|------|---------------|----------------|-------------------|----------------|}
| COL 1-PUJ   | JAFCHS00000000000    | SRR13796410      | 1,582 | 1,679,429       | 38.8          | HpEurope | 72         | 58,624          | 135               | 1,163,956 |
| COL 2-PUJ   | JACSDV00000000000    | SRR13796434      | 1,632 | 1,607,581       | 39            | HpEurope | 46         | 82,542          | 392               | 2,633,600 |
| COL 5-PUJ   | JACSDU00000000000    | SRR13796433      | 1,667 | 1,624,361       | 39.1          | HpEurope | 133        | 21,560          | 105               | 850,310  |
| COL 6-PUJ   | JACSDT00000000000    | SRR13796422      | 1,673 | 1,625,175       | 39.1          | HpEurope | 133        | 20,478          | 95                | 745,268   |
| COL 8-PUJ   | JACSH10000000000     | SRR13796409      | 1,664 | 1,661,424       | 38.9          | HpEurope | 43         | 93,756          | 286               | 2,823,146 |
| COL 9-PUJ   | JAFCHU00000000000    | SRR13796408      | 1,653 | 1,613,788       | 39            | HpEurope | 47         | 84,065          | 311               | 2,073,864 |
| COL 10-PUJ  | JAFCHV00000000000    | SRR13796407      | 1,680 | 1,642,843       | 39.4          | HpEurope | 58         | 107,835         | 375               | 2,741,078 |
| COL 11-PUJ  | JACW00000000000      | SRR13796406      | 1,663 | 1,626,191       | 39.4          | HpEurope | 58         | 86,547          | 221               | 1,639,572 |
| COL 12-PUJ  | JACX00000000000      | SRR13796432      | 1,633 | 1,637,127       | 38.9          | HpEurope | 43         | 96,539          | 368               | 2,773,188 |
| COL 13-PUJ  | JAFCY00000000000     | SRR13796431      | 1,714 | 1,658,899       | 39.3          | HpEurope | 58         | 59,276          | 554               | 4,615,312 |
| COL 14-PUJ  | JACZ00000000000      | SRR13796430      | 1,701 | 1,673,807       | 38.9          | HpEurope | 54         | 80,670          | 256               | 1,758,244 |
| COL 15-PUJ  | JAFCZ00000000000     | SRR13796429      | 1,563 | 1,546,556       | 39.2          | HpEurope | 46         | 60,019          | 133               | 987,226   |
| COL 16-PUJ  | JACSD50000000000     | SRR13796412      | 1,665 | 1,634,541       | 39.3          | HpEurope | 64         | 54,632          | 482               | 2,363,616 |
| COL 18-PUJ  | JAFCB00000000000     | SRR13796428      | 1,662 | 1,658,604       | 38.9          | HpEurope | 37         | 93,756          | 580               | 436,440   |
| COL 19-PUJ  | JAFCI00000000000     | SRR13796427      | 1,654 | 1,656,342       | 38.9          | HpEurope | 42         | 81,649          | 371               | 2,363,966 |
| COL 20-PUJ  | JAFCB00000000000     | SRR13796426      | 1,674 | 1,619,537       | 39            | HpEurope | 56         | 59,593          | 340               | 2,201,658 |
| COL 21-PUJ  | JAFCI00000000000     | SRR13796425      | 1,670 | 1,618,229       | 39            | HpEurope | 55         | 58,359          | 435               | 3,168,290 |
| COL 23-PUJ  | JACDR00000000000     | SRR13796411      | 1,661 | 1,614,131       | 39.1          | HpEurope | 91         | 32,947          | 99                | 762,738   |
| COL 24-PUJ  | JAFCI00000000000     | SRR13796424      | 1,674 | 1,622,444       | 39            | HpEurope | 119        | 25,019          | 119               | 922,348   |
| COL 25-PUJ  | JAFCI00000000000     | SRR13796423      | 1,681 | 1,662,282       | 38.9          | HpEurope | 88         | 37,899          | 113               | 909,004   |
| COL 26-PUJ  | JAFCI00000000000     | SRR13796421      | 1,631 | 1,619,895       | 39            | HpEurope | 52         | 97,025          | 208               | 1,447,706 |
| COL 27-PUJ  | JAFCI00000000000     | SRR13796420      | 1,656 | 1,653,551       | 38.9          | HpEurope | 36         | 80,500          | 359               | 2,247,256 |
| COL 28-PUJ  | JAFCI00000000000     | SRR13796419      | 1,671 | 1,653,809       | 38.9          | HpEurope | 23         | 132,947         | 479               | 3,095,502 |
| COL 29-PUJ  | JAFCI00000000000     | SRR13796418      | 1,585 | 1,586,826       | 39.2          | HpEurope | 31         | 144,624         | 340               | 2,443,034 |
| COL 30-PUJ  | JAFCI00000000000     | SRR13796417      | 1,596 | 1,587,263       | 39.1          | HpEurope | 35         | 92,323          | 339               | 2,478,576 |
| COL 31-PUJ  | JAFCI00000000000     | SRR13796416      | 1,785 | 1,711,739       | 39            | HpEurope | 14         | 41,620          | 363               | 2,430,084 |
| COL 49-PUJ  | JAFCI00000000000     | SRR13796415      | 1,786 | 1,735,252       | 39            | HpEurope | 96         | 59,657          | 255               | 2,107,660 |
| COL 50-PUJ  | JAFCI00000000000     | SRR13796414      | 1,726 | 1,711,312       | 38.9          | HpEurope | 57         | 118,609         | 522               | 4,410,236 |
| COL 51-PUJ  | JAFCI00000000000     | SRR13796413      | 1,656 | 1,666,802       | 38.9          | HpEurope | 39         | 82,225          | 404               | 2,801,844 |

*CDS, coding DNA sequences.*
were analyzed in the Structure 2.3.4 software (13–15) and the MEGA 7.0 software (16). For these analyses, previously reported recommendations (17, 18) were followed, and the results revealed that all Colombian isolates included in this study were classified as HpEurope.

The data reported here provide information on the genetic population structure of Colombian H. pylori. This information will help future functional comparative genomic studies that will greatly enhance the understanding of H. pylori infection dynamics in the Latin American region.

**Data availability.** This whole-genome shotgun project has been deposited in GenBank under accession number PRJNA656306. The accession numbers for the genomes are provided in Table 1.

**ACKNOWLEDGMENTS**

We thank the entities that financially supported the development of this work. A.B.M. is a recipient of a scholarship from the Centro de Estudios Interdisciplinarios Básicos y Aplicados (CEIBA) Foundation, Colombia; and A.B.M., C.A., J.S., and A.A.T.-R. are recipients of a project grant (120380763025/2018) from MinCiencias, Colombia. The work is partially supported by Research Vice-Rectory, Pontificia Universidad Javeriana (PPTA_7676) and F.P.I.T. BanRepCultural (project 3956). F.F.V. is financed by national funds from the Fundación 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; these funds partially supported this work. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

**REFERENCES**

1. Dang BN, Graham DY. 2017. Helicobacter pylori infection and antibiotic resistance: a WHO high priority? Nat Rev Gastroenterol Hepatol 14:383–384. https://doi.org/10.1038/nrgastro.2017.57.
2. Peek RM, Blaser MJ. 2002. Helicobacter pylori and gastrointestinal tract adenocarcinomas. Nat Rev Cancer 2:28–37. https://doi.org/10.1038/nrc703.
3. Camilo V, Sugiyama T, Touati E. 2017. Pathogenesis of Helicobacter pylori infection. Helicobacter 22:e12405. https://doi.org/10.1111/hel.12405.
4. Bravo LE, Cortés A, Canascal E, Jaramillo R, García LS, Bravo PE, Badel A, Bravo PA. 2003. Helicobacter pylori: patología y prevalencia en biopsias gástricas en Colombia. Colomb Med 34:124–131.
5. Atherton JC, Cao P, Peek RM, Tummuru MK, Blaser MJ, Cover TL. 1995. Mosaicism in vacuolating cytotoxin alleles of Helicobacter pylori. Association of specific vacA types with cytotoxin production and peptic ulceration. J Biol Chem 270:17771–17777. https://doi.org/10.1074/jbc.270.30.17771.
6. Bolger AM, Lohse M, Usadel B. 2014. Trimomatic: a flexible trimmer for illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
7. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prijibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
8. 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.
9. Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res 3:124. https://doi.org/10.12688/wellcomeopenres.14826.
10. Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, Blaser MJ, Graham DY, Vacher S, Perez-Perez GI, Yamaoka Y, Mégraud F, Otto K, Reichard U, Kotzowitch E, Wang X, Achtman M, Suerbaum S. 2003. Traces of human migrations in Helicobacter pylori populations. Science 299:1582–1585. https://doi.org/10.1126/science.1080857.
11. Linz B, balloux F, Moodley Y, Manica A, Liu H, Roumagnac P, Falush D, Stamer C, Prugnolle F, van der Merwe SW, Yamaoka Y, Graham DY, Perez-Trallero E, Wadstrom T, Suerbaum S, Achtman M. 2007. An African origin for the intimate association between humans and Helicobacter pylori. Nature 445:915–918. https://doi.org/10.1038/nature05562.
12. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.
13. Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–955.
14. Falush D, Stephens M, Pritchard JK. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecol Notes 7:574–578. https://doi.org/10.1111/j.1471-8286.2007.01758.x.
15. Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587.
16. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054.
17. Vale FF, Vadiivelu J, Oleastro M, Breurec S, Engrand L, Perets TT, Mégraud F, Lehours P. 2015. Dormant phases of Helicobacter pylori reveal distinct populations in Europe. Sci Rep 5:14333. https://doi.org/10.1038/srep14333.
18. Muñoz AB, Trespalacios-Rangel AA, Vale FF. 2021. An American lineage of Helicobacter pylori prophages found in Colombia. Helicobacter 26:e12779. https://doi.org/10.1111/hel.12779.