

Análisis Genómico Comparativo de Aislamientos Colombianos de *Helicobacter pylori*:
Viruloma, Resistoma, Filogenia e Identificación de Profagos.



ANGELA BIBIANA MUÑOZ DELGADO

Tesis presentada como requisito parcial para obtener el título de:
Doctora en Ciencias Biológicas

Directora:

Alba Alicia Trespalacios. PhD.

Codirectora:

Filipa Vale. PhD.

Pontificia Universidad Javeriana
Facultad De Ciencias
Doctorado En Ciencias Biológicas
Bogotá D.C.

2021

NOTA DE ADVERTENCIA

“La Universidad no se hace responsable por los conceptos emitidos por sus alumnos en sus trabajos de tesis. Solo velara porque no se publique nada contrario al dogma y a la moral católica y porque las tesis no contengan ataques personales contra persona alguna, antes bien se vea en ellas el anhelo de buscar la verdad y la justicia”

Artículo 23 de la Resolución N° 13 de julio de 1946.



**Análisis Genómico Comparativo de Aislamientos Colombianos
de *Helicobacter pylori*: Viruloma, Resistoma, Filogenia e
Identificación de Profagos.**

ANGELA BIBIANA MUÑOZ DELGADO

Alba Alicia Trespalcios. PhD.
Decana
Facultad de Ciencias.

Jairo Sutachan Rubio. PhD.
Director de Posgrados
Facultad de Ciencias.

Bogotá, agosto de 2021

**Análisis Genómico Comparativo de Aislamientos Colombianos de *Helicobacter pylori*:
Viruloma, Resistoma, Filogenia e Identificación de Profagos.**

APROBADO:

Alba Alicia Trespalacios. PhD.
Directora.

Filipa Vale. PhD
Co-Directora

Firmado digitalmente por Teresa Alarcón
Fecha: 2021.08.12 18:37:13 +02'00'

Teresa Alarcón Caveró. PhD
Jurado

Digitally signed by Jovanny Zabaleta
DN: cn=Jovanny Zabaleta,
o=Academic, ou=LSUHSC-New Orleans,
email=jzabal@lsuhsc.edu, c=US
Date: 2021.08.17 13:33:19 -05'00'

Jovanny Zabaleta. PhD
Jurado

Mariana Piuri. PhD
Jurado

María Mercedes Bravo. MSc
Jurado

Javier Torres López. PhD
Jurado

AGRADECIMIENTOS

En primer lugar, le agradezco a mis directoras de tesis, las doctoras Alba Alicia Trespalacios y Filipa Vale, por su disposición, su apoyo, sus enseñanzas y sus consejos.

A la Dra. Trespalacios por la confianza que ha depositado en mí y en mi trabajo, por inculcar en mí la rigurosidad en el trabajo y el amor por *Helicobacter pylori*. Gracias por compartir conmigo esta pasión y sus enormes conocimientos.

A la Dra. Vale por su gran capacidad para guiarme, gracias por su paciencia, optimismo y disponibilidad para ayudarme. Gracias por acogerme en Lisboa y hacer de su laboratorio una casa para mí, como dice un famoso fado: “*Dois braços à minha espera, É uma casa portuguesa, com certeza*”

Quiero agradecer también a muchas otras personas e instituciones que de una u otra forma aportaron en el desarrollo de este trabajo:

A la Dra. Olga Villamizar, quien en algún momento actuó como asesora de este proyecto. Gracias por ayudarme a empezar a materializar esta idea.

A la Dra. Martha Vives, quien abrió mi mente al maravilloso mundo de los bacteriófagos, gracias por acogerme en su grupo y por brindarme desinteresadamente sus enseñanzas.

A todo el grupo de trabajo del Laboratorio de Bacteriología Especial, en donde no solo encontré compañeros de trabajo sino también amigos. El desarrollo de este trabajo no hubiera sido posible sin su apoyo. Todos contribuyeron con su compañía y su afecto.

A todo el grupo de trabajo de la línea de investigación de *Helicobacter*, especialmente a mis coequiperas Carmen y Johanna. Gracias por brindarme su apoyo técnico, pero sobre todo gracias por acompañarme en este largo camino.

A las instituciones financiadoras de este proyecto, MinCiencias (proyecto 120380763025/2018); F.P.I.T. Banco de la Republica (proyecto 3956) y la Vicerrectoría de Investigación de la Pontificia Universidad Javeriana (proyecto PPTA_7676).

A la Fundación CEIBA (Centro de Estudios Interdisciplinarios Básicos y Aplicados), entidad que me otorgó la beca que hizo posible cursar mis estudios de doctorado.

A mis amigos, que afortunadamente son tantos que no los puedo listar. También a mis primos, a quienes también cuento como mis amigos. Gracias porque todos aportaron un poco, con una sonrisa, con un café, con una palabra de apoyo.

Por último, pero no menos importante, quiero agradecer a mi familia, ellos son mi mayor felicidad y orgullo, pero también mi soporte y mi consuelo.

A mi esposo y a mi hijo, Andrés y Francisco, mi amor más grande, mi luz y mi verdad. Gracias por estar siempre a mi lado, por sus enormes sacrificios y porque durante este proceso conocieron lo mejor y lo peor de mí y aun así me siguen amando.

A mis padres, gracias por su apoyo incondicional y por ser los mejores abuelos. Gracias a mi madre por darme tanto amor y cuidado, gracias por hacer hasta lo imposible para garantizar mi educación. ¡Lo hemos logrado!

Dedicatoria

A Andrés y Francisco.

INDICE GENERAL

RESUMEN	15
ABSTRACT	17
1. INTRODUCCIÓN Y PLANTEAMIENTO DEL PROBLEMA	19
2. OBJETIVOS	23
2.1 OBJETIVO GENERAL	23
2.2 OBJETIVOS ESPECÍFICOS	23
3. CAPITULO 1.	25
ANTECEDENTES Y MARCO TEORICO	25
3.1 <i>Helicobacter pylori</i> : Generalidades.....	25
3.2 Genoma, diversidad genética y estructura poblacional de <i>Helicobacter pylori</i>	25
3.3 Patogenia e infección por <i>Helicobacter pylori</i>	27
3.3.1 Factores de virulencia de <i>H. pylori</i>	28
3.4 Tratamiento de erradicación de <i>H. pylori</i> y resistencia antimicrobiana.	33
3.4.1 Mecanismos moleculares involucrados en la resistencia a los antimicrobianos.	35
3.5 Caracterización de bacteriófagos de <i>H. pylori</i>	40
3.5.1 Profagos de <i>H. pylori</i>	40
3.5.2 Diversidad y filogenia de profagos de <i>H. pylori</i>	42
4. CAPITULO 2.	44
CARACTERIZACIÓN DE FACTORES DE VIRULENCIA Y GENES RELACIONADOS CON RESISTENCIA A LOS ANTIMICROBIANOS EN GENOMAS COLOMBIANOS DE <i>H. pylori</i>	44
4.1 INTRODUCCIÓN.....	44
4.2 METODOLOGÍA.....	45
4.2.1 Obtención de los genomas.....	45

4.2.2	Determinación <i>in silico</i> de variantes genéticas en los factores de virulencia	48
4.2.3	Detección <i>in silico</i> de mutaciones en genes relacionados con resistencia a los antimicrobianos.....	51
4.2.4	Análisis estadístico.....	52
4.3	Resultados.....	53
4.3.1	Características de los genomas y datos demográficos de los pacientes a partir de los cuales se obtuvieron los aislamientos de <i>H. pylori</i>	53
4.3.2	Distribución de genotipos de virulencia en aislamientos colombianos de <i>H. pylori</i>	55
4.3.3	Determinación <i>in silico</i> de las mutaciones asociadas con resistencia a los antimicrobianos.....	61
4.4	DISCUSIÓN.....	66
4.5	CONCLUSIONES.....	79
5.	CAPITULO 3.	80
	ANÁLISIS DE DIVERSIDAD Y FILOGENIA DE GENOMAS COLOMBIANOS DE <i>H. pylori</i> .	80
5.1	INTRODUCCIÓN.....	80
5.2	METODOLOGÍA.....	82
5.2.1	Análisis filogenético por MLST.....	82
5.2.2	Análisis filogenético por búsqueda de polimorfismos de nucleótido simple.....	83
5.2.3	Estimación del Pangenoma	83
5.3	RESULTADOS	84
5.3.1	Análisis filogenético por MLST.....	84
	Análisis filogenético por SNP's.....	86
	Caracterización del pangenoma	87

5.4 DISCUSIÓN.....	91
5.5 CONCLUSIONES.....	94
6. CAPITULO 4.	95
CARACTERIZACIÓN DE PROFAGOS DE <i>H. pylori</i> EN AISLAMIENTOS COLOMBIANOS	95
6.1 INTRODUCCIÓN.....	95
6.2 METODOLOGÍA.....	96
6.2.1 Cepas de <i>H. pylori</i>	96
6.2.2 Búsqueda de profagos de <i>H. pylori</i>	97
6.2.3 Análisis de Prophage Sequence Typing (PST)	98
6.3 RESULTADOS	99
6.3.1 Identificación de profagos de <i>H. pylori</i>	99
6.3.2 Estructura poblacional de los profagos de <i>H. pylori</i>	103
6.4 DISCUSIÓN.....	108
6.5 CONCLUSIONES.....	112
7. CONCLUSIONES GENERALES	113
8. RECOMENDACIONES	114
BIBLIOGRAFÍA	116
ANEXOS	134

INDICE DE FIGURAS

Figura 1. Alineamiento de la secuencia de aminoácidos de la región 3' de <i>cagA</i> para cuatro genomas colombianos de <i>H. pylori</i>	55
Figura 2. Frecuencia de mutaciones asociadas a RAM	64
Figura 3. Distribución de RAM en el tiempo en los genomas colombianos analizados.	66
Figura 4. Poblaciones de <i>H. pylori</i> por análisis de MLST en el software Structure	85
Figura 5. Análisis filogenético por MLST de los genomas colombianos y 741 genomas de referencia.	85
Figura 6. Caracterización filogenética por SNP de los genomas colombianos	87
Figura 7. Caracterización del pangenoma colombiano de <i>H. pylori</i>	88
Figura 8. Número de genes que hacen parte del pangenoma de acuerdo con el número de genomas incluidos en el análisis.....	88
Figura 9. Número total de genes respecto a número de genes conservados en función del número de genomas analizados.	89
Figura 10. Diagrama de Venn para el análisis de pangenoma en los diferentes grupos poblacionales de los genomas colombianos analizados.	90
Figura 11. Alineamiento de los profagos colombianos de <i>H. pylori</i>	101
Figura. 13. Árbol filogenético de los genes de integrasa y holin	106
Figura 14. Árbol filogenético de genomas completos de 40 profagos	107

INDICE DE TABLAS

Tabla 1.	47
Genomas colombianos secuenciados en este estudio.	47
Tabla 2.	48
Cebadores usados para la determinación <i>in silico</i> de factores de virulencia.	48
Tabla 3.	49
Secuencia de aminoácidos de los motivos EPIYA de la proteína CagA.	49
Tabla 4.	51
Mutaciones asociadas a mecanismos de resistencia a los antimicrobianos.	51
Tabla 5.	54
Año de obtención de los aislamientos de los cuales se obtuvieron los genomas.	54
Tabla 6.	56
Análisis de Motivos EPIYA en genomas <i>cagA</i> positivos	56
Tabla 7.	56
Estructura del islote de patogenicidad <i>cagPAI</i> en los genomas analizados.	56
Tabla 8.	57
Positividad de genes componentes de <i>cagPAI</i>	57
Tabla 9.	58
Combinaciones alélicas del gen <i>vacA</i> en los genomas analizados.	58
Tabla 10.	59
Frecuencia de alelos en las regiones analizadas del gen <i>vacA</i>	59
Tabla 11.	60
Frecuencia de genes de adherencia en los genomas analizados	60
Tabla 12.	61

Análisis de genes de adherencia frente a la presencia de <i>cagPAI</i>	61
Tabla 13.....	62
Frecuencia de mutaciones asociados a RAM en los genomas analizados.....	62
Tabla 14.....	65
Distribución de RAM en el tiempo en los genomas colombianos analizados.	65
Tabla 15.....	89
Análisis de pangenoma por grupo poblacionales	89
Tabla 16. Iniciadores utilizados para la detección de los genes holin e integrasa.....	98
Tabla 17.....	100
Características de los profagos reportados al NCBI.....	100
Table 18. Sitios de Inserción para profagos colombianos.....	102
Tabla 19. Resultados del análisis de PST en el software Structure.....	103
Table 20. Resultados para el analisis en Structure Harvester.....	105

INDICE DE ANEXOS

ANEXO 1. Bacteriophages of <i>Helicobacter pylori</i>	135
ANEXO 2. Draft genome sequences of 29 <i>Helicobacter pylori</i> strains isolated from Colombia.	146
ANEXO 3. An American lineage of <i>Helicobacter pylori</i> prophages found in Colombia..	150
ANEXO 4. Información de genomas colombianos de <i>H. pylori</i> recolectados de bases de datos públicas.	163

RESUMEN

Helicobacter pylori es una bacteria patógena capaz de infectar el estómago humano. Se considera la infección bacteriana crónica más frecuente en todo el mundo. Todos los individuos infectados con *H. pylori* presentan gastritis crónica, y el 20% pueden progresar a enfermedades como úlcera péptica, adenocarcinoma o linfoma gástrico tipo MALT. Numerosos factores se han descrito como responsables de la progresión de la infección, dentro de ellos, varios factores de virulencia de *H. pylori*. Un desafío crítico que presenta la infección por *H. pylori*, son los altos niveles de resistencia a los antibióticos. Se han informado tasas de efectividad del tratamiento antibiótico tan bajas como del 57%, cuando la mínima aceptable para un tratamiento empírico es del 90%. Además de ser una bacteria virulenta y resistente a los antibióticos, *H. pylori* es una bacteria genéticamente diversa. Fenómeno que puede estar influenciado entre otras causas, por la presencia de elementos genómicos móviles, como los profagos. Esta diversidad bacteriana se ha asociado con el origen geográfico de las poblaciones y cuando se ha analizado en conjunto con la diversidad de los profagos, la mayoría de las observaciones indican una concordancia filogeográfica; sin embargo, algunos profagos de *H. pylori* se han asignado a poblaciones distintas de su anfitrión, mientras que otros han exhibido signos de recombinación entre poblaciones.

Teniendo en cuenta estos antecedentes, el objetivo de este trabajo fue caracterizar molecularmente genomas colombianos de *H. pylori*, incluyendo filogenia, resistoma, viruloma y caracterización de profagos. Para ello, se analizaron 221 genomas colombianos de *H. pylori*, los cuales incluyeron 29 genomas secuenciados en este estudio y 192 genomas recolectados de bases de datos públicas. Para la determinación del viruloma, se realizó la caracterización de los principales factores de virulencia de *H. pylori*: *cagA* (motivos EPIYA), *cagPAI*, *vacA* (regiones *m*, *s* e *i*), *iceA*, *babA* y genes de adherencia. Para conocer el resistoma, se realizó la búsqueda de mutaciones en los genes y/o proteínas relacionadas con resistencia a Claritromicina (CLA), Tetraciclina (TET), Levofloxacina (LEV), Amoxicilina (AMX) y Metronidazol (MTZ). Para el análisis de filogenia, se realizó caracterización por MLST (Multi-locus sequence typing) y análisis mediante SNP (polimorfismos de nucleótido simple), adicionalmente se realizó una caracterización del pangenoma colombiano de *H. pylori* utilizando el software Roary. La búsqueda y caracterización de profagos, se realizó

utilizando los genes marcadores integrasa y holin, así como de la secuencia completa de los profagos. Se realizó una caracterización genómica y filogenética de los profagos colombianos de *H. pylori*.

El análisis de los genomas mostró que estos provenían de aislamientos de *H. pylori* obtenidos en diferentes regiones de Colombia entre los años 2000 y 2018. Al analizar el viruloma se logró evidenciar que el genotipo de virulencia más frecuentemente encontrado fue *vacAs1b i1 ml / babA2 / iceA2 / cagPAI* parcialmente deletado con *cagA+* (EPIYA-ABC), un genotipo definido como de alta virulencia. El análisis de resistoma permitió evidenciar la presencia de mutaciones asociadas a resistencia a CLA en el 3,6% de los genomas analizados, a TET en el 3,61%, a AMX en el 7,23%, a LEV en el 10,26% y a MTZ en el 75,7%. Se encontraron 32 (14,4%) genomas con mutaciones ligadas a resistencia frente a más de un antibiótico. Estos análisis reflejan las características de virulencia y las mutaciones asociadas a resistencia de las cepas que han circulado en el país a lo largo de las dos últimas décadas.

El estudio de la estructura poblacional reveló la presencia de un grupo clonal compuesto exclusivamente por aislamientos colombianos. El análisis de pangenoma permitió evidenciar que el pangenoma colombiano de *H. pylori* es abierto y muy diverso, lo cual está relacionado con la amplia diversidad genética y la historia evolutiva de esta bacteria humana. Los análisis de búsqueda y caracterización de profagos permitieron encontrar profagos o parte de ellos en 11 de los genomas analizados, incluyendo los genomas secuenciados en este estudio y los colombianos disponibles en bases de datos públicas. Los análisis de filogenia para estos profagos mostraron evidencia de una nueva población filogenética formada exclusivamente por profagos de origen colombiano. Este nuevo grupo clonal fue denominado hpColombia.

Los resultados obtenidos aquí permiten ampliar el conocimiento de la genómica de aislamientos colombianos de *H. pylori*, dar a conocer la presencia de profagos en éstos, comprobar la presencia de un grupo clonal específico para los profagos colombianos de *H. pylori* y también para sus bacterias hospedadoras. Adicionalmente, los análisis realizados aquí permiten verificar que la genómica comparativa es una herramienta útil en la evaluación de varios elementos que contribuyen a la diversidad entre los genomas de *H. pylori*.

ABSTRACT

Helicobacter pylori is a pathogenic bacterium capable of infecting the human stomach. This infection is considered the most common chronic bacterial infection in the world. All individuals infected with *H. pylori* develop chronic gastritis, and 20% of them may progress to diseases such as peptic ulcer, adenocarcinoma, or gastric MALT lymphoma. Numerous factors have been described as responsible for the progression of the infection. Among them, several factors of virulence of *H. pylori*. Another critical challenge presented by *H. pylori* infection is its high levels of resistance to antibiotics. Antibiotic treatment effectiveness rates as low as 57% have been reported, when the minimum acceptable for empirical treatment is 90%. In addition to being a virulent and antibiotic-resistant bacterium, *H. pylori* is a genetically diverse bacterium, which can be influenced by the presence of mobile genomic elements, such as prophages. This diversity has been associated with the geographic origin of the populations. Most observations indicate a phylogeographic agreement between prophage and bacterial genes; however, some *H. pylori* prophages have been assigned to populations other than their host, while others show signs of recombination between populations.

Considering these antecedents, the objective of this work was to characterize molecularly Colombian *H. pylori* genomes. The characterization included phylogeny, resistome, virulome, and the presence of prophages. For this, 221 Colombian *H. pylori* genomes were analyzed, including 29 genomes sequenced in this study and 192 genomes collected from public databases. To determine the virulome, the *H. pylori* major virulence factors were characterized: *cagA* (including EPIYA motifs), *cagPAI*, *vacA* (*m*, *s* and *i* regions), *iceA*, *babA*, and some adherence genes. A search for mutations in genes and/or proteins related to resistance to Clarithromycin (CLA), Tetracycline (TET), Levofloxacin (LEV), Amoxicillin (AMX), and Metronidazole (MTZ) was made to determine the resistome. For the phylogeny analysis, the characterization was performed by MLST (Multi-locus sequence typing) and SNP (single nucleotide polymorphisms); additionally, a description of the Colombian *H. pylori* pangenome was performed using Roary software. The characterization of prophages was carried out by searching the integrase and holin genes and the genomic and phylogenetic characterization of the complete sequence of the prophages.

The genomes analysis showed that they came from *H. pylori* isolates obtained in different regions of Colombia between 2000 and 2018. Virulome analysis showed that the most frequently found virulence genotype was *vacAs1bi1m1 / babA2 / iceA2 / cagPAI* partially deleted with *cagA+* (EPIYA-ABC), a genotype defined as high virulence. The resistome analysis revealed the presence of mutations associated with resistance to CLA in 3.6% of the genomes analyzed, to TET in 3.61%, to AMX in 7.23%, to LEV in 10, 26%, and MTZ at 75.7%. 32 (14.4%) genomes were found with mutations linked to resistance to more than one antibiotic. These analyzes reflect the virulence characteristics and the resistance-associated mutations present in *H. pylori* strains that have circulated in the country over the last two decades.

The study of the population structure by MLST showed that the Colombian isolates belong to the clonal group hpAfrica1 and mainly to the group HpEurope; the complementary analyzes showed the presence of a clonal group composed exclusively of Colombian isolates. The pangenome analysis showed that the Colombian *H. pylori* pangenome is open and very diverse. In these analyzes, the presence of some marker genes of geographic distribution was not found. The analysis of prophages allowed us to find three complete prophages and two incomplete prophages in the sequenced isolates and the presence of prophages in six genomes recovered from public databases. Phylogenetic analyzes for these prophages showed evidence of a new phylogenetic population made up exclusively of prophages of Colombian origin. This new clonal group was named hpColombia.

The results obtained here allow us to expand the knowledge of the genomics of Colombian *H. pylori* isolates. Likewise, the results will enable us to inform the presence of prophages in these genomes and verify the existence of a specific clonal group for this region of both prophages and their host bacteria. The analyses carried out here allow us to confirm that comparative genomics is a helpful tool in evaluating various elements that contribute to the diversity among *H. pylori* genomes.

1. INTRODUCCIÓN Y PLANTEAMIENTO DEL PROBLEMA

Helicobacter pylori (*H. pylori*) es una bacteria Gram negativa, capaz de colonizar y persistir en el estómago humano; su infección se ha asociado con el desarrollo de patologías como gastritis crónica, úlcera péptica y cáncer gástrico; desde 1994 fue clasificada por la Agencia Internacional para la Investigación en Cáncer (IARC) de la Organización Mundial de la Salud (OMS) como carcinógeno del grupo I (Dang & Graham, 2017).

La infección por *H. pylori* es una de las infecciones crónicas más comunes en el mundo, lo que convierte a esta bacteria en uno de los patógenos humanos más exitosos (Hooi et al., 2017; Peek & Blaser, 2002). La prevalencia de la infección por *H. pylori* alcanza cifras de hasta el 80% en los países en vía de desarrollo, como Colombia (Hunt et al., 2011). Sin embargo, se han observado diferencias en la prevalencia de infección dentro de diferentes ubicaciones geográficas de un mismo país, incluso se presentan diferencias en las tasas de resistencia a los antimicrobianos y en los genotipos circulantes entre regiones (Ghotaslou et al., 2015).

Uno de los principales problemas que conlleva la infección por *H. pylori* es el riesgo que representa en el desarrollo de cáncer gástrico; sin embargo, la incidencia de esta patología tiene una amplia variación geográfica. Las regiones de alta incidencia incluyen Asia oriental, Europa oriental, América Central y América del Sur (Wong et al., 2021). De acuerdo con los datos reportados en 2020 por la IARC, el cáncer gástrico es el tercer cáncer con mayor incidencia en Colombia (7.3 por 100000 habitantes) (IARC, 2020). Muchos estudios han demostrado asociación entre factores de virulencia y desarrollo de patologías, sin embargo, en Colombia ha sido más difícil comprobar esta hipótesis. A pesar de que estima que más del 80% de la población del país se encuentra infectada con *H. pylori* (Camargo et al., 2015; Correa & Piazuelo, 2010; Park et al., 2018), las tasas de incidencia de cáncer gástrico en Colombia difieren marcadamente entre las personas que residen en la Cordillera de los Andes en comparación con aquellas que viven en las costas (Torres et al., 2013). Adicionalmente los estudios que han caracterizado los factores de virulencia en el país, han utilizado tamaños de muestra pequeños y generalmente se han realizado sobre grupos poblacionales específicos; por esta razón, aunque existen un gran número de investigaciones, las

Introducción y Planteamiento del problema

conclusiones obtenidas por los diferentes estudios han sido contradictorias (Carlosama-Rosero et al., 2021).

Es bien conocido que la progresión de la enfermedad mediada por *H. pylori* depende de varios factores, incluidos condiciones propias del hospedero, el genotipo de la cepa bacteriana y ciertas condiciones ambientales (FitzGerald & Smith, 2021). Respecto al genotipo de *H. pylori*, los estudios han mostrado que esta bacteria exhibe una heterogeneidad genética extremadamente diversa. Diferentes determinantes de patogenicidad de *H. pylori* se han asociado con el progreso de la infección a patologías más severas. Los más estudiados son una gran variedad de adhesinas donde se destaca BabA, responsable de la adhesión y colonización de *H. pylori* en la mucosa gástrica (Kusters et al., 2006); la citotoxina vacuolizante VacA, presente en todas las cepas pero con presencia de polimorfismos significativos que afectan su actividad citotóxica (Matos et al., 2013); la oncoproteína CagA, presente en algunas cepas y ubicada en la isla de patogenicidad Cag (*cag*-PAI), la cual codifica un sistema de secreción de tipo IV que permite la translocación de CagA y otras moléculas efectoras bacterianas que actúan directamente en las células gástricas, dando como resultado, la inducción de vías de señalización intracelular que provocan una motilidad y proliferación celular aberrante (Salimzadeh et al., 2015); y el gen inducido por contacto por el epitelio *iceA* que codifica una proteína que actúa como endonucleasa y se activa al entrar en contacto con las células epiteliales lo que le permite a la bacteria proteger su genoma de diferentes agentes (Figueiredo et al., 2000).

Considerando las implicaciones de la infección por esta bacteria, la erradicación de *H. pylori* comúnmente se prescribe para el tratamiento o la prevención de estos trastornos; sin embargo, la emergente resistencia a los antimicrobianos (RAM) está aumentando en todo el mundo y es el principal factor del fracaso del tratamiento (Lim et al., 2016). Existen diferentes regímenes de tratamiento anti - *H. pylori*, típicamente basados en el uso de un inhibidor de la bomba de protones (IBP) más uno o dos antibióticos, dentro de los más comúnmente usados encontramos Claritromicina (CLA), Metronidazol (MTZ), Levofloxacin (LEV), Tetraciclina (TET) o Amoxicilina (AMX) (Malfertheiner et al., 2017). En Colombia, las cifras de resistencia primaria a los antibióticos más usados es alta, 13,6% para CLA (Trespacios et al., 2013), 81% para MTZ (Trespacios et al., 2010) y 27% para LEV (Trespacios-Rangél et al., 2016). Considerando esta problemática, es cada

Introducción y Planteamiento del problema

vez más evidente la necesidad de tener esquemas de tratamiento guiados por pruebas de susceptibilidad antibiótica (Malfertheiner et al., 2017). Sin embargo, *H. pylori* es un microorganismo de difícil crecimiento, lo cual dificulta la obtención de un perfil de susceptibilidad; por lo tanto, la detección molecular de la resistencia de *H. pylori* puede proporcionar un medio más rápido y fácil para este fin.

En general, es conocido que *H. pylori* se hace resistente a estos antibióticos principalmente mediante mutaciones puntuales en un par de bases del gen que codifica para el sitio de acción del antibiótico (Arévalo et al., 2019). El conocimiento de estas mutaciones puede proporcionar una guía para el manejo y las opciones de tratamiento, mediante el uso de herramientas de genómica que permitan predecir la RAM y llevar esta información a entornos clínicos. Este estudio analiza las mutaciones asociadas a RAM presentes en genomas colombianos de *H. pylori* y aporta conocimiento sobre los genotipos de resistencia que han estado circulando en el país a través del tiempo, esto a su vez aportará información que pudiera ser usada en el desarrollo de aplicaciones que permitan un diagnóstico rápido de la resistencia de *H. pylori* a los antimicrobianos.

Otro elemento recientemente descrito en los genomas de *H. pylori*, son los profagos, los cuales han mostrado tener una población filogenética estructurada, la cual al parecer contribuye a la diversidad genética de la bacteria (Vale et al., 2015). Hasta ahora no se han descrito profagos presentes en aislamientos de *H. pylori* de origen americano. En este estudio identificamos, caracterizamos y presentamos el análisis filogenético de los profagos presentes en *H. pylori* de origen colombiano.

Por otra parte, los estudios han mostrado que *H. pylori* es una bacteria que puede mostrar procesos adaptativos locales muy rápidos mediante mutación y recombinación homóloga con otras cepas (Furuta et al., 2015). Estas características han sido especialmente evidentes en Colombia, donde varios estudios recientes han descrito la presencia de un componente poblacional presente solamente en cepas de *H. pylori* aisladas de población mestiza colombiana (Gutiérrez-Escobar et al., 2017; Muñoz-Ramírez et al., 2017; Shiota et al., 2014), y han sugerido que en Colombia, las poblaciones bacterianas evolucionan rápidamente y han formado nuevas subpoblaciones de origen europeo (Muñoz-Ramírez et al., 2020). Estos

Introducción y Planteamiento del problema

hallazgos destacan la necesidad de realizar estudios filogenéticos más amplios en genomas colombianos de *H. pylori*.

Los múltiples y diversos estudios realizados en Colombia con genomas de *H. pylori*, han aportado un total de 192 genomas colombianos disponibles en bases de datos públicas. En este estudio, se utilizaron estos genomas junto con otros recientemente secuenciados, para desentrañar los perfiles genómicos de aislamientos colombianos de *H. pylori*. Se tuvieron en cuenta varios enfoques de tipificación *in silico* para analizar su contenido de resistoma, viruloma, presencia de profagos y caracterización filogenética mediante estrategias como SNP y MLST. El objetivo de este estudio fue realizar un análisis de epidemiología genómica que permitiera ampliar el conocimiento de posibles biomarcadores de la infección por *H. pylori* en Colombia y, por otra parte, el estudio pretende demostrar que los datos de WGS permiten hacer una caracterización rápida de todos los factores que pueden estar involucrados en la capacidad patógena de *H. pylori*. Lo cual se considera un aporte importante teniendo en cuenta el impacto que tiene la infección por *H. pylori* en las tasas de cáncer gástrico y el aumento de la expresión de RAM por *H. pylori* en Colombia.

2. OBJETIVOS

2.1 OBJETIVO GENERAL

Caracterizar la filogenia, el resistoma, el viruloma y la presencia de profagos en genomas colombianos de *Helicobacter pylori*.

2.2 OBJETIVOS ESPECÍFICOS

- Caracterizar los factores de virulencia y los genes relacionados con resistencia a antimicrobianos en genomas colombianos de *H. pylori*
- Determinar la diversidad y las relaciones filogenéticas de genomas colombianos de *H. pylori*.
- Describir las características de profagos presentes en genomas colombianos de *H. pylori*.

De acuerdo con los objetivos propuestos, la tesis estará dividida en 4 capítulos para su mejor comprensión:

Capítulo 1. Antecedentes y Marco teórico.

En este capítulo se describirán las características generales de *H. pylori*, los factores de virulencia más importantes, los mecanismos de resistencia a los antimicrobianos y los antecedentes que existen sobre la presencia de profagos en *H. pylori*. Este último tema se amplía en el artículo de revisión (Anexo 1):

Muñoz, A. B., Stepanian, J., Trespalacios, A. A., & Vale, F. F. (2020). Bacteriophages of *Helicobacter pylori*. *Frontiers in microbiology*, *11*, 549084. <https://doi.org/10.3389/fmicb.2020.549084>

Capítulo 2. Caracterización de factores de virulencia y genes relacionados con resistencia a los antimicrobianos en genomas colombianos de *H. pylori*.

En este capítulo se presentarán los resultados de la caracterización de los principales factores de virulencia y de búsqueda de mutaciones asociadas en RAM en genomas colombianos de *H. pylori*. Parte del trabajo desarrollado en este capítulo, específicamente lo relacionado con la secuenciación de genomas colombianos de *H. pylori*, se describe a profundidad en la publicación (Anexo 2):

Muñoz AB, Stepanian J, Acosta C, Solano-Gutierrez JS, Vale FF, Trespalacios-Rangel AA. 2021. Draft genome sequences of 29 *Helicobacter pylori* strains isolated from Colombia. *Microbiol Resour Announc* 10:e00218-21. <https://doi.org/10.1128/MRA.00218-21>

Capítulo 3. Análisis de diversidad y filogenia de genomas colombianos de *H. pylori*.

En este capítulo se mostrarán los resultados de la caracterización del pangenoma de 221 genomas colombianos de *H. pylori*. Así mismo se presentarán y discutirán los resultados del análisis de relaciones filogenéticas de estos genomas.

Capítulo 4. Caracterización de profagos de *H. pylori* en aislamientos colombianos.

En este capítulo, se analizan los resultados de la búsqueda de profagos en aislamientos colombianos de *H. pylori*. Se muestra la prevalencia de estos elementos en las cepas colombianas y se identifica la población filogenética de los profagos encontrados. Los resultados presentados en este capítulo pueden ser revisados también en la publicación (Anexo 3):

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>

3. CAPITULO 1.

ANTECEDENTES Y MARCO TEORICO

3.1 *Helicobacter pylori*: Generalidades.

H. pylori es una Proteobacteria, perteneciente al orden *Campylobacteriales*, de la familia *Helicobacter* (Goodwin & Worsley, 1993). Es una bacteria Gram negativa, microaerofílica, helicoidal, flagelada, que mide de 2 a 4 μm de largo y de 0,5 a 1 μm de ancho, tiene de 2 a 6 flagelos unipolares de 3 μm de largo los cuales le confieren movilidad y le permiten colonizar y persistir en el estómago humano (Kusters et al., 2006; Tomb et al., 1997). Esta bacteria tiene la capacidad de convertir su forma de hélice a forma cocoide, lo cual parece conferir ventajas para la supervivencia bacteriana en el microambiente gástrico (Baj et al., 2020).

H. pylori es una bacteria nutricionalmente exigente, de crecimiento lento; es catalasa, ureasa y oxidasa positiva (Kao et al., 2016). Además, *H. pylori* es capaz de formar biopelículas para disminuir su susceptibilidad a varios antibióticos, lo que genera mutaciones de resistencia a los antibióticos y mayores dificultades con la erradicación bacteriana (Baj et al., 2020).

Desde 1994, *H. pylori* ha sido clasificado como carcinógeno de clase I asociado con la aparición de cáncer gástrico (CG) por la Organización Mundial de la Salud (OMS), así como por la Agencia Internacional para la Investigación del Cáncer (IARC) (IARC, 1994; Ishaq & Nunn, 2015). *H. pylori* es un patógeno bacteriano altamente exitoso que infecta a más de la mitad de la población humana en todo el mundo (Camilo et al., 2017; Hooi et al., 2017; Peek & Blaser, 2002). Las principales rutas de transmisión descritas hasta ahora son oral - oral y fecal – oral. La prevalencia de las infecciones inducidas por *H. pylori* se estima en un 85-95% en los países en desarrollo y aproximadamente un 30-50% en los países desarrollados (Burucoa & Axon, 2017).

3.2 Genoma, diversidad genética y estructura poblacional de *Helicobacter pylori*

H. pylori tiene un genoma circular de 1,6 Mpb compuesto por aproximadamente 1901 secuencias codificantes. Se han descrito 1590 genes y el tamaño promedio de cada gen es 945pb; adicionalmente, se ha documentado la presencia de un sistema de regulación génica

Antecedentes y Marco Teórico

que incluye cuatro proteínas censoras de estímulos ambientales (Kusters et al., 2006). Aproximadamente el 75% de los genes de *H. pylori* son compartidos por todas las cepas y algunos están involucrados en la virulencia, por lo cual están sujetos a una fuerte presión de selección debido a su interacción con el hospedero (Rubio & Pérez-Pulido, 2021).

La tasa de mutación y recombinación de *H. pylori* es una de las más rápidas entre las bacterias, posiblemente para permitir su adaptación flexible al hospedador. Debido a su rápida evolución, sus huellas genómicas han servido como marcadores para caracterizar la historia humana (Linz et al., 2007). Durante su larga coevolución con los humanos, *H. pylori* ha ganado un nivel de diversidad y variabilidad genética excepcionalmente alto (Humbert et al., 2011). En la actualidad, se observan diferencias genéticas incluso entre aislamientos de un mismo paciente. Los mecanismos subyacentes a este fenómeno incluyen mecanismos de reparación del ADN relativamente ineficaces, así como competencia natural para la transformación y la capacidad de integrar pequeños fragmentos de ADN homólogo en el cromosoma (Ailloud et al., 2019). Esta capacidad de adaptación a nichos individuales y la alta recombinación genética han generado en *H. pylori* una protección bien estructurada contra la inmunidad del hospedador e incluso contra los antimicrobianos (Ailloud et al., 2019).

Recientemente, la disponibilidad de técnicas de secuenciación de genoma completo ha facilitado la comparación de secuencias para estudiar la evolución in vivo de *H. pylori* en diferentes etapas de la infección. Los estudios han mostrado que los genes que codifican las proteínas de la membrana externa (OMP) parecen evolucionar más rápido que los genes de otras categorías funcionales, lo que sugiere una adaptación a los huéspedes individuales; así mismo, los análisis han revelado que cepas del mismo huésped pueden diferenciarse genéticamente (Kennemann et al., 2011).

La caracterización filogenética de la bacteria ha permitido su clasificación en grandes poblaciones filogeográficas con distribuciones geográficas distintas, lo cual ha permitido el uso de genotipos de *H. pylori* para rastrear los eventos históricos de migración humana (Falush, Wirth, et al., 2003; Linz et al., 2007; Linz et al., 2014). Se ha reportado que en la población urbana latinoamericana suelen estar presentes cepas de origen europeo y africano, en diferentes proporciones (Thorell, Yahara, et al., 2017). Así mismo, se ha

reportado para cepas colombianas, peruanas y mexicanas una alta tasa de intercambio génico, las cuales probablemente ocurrieron durante la conquista española, fenómeno que implicó la migración masiva de poblaciones, principalmente de origen español y africano (Camorlinga-Ponce et al., 2011; de Sablet et al., 2011).

En Colombia, se ha reportado que en las zonas montañosas del país es más frecuente encontrar cepas de origen Europeo, mientras que en zonas costeras coexisten las poblaciones hpEuropa y hpAfrica1 (de Sablet et al., 2011). Estudios recientes han reportado que las cepas colombianas se agrupan en un clúster separado, lo cual ha sido atribuido al proceso de mestizaje que ha tenido lugar en Colombia y a la rápida evolución de la bacteria, producto de la co-adaptación con el hospedero (Muñoz-Ramírez et al., 2017). Se ha sugerido que esta línea evolutiva independiente de cepas colombianas no solo está influenciada por la población hpEurope sino también por grupos africanos (Gutiérrez-Escobar et al., 2017). Esta nueva población de cepas colombianas que ha sido denominada hspEuropeColombia y ha mostrado tener un perfil intermedio entre las poblaciones filogenéticas hspAfrica1 y hpEurope, lo cual evidencia la rápida evolución y dinamismo de *H. pylori*, y sugiere una amplia correlación entre la variación genética de la bacteria y su hospedero humano (Muñoz-Ramírez et al., 2020; Thorell, Yahara, et al., 2017).

3.3 Patogenia e infección por *Helicobacter pylori*

La infección por *H. pylori* por lo general se adquiere durante la infancia y puede establecer una infección crónica de por vida. Los pacientes infectados son asintomáticos en la mayoría de los casos, pero las complicaciones clínicas pueden variar desde úlceras gástricas o duodenales hasta atrofia gástrica y, en última instancia, cáncer gástrico o linfoma de tejido linfóide asociado a mucosas (MALT) (Suerbaum & Michetti, 2002).

La patogenicidad de *H. pylori* se asocia con varios mecanismos, algunos relacionados con el hospedero, entre los que destacan las alteraciones de las vías de señalización, la respuesta inflamatoria indirecta inducida dentro de la mucosa gástrica y los resultados epigenéticos directos en las células epiteliales gástricas (Machlowska et al., 2020); y otros relacionados con la capacidad que tiene *H. pylori* para desencadenar patología después de ingresar al

estómago del hospedero, en donde se conocen cuatro pasos críticos: 1- Supervivencia en condiciones ácidas del estómago, para lo cual, *H. pylori* utiliza su actividad ureasa para neutralizar la condición ácida hostil; 2- movimiento hacia las células del epitelio gástrico mediada por flagelos; 3- unión a los receptores del hospedero mediante adhesinas, lo que conduce a una colonización exitosa y una infección persistente; y 4- daño tisular por liberación de varias proteínas efectoras y toxinas, incluida la citotoxina A vacuolante (VacA) y la oncoproteína CagA, lo que daña el tejido infectado (Kao et al., 2016).

3.3.1 Factores de virulencia de *H. pylori*

La severidad de las enfermedades relacionadas con *H. pylori* se asocia con numerosos factores de virulencia; los cuales interactúan con el hospedero y el microambiente gástrico, determinando lo que será el curso de la infección. Los factores de virulencia de *H. pylori* no solo están involucrados en la inducción de respuestas inflamatorias, sino que también controlan y regulan esas respuestas, permitiendo que la inflamación sea crónica. Los factores de virulencia de *H. pylori* permiten la colonización y supervivencia de la bacteria dentro de la mucosa gástrica, lo que conduce a un mayor escape inmunológico y, en última instancia, a la inducción de alteraciones premalignas (Baj et al., 2020).

i. Factores de Adherencia

La adherencia de *H. pylori* en la mucosa juega un papel fundamental en la infección. Se estima que un 4% del genoma de este microorganismo codifica para OMP, muchas de ellas involucradas en la adhesión, y se han agrupado en una familia de proteínas conocidas como Hop (*Helicobacter* outer membrane protein) (Kusters et al., 2006). Dentro de las proteínas más conocidas de la familia Hop se encuentran HopP (SabA), HopH (OipA), HopQ, HopZ y HopS también conocida como BabA, siendo esta última la OMP mayormente estudiada. BabA tiene un peso de 78kD, permite la unión bacteriana de alta afinidad a los antígenos del grupo sanguíneo ABO / Lewis^b (Le^b) que abundan en las células epiteliales gástricas y en las mucinas, facilitando la liberación de factores proinflamatorios (Ansari & Yamaoka, 2017). Por lo tanto, como factor de virulencia, BabA promueve la infección e inflamación crónica de la mucosa, lo que da como resultado la persistencia bacteriana y el riesgo de progresión

de la enfermedad. Adicionalmente, se ha determinado que la expresión del gen *babA* está fuertemente asociada con la presencia del gen *cagA* y que depende del área geográfica particular donde las cepas sean aisladas (Thorell et al., 2016)

BabA pertenece a una familia que también incluye las proteínas BabB y BabC, que son muy similares, pero carecen de las propiedades adhesinas de unión a Le^b; la función exacta de BabB y BabC aún no se conoce. El análisis de la secuencia del gen *bab* reveló que existe una amplia homología en las regiones 5' y 3', particularmente entre *babA* y *babB* pero la variabilidad en la región media de la secuencia sugiere que la región divergente media de *babA* probablemente sea primordial en la función de unión (Ansari & Yamaoka, 2017).

Se ha estudiado la capacidad de BabA de unirse a los antígenos Le^b, ALe^b y BLe^b encontrando que el 95% de las cepas son capaces de unirse a los 3 antígenos, mientras que una pequeña proporción de cepas solo se unen a ALe^b conocidas como cepas “especializadas”. Se ha reportado que un 60% de cepas especializadas pertenecían a individuos de origen suramericano y fueron catalogadas como cepas de la población hpEurope (Kersulyte et al., 2000; Yamaoka et al., 2002), lo que probablemente pueda deberse a las mutaciones y/o recombinaciones de los últimos 500 años en respuesta a la presión medioambiental y a los patrones de glicosilación del hospedero.

BabA es codificada por el gen *babA*, un gen de única copia con dos alelos, *babA1* y *babA2*, los cuales se diferencian principalmente en la capacidad que brindan para adherirse, las cepas portadoras de *babA2* son capaces de adherirse a sus receptores celulares (Ilver et al., 1998). La incapacidad de adhesión en las cepas portadoras del alelo *babA1* ocurre por la presencia de una delección de 10pb en el péptido señal de este alelo, lo cual genera la eliminación del codón inicio de traducción y tiene como consecuencia la inhibición de la síntesis de la proteína BabA. Adicionalmente, se ha reportado que en los alelos de *babA* (*babA1* y *babA2*) puede ocurrir un cambio en el marco de lectura que provoca un aminoácido de parada en la posición 55 (Hennig et al., 2006). La mayoría de los estudios de esta adhesina utilizan técnicas basadas en PCR para detectar los alelos presentes, no obstante, debido a la similitud de secuencias entre *babA1*, *babA2* y *babB*, el uso de un solo par de cebadores puede no reflejar el estado de expresión real de BabA. Esto puede explicar los resultados contradictorios de los estudios que exploran la correlación entre la positividad de *babA2* y el

cáncer gástrico (Chang et al., 2018), por lo que, actualmente se recomiendan métodos de interacción antígeno-anticuerpo como los inmunoblots (Yamaoka, 2008).

ii. Isla de patogenicidad cag (cag PAI) y la oncoproteína CagA

La isla de patogenicidad *cag* (*cag PAI*) se encuentra en aproximadamente el 50% al 70% de las cepas de *H. pylori*. Mide 35-40 kb y contiene más de 30 genes (Alm et al., 1999). *cag PAI* no es fundamental en la infección, pero confiere ventajas competitivas para la bacteria. Dentro de los genes que lo componen, al menos 17 codifican para el sistema de secreción de tipo IV (T4SS), útil para codificar el sistema que transloca la proteína bacteriana CagA hacia el citoplasma de la célula gástrica del hospedero (Cover et al., 2020; Odenbreit et al., 2000). Además, otras proteínas como CagL y CagY, que están presentes en el T4SS utilizan la integrina $\beta 1$ como receptor para administrar CagA a la célula hospedera (Conradi et al., 2012).

CagA es una proteína efectora de 125-145 kDa, y las cepas que expresan esta proteína se consideran cepas altamente virulentas (Ansari & Yamaoka, 2020). Sin embargo, la presencia de *cagA* (un marcador de *cag PAI*) no predice resultados clínicos en poblaciones de alto riesgo, ya que la mayoría de *H. pylori* son *cagA* positivas, y entre los sujetos infectados menos del 3% desarrolla cáncer gástrico (Rizzato et al., 2020).

Una vez dentro de la célula epitelial la proteína CagA puede actuar dependiente o independiente de fosforilación (Backert & Selbach, 2008). En el primer caso, desencadena una cascada de fosforilación que conlleva a un descontrol de los procesos de morfogénesis y ciclo celular, además induce la proliferación (Arévalo et al., 2009). Cuando CagA actúa independiente de fosforilación afecta proteínas de unión estrecha como zonulina (ZO-1) y proteínas de adhesión como E-caderina, provocando alteraciones del complejo apical, destrucción de la barrera epitelial, cambios displásicos en la morfología de las células epiteliales, entre otros (Hatakeyama & Higashi, 2005).

CagA contiene un motivo bien caracterizado que comprende cinco aminoácidos, conocido como motivo EPIYA (ácido glutámico-prolina-isoleucina-tirosina-alanina). Este motivo se ubica en la región C-terminal, y junto con la secuencia vecina es responsable de la actividad biológica de CagA (Yamaoka, 2010). Estudios genómicos de *H. pylori* han revelado que

existe diversidad genética del motivo EPIYA entre cepas de *H. pylori* aisladas de diferentes zonas geográficas. De acuerdo con estos hallazgos, las cepas de *H. pylori* pueden poseer cuatro tipos diferentes de secuencias EPIYA, es decir, EPIYA-A, -B, -C y -D. Las secuencias A y B están presentes en todas las cepas mientras que la secuencia C se ha asociado con aislamientos occidentales y la D con cepas de origen asiático, y adicionalmente se ha reportado en cepas amerindias secuencias híbridas formadas por la mezcla de cepas africanas y europeas, estas secuencias han sido designadas EPIYA-DC, el cual aparenta ser menos virulento que los motivos no-híbridos (Bridge & Merrell, 2013). El motivo EPIYA-C puede tener múltiples repeticiones y el número de las mismas se correlaciona directamente con los niveles de fosforilación de la tirosina, es decir que las proteínas CagA con mayor número de repeticiones EPIYA-C son biológicamente más activas, generan cambios morfológicos más pronunciados en las células epiteliales y se han relacionado con mayor riesgo de desarrollo de cáncer (Xia et al., 2009).

La secuencia EPIYA-D tiene una mayor capacidad de desregulación de los procesos celulares que la secuencia EPIYA-C (Higashi et al., 2002). A pesar del papel de las secuencias vecinas de motivos EPIYA, se ha encontrado que la variación de aminoácidos en el propio motivo de fosforilación de tirosina EPIYA-B reduce el riesgo de desarrollar complicaciones graves. Se encontró que las cepas con polimorfismos de aminoácidos dentro del motivo EPIYA-B específico de occidente, como EPIYT-B, influyen en la actividad CagA, reduciendo la inducción de IL-8, lo cual puede traducirse en un riesgo más elevado de úlcera duodenal pero un riesgo menor de desarrollo de cáncer gástrico (Zhang et al., 2015).

Un estudio que caracterizó los motivos EPIYA de aislamientos de *H. pylori* colombianos, reportó que todos los aislamientos analizados eran del tipo *cagA* occidental. El 40% de los aislamientos presentó secuencias EPIYA ABC, siendo el motivo más frecuente y un 51,8% presentó múltiples repeticiones EPIYA-C (30,6% EPIYA ABCC y 21,2% EPIYA ABCCC), (Rodríguez Gómez et al., 2020). Varios reportes han mostrado que no se evidencia una relación directa entre la patología del paciente y el motivo EPIYA-C (Acosta et al., 2010; Khaledi et al., 2020; Rizzato et al., 2012), demostrando la importancia de utilizar *cagA* junto con otros factores de virulencia como biomarcadores específicos en el pronóstico de cáncer y otras enfermedades relacionadas con la infección por *H. pylori* (Bridge & Merrell, 2013).

iii. Citotoxina vacuolizante VacA

La citotoxina vacuolizante (VacA) es uno de los principales factores de virulencia de *H. pylori* y es la probablemente la toxina de *H. pylori* más estudiada. Recibió su nombre gracias a su capacidad para inducir la formación de vacuolas en células eucariotas (Chang et al., 2018), sin embargo, es una toxina multifuncional. VacA ayuda a *H. pylori* a colonizar el estómago ejerciendo múltiples efectos sobre las células epiteliales del hospedero humano y además de llegar a las células del epitelio gástrico también causa diferentes efectos en las células del sistema inmunológico (Chauhan et al., 2019).

VacA es una proteína formadora de poro de 140kDa en su forma inmadura y 88kDa en su forma madura, la cual es posteriormente clivada para la formación de dos fragmentos de 33kDa y 55kDa denominados p33 y p55, respectivamente (Kim & Blanke, 2012). El principal efecto de VacA está relacionado con su capacidad de internalizarse en las células epiteliales. Se ha sugerido que *H. pylori* secreta VacA cerca de la membrana apical de las células, donde se une a la esfingomielinina presente en los receptores de la superficie celular como RPTP β , RPTP α , para posteriormente internalizarse en la célula. Posteriormente, VacA ejerce efectos multifuncionales sobre los orgánulos celulares. VacA juega un papel vital en la supervivencia de *H. pylori* al inducir un flujo de iones y nutrientes desde la mucosa hacia la luz del estómago, lo que da como resultado la alteración de la integridad de las células del epitelio gástrico, la inducción de apoptosis y la formación de grandes vacuolas intracelulares (Chauhan et al., 2019; Cover & Blanke, 2005).

Todas las cepas de *H. pylori* llevan el gen *vacA* que codifica la producción de la toxina VacA; sin embargo, es un gen altamente heterogéneo y se presentan varios genotipos determinados por las combinaciones alélicas de las regiones que lo componen. La diferencia en las regiones del gen *vacA* está relacionada con la capacidad de vacuolización que tendrá la toxina al ingresar en la célula hospedera; las principales variantes alélicas que se conocen son en la secuencia señal (*s1a*, *s1b*, *s1c* y *s2*), la región media (*m1* y *m2*) y la región intermedia (*i1* e *i2*) (Ferreira et al., 2012). Para el genotipo *s1/m1*, la expresión de VacA es muy activa y puede dañar las células de forma más aguda. Se ha demostrado que este genotipo se asocia con altos niveles de inflamación en la mucosa gástrica y un mayor riesgo de atrofia gástrica

y carcinoma, en comparación con las cepas *vacA s2* y *m2* (González et al., 2011; Kao et al., 2016). El genotipo *vacA i1* está fuertemente asociado con cepas *vacAs1/m1* y *cagA* positivas, mientras que el genotipo *vacA i2* está estrechamente asociado con *vacA s2*, *m2*, y *cagA* genotipos negativos (Ferreira et al., 2012). Al parecer, los subtipos *vacA s1a* y *s1c* son los menos frecuentes en pacientes con cáncer gástrico (Lin et al., 2004). A pesar de que diversos estudios han reportado asociaciones entre los genotipos de *vacA* y la enfermedad, esta información no es consistente entre diferentes países.

iv. El gen inducido por contacto con el epitelio iceA

Descrito por primera vez en 1998 y del cual se conocen dos variantes alélicas *iceA1* e *iceA2* (Chomvarin et al., 2008), aunque aún se desconoce la función exacta de este gen, se sabe que los niveles de IceA aumentan significativamente durante la unión de *H. pylori* con las células epiteliales gástricas (Baj et al., 2020). Algunos estudios han informado que no existe relación entre la expresión de IceA y la presencia de otros factores de virulencia como CagA o VacA; sin embargo, estos datos son contradictorios entre autores (Boyanova et al., 2010; Shiota et al., 2012). Se ha demostrado que la expresión de *iceA*, conduce a un peor resultado clínico de los pacientes infectados. La infección por cepas que portan el gen *iceA1* se ha asociado con el desarrollo de úlcera péptica y aumento de los niveles mucosos de IL-8. Un metanálisis mostró que la presencia del gen *iceA1* se asocia con aumento en el riesgo de desarrollar úlcera péptica (OR = 1,28) en los países occidentales (Shiota et al., 2012). Además, las cepas *iceA1* son más propensas a inducir daño oxidativo en el ADN, lo que puede considerarse un marcador de la progresión o un factor de riesgo relacionado con el desarrollo de cáncer gástrico (Ladeira et al., 2008; Ladeira et al., 2004). Por el contrario, la expresión de *iceA2* se asocia más a menudo con dispepsia ulcerosa no péptica (Feliciano et al., 2015). Se sabe también que el estatus del gen *iceA* difiere significativamente según áreas geográficas particulares (Yamaoka et al., 1999).

3.4 Tratamiento de erradicación de *H. pylori* y resistencia antimicrobiana.

Desde el 2013, la IARC reconoció la evidencia que soporta que la erradicación de *H. pylori* es la mejor estrategia para prevenir el cáncer gástrico, y recomendó la inclusión de esta estrategia en programas nacionales para el control del cáncer (Group, 2014). Teniendo en

cuenta que la infección por *H. pylori* es adquirida principalmente en la infancia, se ha sugerido que la erradicación temprana de este microorganismo disminuye el riesgo de desarrollar patologías asociadas (Malfertheiner et al., 2017).

Hasta ahora el tratamiento que se ha utilizado para la erradicación de este microorganismo alrededor del mundo es la terapia basada en antibióticos (O'Connor et al., 2009), la cual está recomendada para los pacientes con úlceras gástricas, gastritis atrófica, linfomas de tipo MALT de bajo grado, anemias inexplicadas por deficiencia de hierro y púrpura trombocitopénica idiopática (Acosta et al., 2014). No existe un régimen de tratamiento universalmente aceptado, aunque todos los que existen tienen como objetivo la reversión de la sintomatología y la curación de la mucosa dañada por el proceso de infección (Safavi et al., 2016). Algunos antibióticos como Amoxicilina (AMX), Claritromicina (CLA), Metronidazol (MTZ), Levofloxacina (LEV) y Tetraciclina (TET) se han usado más comúnmente, pero también se han empleado otros como, Doxiciclina, Furazolidona, Rifabutina entre otros (Otero R et al., 2018).

Estos esquemas de erradicación de *H. pylori* consisten en la combinación de un inhibidor de la bomba de protones (IBP) con los antibióticos seleccionados. Dependiendo del tipo de combinación, las terapias de erradicación se clasifican como duales, triple estándar, triples con quinolonas, secuenciales, concomitantes, híbridas y cuádruples con bismuto y misceláneas (Graham et al., 2014). En la actualidad, las tasas de erradicación con estos esquemas han disminuido alcanzando cifras de hasta el 60% en algunos países, y se correlacionan inversamente con las tasas de resistencia, las cuales alcanzan niveles cada vez más altos (Mascellino et al., 2017; Thung et al., 2016).

Debido al aumento de la infección por *H. pylori* resistente a los antimicrobianos, las directrices recientes para la erradicación de *H. pylori* del “Maastricht-V Consensus” no recomiendan esta terapia en regiones con alta prevalencia de resistencia (Malfertheiner et al., 2017). La evidencia muestra que en regiones donde la resistencia a los antibióticos es superior al 15%, la tasa de éxito se reduce hasta en un 35%. Niveles por encima de este rango se han informado en algunos países europeos, asiáticos y de América Latina (Camargo et al., 2014) mientras que los resultados obtenidos en población colombiana muestran que la resistencia a

CLA es de 17,7% (Trespalacios et al., 2010) y 27.3% para LEV (Trespalacios-Rangél et al., 2016).

Se sabe que las tasas de curación han sido constantemente bajas durante los últimos 15 años, sin embargo, las recomendaciones de tratamiento no han cambiado radicalmente (Chey et al., 2017). Según los reportes recientes, las causas principales de falla terapéutica son la resistencia de *H. pylori* a los antibióticos (principalmente CLA, MTZ y LEV) y la no adherencia al tratamiento (Malfertheiner et al., 2012). Este fenómeno de falla terapéutica se ha visto acrecentado porque estos esquemas corresponden a tratamientos empíricos y no guiados por reportes de resistencia individual; adicionalmente en muchos países no existen guías locales de tratamiento basadas en la epidemiología y las tasas de resistencia locales, sino que se utilizan protocolos globales que pueden pasar por alto situaciones particulares (Arama et al., 2016; Savoldi et al., 2018).

3.4.1 Mecanismos moleculares involucrados en la resistencia a los antimicrobianos.

Los antibióticos empleados en la erradicación de *H. pylori* funcionan a través de diferentes mecanismos. Algunos bloquean la síntesis de ácidos nucleicos, como las quinolonas (LEV), rifampicina y nitroimidazoles (MTZ); otros bloquean la traducción de proteínas bacterianas mediante la unión a las subunidades ribosomales, como los macrólidos (CLA) y las tetraciclinas (TET); y algunos influyen en la síntesis de la pared celular bacteriana, como los β -lactámicos (AMX) (Mégraud et al., 2015)

La resistencia a los antimicrobianos en *H. pylori* es debida principalmente a cambios mutacionales (sin sentido, cambio de marco, inserciones o deleciones) en las secuencias genómicas bacterianas, las cuales alteran la diana del fármaco o inhiben la activación del fármaco dentro de las células bacterianas (Tshibangu-Kabamba & Yamaoka, 2021). Los estudios sugieren que, a diferencia de otras bacterias, estas mutaciones están codificadas cromosómicamente y no en elementos extracromosómicos como plásmidos (Mégraud et al., 2015; Mégraud & Lehours, 2007). El vínculo de causalidad o correlación entre las mutaciones genéticas y el fenotipo de resistencia se ha establecido tanto experimentalmente mediante el uso de mutantes resistentes generados en el laboratorio, como estadísticamente mediante la investigación de aislamientos clínicos con fenotipos resistentes y susceptibles. Sin embargo, se han informado muchas mutaciones en unos pocos aislamientos

resistentes que aún requieren exploración para establecer su relevancia biológica (Tshibangu-Kabamba & Yamaoka, 2021).

Además de las mutaciones, se han descrito mecanismos de resistencia adicionales, pero con una relevancia clínica limitada o hipotética (Tshibangu-Kabamba & Yamaoka, 2021). Dentro de estos mecanismos, se han descrito, por ejemplo, la permeabilidad reducida de la membrana celular, dada por la familia de proteínas de membrana externa, que cuando se sobre expresan forman una barrera natural para reducir la acumulación intracelular de compuestos antibióticos (Smiley et al., 2013); la producción de biopelículas, que para *H. pylori* al parecer, aumenta la resistencia a CLA, aunque el mecanismo biológico por el cual lo hacen aún sigue sin dilucidarse (Yonezawa et al., 2013); las proteínas de eflujo que pueden contribuir a la pérdida de los antibióticos y que para *H. pylori* al parecer median en la resistencia a AMX, MTZ, CLA y TET (Iwamoto et al., 2014; Qureshi et al., 2014). Sin embargo, de la gran gama de proteínas transportadoras que tiene *H. pylori* solo unas pocas han sido relacionadas con RAM, por lo tanto, aún se requiere mayor investigación al respecto (Gong & Yuan, 2018). También se han propuesto otros mecanismos de resistencia, como la expresión de enzimas tipo β -lactamasas, que podrían inactivar compuestos antibióticos como la AMX, o la expresión de factores de virulencia, como CagA, VacA y DupA, que podrían afectar la actividad antibiótica promoviendo la liberación de factores inflamatorios y provocando mutaciones en genes relacionados con la resistencia bacteriana (Gong & Yuan, 2018); o la transformación a formas cocoides, que de acuerdo con algunos estudios, persisten en la mucosa gástrica durante los tratamientos de erradicación de *H. pylori* y pueden restaurar el crecimiento de 2 a 4 semanas después, llevando al fracaso de la erradicación (Thung et al., 2016). A pesar de que se han relacionado muchos y muy diversos mecanismos asociados a la RAM en *H. pylori*, solamente los cambios mutacionales han tenido un análisis amplio que ha demostrado un mecanismo biológico asociado. Por esta razón este documento solo se centrará en la descripción y el análisis de este mecanismo.

i. Mutaciones en el dominio V del gen que codifica el ARNr 23S

El ARN ribosómico (ARNr) funciona como un andamio para el ARN mensajero (ARNm) y promueve la síntesis de la cadena peptídica. Los ARNr de procariotas incluyen 5S, 16S y 23S. El ARNr 23S forma parte de la subunidad grande 50S de los ribosomas y su dominio V

Antecedentes y Marco Teórico

tiene actividad peptídica aciltransferasa. Los antibióticos de tipo macrólido como CLA se unen al dominio V del ARNr 23S e inhiben la síntesis de proteínas bacterianas. Se ha demostrado que mutaciones puntuales dentro del dominio V del gen *23S rRNA* pueden alterar la conformación del bucle de la peptidil transferasa e inhibir la unión entre el CLA y el ARNr 23S, reduciendo su eficacia y dando lugar a un fenotipo de resistencia (Kim et al., 2002; Marques et al., 2020).

Dos copias del operón de ARNr 23S están presentes en el genoma de *H. pylori* y, para la mayoría de las cepas, las mutaciones generalmente se encuentran en ambas copias; sin embargo, un fenotipo heterocigoto es suficiente para conferir una resistencia intermedia a CLA. Las cepas resistentes se pueden dividir en dos grupos: un alto nivel de resistencia ($\text{CMI} > 64 \text{ mg l}^{-1}$) y un nivel bajo de resistencia ($0.5 \leq \text{CMI} \leq 1 \text{ mg l}^{-1}$) (Marques et al., 2020).

Tres mutaciones puntuales en el dominio V del *ARNr23S* son responsables de $> 90\%$ de las cepas resistentes a CLA. Ocurren en dos posiciones específicas de nucleótidos adyacentes, una transición de adenina a guanina en la posición 2142 (A2142G) o 2143 (A2143G) o, con menos frecuencia, una adenina a transversión de citosina en la posición 2142 (A2142C) (Mégraud, 2004; Trespalacios et al., 2013). Diversos estudios han propuesto otras mutaciones puntuales simultáneamente en el mismo gen, pero su papel aún es controvertido. De acuerdo con varios autores, es poco probable que estas otras mutaciones contribuyan a la resistencia (Marques et al., 2020).

ii. Mutaciones en el gen que codifica el ARNr 16S

El ARNr 16S es parte de la subunidad pequeña ribosómica procariota 30S y tiene una función similar a la del ARNr 23S. Antibióticos como TET actúan sobre la subunidad 30S inhibiendo la unión del tARN lo cual impide la adición de aminoácidos en la síntesis proteica. Sin embargo, mutaciones en el gen *ARNr 16S* pueden disminuir la capacidad de unión de TET al ribosoma, causando resistencia (Saranathan et al., 2020). Las mutaciones asociadas se encuentran predominantemente en las posiciones 926–928 (Gong & Yuan, 2018); sin embargo, la mutación más comúnmente asociada es la A926G (Saranathan et al., 2020).

iii. Mutaciones en los genes de la ADN girasa

La ADN girasa, codificada por los genes *gyrA* y *gyrB*, es una enzima necesaria para mantener la estructura de la hélice del ADN, desempeñando un papel en la replicación, recombinación y transcripción del ADN (Gong & Yuan, 2018). Algunos medicamentos como LEV pueden inhibir la ADN girasa bacteriana, lo que da como resultado un daño irreversible del ADN bacteriano. Las mutaciones en los genes *gyrA* y *gyrB* pueden evitar la unión de los antibióticos a las enzimas, lo cual es la principal razón de la resistencia a las quinolonas en *H. pylori*. En el gen *gyrA* dos puntos de acceso que confieren resistencia a quinolonas se han descubierto en esta región: 87 (Asn a Lys, Tyr, Ile) y 91 (Asp a Gly, Asn, Tyr) (Lee et al., 2011; Sanches et al., 2016; Secka et al., 2013; Trespalacios et al., 2015; Trespalacios-Rangél et al., 2016). La resistencia a las quinolonas de bajo nivel puede estar asociada con una mutación puntual en la posición 91, mientras que una mutación en la posición 87 puede resultar en un alto nivel de resistencia (Murakami et al., 2009). Del mismo modo, mutaciones en el gen *gyrB* están implicadas en la resistencia a quinolonas, aunque generalmente coexisten con mutaciones en *gyrA* y la incidencia de mutaciones en *gyrB* es significativamente menor que la de mutaciones en *gyrA* (Miftahussurur et al., 2016). Las mutaciones más conocidas en el gen *gyrB* que puede conducir a resistencia incluyen D481E, R484K, D484K, F438S y E463K (Gong & Yuan, 2018).

iv. Mutaciones en genes que codifican enzimas redox

Algunos antibióticos como MTZ, un antibiótico de la familia de los imidazoles, son metabolizados por nitroreductasas y en su degradación se producen compuestos altamente oxidantes como O₂- los cuales son capaces de causar daño en el ADN bacteriano. En *H. pylori*, los principales genes que codifican enzimas redox son *rdxA* (que codifica una nitroreductasa NADPH insensible al oxígeno) y *frxA* (que codifica una NADPH: flavin oxidoreductasa). Las mutaciones de desplazamiento, inserción y delección en estos genes que codifican enzimas redox pueden causar pérdida de actividad enzimática y han sido el mecanismo principal atribuido a la resistencia frente MTZ para *H. pylori* (Gong & Yuan,

2018). Muchas mutaciones puntuales han sido descritas para los genes *rdxA* y *frxA*, pero los sitios de estas mutaciones no se conservan en todas las regiones geográficas. Además, algunos autores han informado que la mayoría de las cepas resistentes al metronidazol tienen mutaciones distribuidas en el gen *rdxA* que probablemente no contribuyan a su fenotipo resistente, sino que reflejan una diversidad de secuencias bien reconocida de este organismo (Secka et al., 2013).

v. *Mutaciones en genes asociados con la síntesis de la pared celular*

Las proteínas de unión a penicilina (PBP) son un grupo de proteínas con actividad glicosiltransferasa y aciltransferasa involucradas en las etapas finales de la síntesis de peptidoglicano, un componente principal de las paredes celulares bacterianas. En general, la unión covalente de antibióticos y PBP bloquea la síntesis de peptidoglicanos e inhibe la formación de la pared celular. En *H. pylori* se ha descrito que mutaciones en los genes que codifican estas PBP, específicamente en *pbp1A*, son el mecanismo más común de resistencia a AMX. Estas mutaciones codifican cambios en los motivos conservados de PBP, lo que provoca una afinidad reducida del antibiótico por PBP1 (Gerrits et al., 2006). Los principales cambios mutacionales descritos en PBP1 son N107R, A201V, V250I y S543T (Saranathan et al., 2020). Mutaciones en PBP2 y PBP3 también podrían estar relacionadas con la resistencia a AMX en *H. pylori* al facilitar el efecto de las mutaciones en PBP1; se ha sugerido que una combinación de mutaciones en PBP1, PBP2 y PBP3 pueden contribuir en conjunto a una alta resistencia a AMX (Gong & Yuan, 2018). En PBP2 se ha descrito la mutación I259T, y en PBP3 los cambios D2N, A50S, F490Y, A541T y V374I (Saranathan et al., 2020).

3.5 Caracterización de bacteriófagos de *H. pylori*

Este tema se amplía en el artículo de revisión (Anexo 1):

Muñoz, A. B., Stepanian, J., Trespalacios, A. A., & Vale, F. F. (2020). Bacteriophages of *Helicobacter pylori*. *Frontiers in microbiology*, *11*, 549084. <https://doi.org/10.3389/fmicb.2020.549084>

Los bacteriófagos son partículas virales capaces de infectar bacterias. Pueden ser líticos, lisogénicos o pseudo-lisogénicos. Los fagos líticos reconocen las superficies bacterianas e inyectan sus ácidos nucleicos en la célula huésped; luego, se ensamblan, se multiplican y finalmente rompen la célula para liberar la progenie del fago que infecta nuevas células bacterianas (Sulakvelidze, 2005). En los fagos lisogénicos, el ADN del fago se integra en el genoma bacteriano, formando profagos (también conocidos como fagos templados). Los profagos promueven la evolución bacteriana a través de la transferencia horizontal de genes e inducen la transducción de varios genes involucrados en los comportamientos biológicos de la bacteria (Brüssow et al., 2004; Paterson et al., 2010; Touchon et al., 2016). Los fagos pseudo-lisogénicos están presentes como episomas (es decir, el material genético no está integrado en el genoma bacteriano) y posponen la lisis celular en huéspedes con agotamiento de nutrientes (Uchiyama et al., 2013); tanto los fagos líticos como lisogénicos, bajo ciertas condiciones, pueden adquirir formas episomales (Łoś & Węgrzyn, 2012).

El descubrimiento de los profagos de *H. pylori* ha sido de particular importancia para explicar la tremenda diversidad genética de esta bacteria.

3.5.1 Profagos de *H. pylori*

Las primeras observaciones de partículas intracelulares similares a fagos en preparaciones de *H. pylori* se realizaron poco después del descubrimiento de esta bacteria (Goodwin & Worsley, 1993; Marshall et al., 1987). A principios de la década de 1990, se describió una cepa lisogénica de *H. pylori* que producía partículas de fagos de forma espontánea (Schmid et al., 1990). Tres años más tarde, se reprodujo el ciclo lítico de este fago. La microscopía electrónica de tinción negativa reveló que las cabezas de los fagos de alrededor de 50 a 60 nm y la longitud del ADN se estimó en 22.000 pb (Heintschel von Heinegg et al., 1993).

Posterior a estos hallazgos, los avances en las tecnologías de secuenciación han permitido corroborar la presencia de profagos en el genoma de *H. pylori* y analizar sus funciones en este contexto. Los estudios informan que los profagos están presentes en alrededor del 20% de los aislados de *H. pylori* (Muñoz et al., 2020). El primer análisis genómico de profagos en *H. pylori*, corresponde al análisis del fago PhiHp33, el cual fue encontrado en la cepa *H. pylori* B45 que fue aislada de un paciente con linfoma tipo MALT (Lehours et al., 2011). Gracias a este estudio, posteriormente se logró el análisis de secuencias de dos genes utilizados como marcadores de presencia de profagos en las cepas de *H. pylori*, integrasa y holin. El gen de la integrasa es responsable de la integración del genoma del fago en el cromosoma bacteriano. Por lo general, se coloca en el extremo 5', mientras que el gen holin participa en la lisis celular cuando se produce un ciclo lítico y generalmente se coloca en el extremo 3' (Brüssow et al., 2004). El análisis de estos genes demostró que las cepas de *H. pylori* de diferentes regiones portan secuencias de profago completas o parciales; así mismo, permitió diferenciar cepas según su origen geográfico, generando resultados consistentes con la clasificación basada en tipificación de secuencias multilocus (MLST) (Lehours et al., 2011; Secka et al., 2017; Vale et al., 2015).

En 2012, se informó del aislamiento de un fago (1961P) a partir de cepas taiwanesas de *H. pylori*. Este fago mostró características compatibles con la familia *Podoviridae*. Además, el análisis de secuencia identificó otros profagos similares integrados en los genomas de diferentes cepas de *H. pylori* que se habían secuenciado previamente (Luo et al., 2012). Además, en el mismo año, se informaron las secuencias genómicas completas de dos bacteriófagos de *H. pylori* (KHP30 y KHP40) aislados de pacientes japoneses. Estos profagos se obtuvieron después de la liberación espontánea de partículas virales por cepas de *H. pylori* (Uchiyama et al., 2012). Este estudio demostró que el fago KHP30 era un virus esférico con una envoltura lipídica. Además se demostró que estaba presente en las bacterias infectadas en forma de episoma (Uchiyama et al., 2013). El análisis de KHP30 ha continuado, analizando sus efectos sobre los comportamientos de la cepa hospedera y han demostrado que la presencia de bacteriófagos induce alteraciones en el genoma de *H. pylori*, lo que lleva a la producción continua de derivados libres de profagos que coexisten en las comunidades microbianas de *H. pylori* y confieren diversidad en morfología, motilidad, viabilidad y patogenicidad. Además, este estudio informó que el profago KHP30 podría suprimir

específicamente la expresión de CagA (Takeuchi et al., 2018). Estudios relacionados han demostrado la presencia de fagos similares a KHP30 en otras cepas de *H. pylori*; insertados entre dos genes de virulencia putativos, *oipA* y *hombB*, los cuales codifican factores de adherencia. Los autores sugieren que la presencia de un profago insertado en esta región indica que HomB y OipA pueden actuar como receptores del fago (You et al., 2015).

3.5.2 Diversidad y filogenia de profagos de *H. pylori*

Además de la caracterización de los profagos de *H. pylori*, se ha discutido ampliamente la relación entre la presencia de profagos y la diversidad genética bacteriana. Como se ha mencionado previamente, la diversidad de *H. pylori* se ha asociado con el origen geográfico de las poblaciones (Matta et al., 2017), y al parecer, entre otros factores, está influenciada por la presencia de elementos genómicos móviles, como los profagos, (Vale & Lehours, 2018). La mayoría de las observaciones han mostrado una concordancia filogeográfica entre los genes de fagos y los bacterianos, lo que sugiere un modelo de coevolución entre el virus y su hospedero bacteriano. No obstante, algunos profagos se asignaron a poblaciones distintas de su huésped, mientras que otros exhibieron signos de recombinación entre poblaciones (Yahara et al., 2019).

La distribución filogeográfica de los profagos de *H. pylori* puede caracterizarse mediante un análisis de agrupamiento bayesiano, basado en un esquema de tipificación que incluye las secuencias de dos genes profagos: integrasa (que es responsable de la integración del genoma del fago en el cromosoma bacteriano) y holin (que participa en la lisis celular durante un ciclo lítico). Teniendo en cuenta esta metodología se han descrito cuatro poblaciones de profagos: hpAfrica1, hpEastAsia, hpNEurope y hpSWEurope (Vale et al., 2015), subdividiendo notablemente la población europea que hasta hace poco era solo caracterizada como hpEurope. Recientemente, analizando el genoma completo de *H. pylori*, también se observaron dos subpoblaciones de hpEurope (Thorell, Yahara, et al., 2017).

La información disponible sobre la diversidad de profagos de *H. pylori* indica que los profagos y las bacterias comparten una historia evolutiva compleja (Vale et al., 2017) y que el genoma bacteriano se ha modificado ampliamente, en diversas regiones, a través de la transferencia horizontal de genes (Takeuchi et al., 2018). Por otra parte, análisis *in-silico* han

Antecedentes y Marco Teórico

revelado cierta asociación entre la presencia de genes ortólogos de profagos y genes de virulencia, específicamente *cagA* o *vacA* (Kyriillos et al., 2016). Los informes sugieren que los profagos insertados en el genoma de *H. pylori* representan elementos esenciales para la adaptación de esta bacteria a ambientes hostiles, porque las ventajas metabólicas o los mecanismos de tolerancia que pueden ser influenciados por los profagos pueden mejorar la competitividad de *H. pylori*. Sin embargo, hasta el momento no se han reconocido funciones específicas de profago en *H. pylori*, y rara vez se han registrado las funciones de profago en la enfermedad (Vale et al., 2017).

En general se sabe que los fagos participan activamente en los procesos de recombinación e intercambio genético horizontal. En este contexto, un estudio reciente presentó un análisis cuantitativo de los eventos de recombinación descritos de los genes de unos de los profagos descritos de *H. pylori*, revelando la frecuente recombinación entre los genes centrales del fago previamente reportados como menos propensos a la recombinación. Esta observación parece reflejar las relaciones co-evolutivas de estos fagos con sus huéspedes, revelando la carrera armamentista que existe, ya sea para contribuir al escape de los fagos de la inmunidad bacteriana o para proteger a las bacterias hospederas. La alta tasa de recombinación en los fagos de *H. pylori* sugiere que se encuentran entre los fagos más recombinogénicos del mundo (Yahara et al., 2019).

A pesar del conocimiento que ahora existe sobre algunas características biológicas y genómicas de los profagos de *H. pylori*, aún existe mucho por investigar acerca de la relación que existe entre profagos y bacterias hospederas, especialmente para establecer si los profagos pueden beneficiar a los patógenos de formas que aún no han logrado determinarse.

4. CAPITULO 2.

CARACTERIZACIÓN DE FACTORES DE VIRULENCIA Y GENES RELACIONADOS CON RESISTENCIA A LOS ANTIMICROBIANOS EN GENOMAS COLOMBIANOS DE *H. pylori*

4.1 INTRODUCCIÓN

H. pylori coloniza el estómago de aproximadamente 4.4 billones de personas en el mundo (Hooi et al., 2017), alcanzando prevalencias más altas en los países en vía de desarrollo. En Colombia se han reportado tasas de infección por *H. pylori* de hasta el 80% para población adulta (Camargo et al., 2014; Gómez et al., 2015). La presencia de diferentes genotipos en diferentes regiones geográficas ha sido ampliamente demostrada (Thorell, Yahara, et al., 2017) y la severidad de las enfermedades gástricas asociadas con la infección por *H. pylori* se ha relacionado con el genotipo predominante en cada región (Qumar et al., 2021).

Muchos aspectos pueden ser responsables de la progresión de la infección por *H. pylori* hacia enfermedades severas. Dentro de estos, la presencia de factores de virulencia bacterianos ha sido especialmente asociada con la malignidad y el mayor riesgo de progresión de la enfermedad (Miftahussurur et al., 2017). Entre otros, los factores de virulencia más frecuentemente asociados son la oncoproteína CagA, la citotoxina vacuolizante A (vacA), la adhesina fijadora de antígenos del grupo sanguíneo (BabA) y el gen inducido por contacto con el epitelio (IceA) (Baj et al., 2020; Denic et al., 2020).

Además de la virulencia, otro factor que impone desafíos cada vez más grandes en el manejo de la infección por *H. pylori* es la aparición de farmacorresistencia, un fenómeno que cada vez más resulta en la presentación de fracasos del tratamiento (Camargo et al., 2014; Talebi Bezmin Abadi, 2017). En Colombia, las tasas reportadas de resistencia a antibióticos son preocupantes: la resistencia primaria a CLA, LEV y MTZ ha sobre pasado los límites aceptados, presentando tasas de resistencia de hasta 17.7%, 27.3% y 83%, respectivamente

(Talebi Bezmin Abadi, 2017; Trespalacios et al., 2013; Trespalacios-Rangél et al., 2016). A pesar de ello, el tratamiento empírico combinando con estos antibióticos continúa siendo una práctica común.

Se sabe que *H. pylori* desarrolla resistencia a los antibióticos, principalmente, mediante mutaciones puntuales en los genes que codifican para el sitio de acción del antibiótico; analizar estos genes en los genomas completos de *H. pylori* reportados en Colombia, aportará conocimiento sobre el genotipo de resistencia que ha estado circulando en el país a través del tiempo y aportará información para la vigilancia del fenómeno de la resistencia a los antimicrobianos en Colombia.

Considerando la problemática de la infección por *H. pylori* en Colombia, el objetivo de este capítulo fue caracterizar los principales factores de virulencia y analizar las mutaciones relacionadas con resistencia a los antimicrobianos. Para abordar este objetivo se utilizaron herramientas de epidemiología molecular que permitieron analizar los genomas de 29 aislamientos colombianos de *H. pylori* pertenecientes a la colección de Microorganismos de la Pontificia Universidad Javeriana; así como los genomas colombianos de *H. pylori* disponibles en bases de datos públicas.

4.2 METODOLOGÍA

4.2.1 Obtención de los genomas

Esta metodología se describe a profundidad en la publicación (Anexo 2):

Muñoz AB, Stepanian J, Acosta C, Solano-Gutierrez JS, Vale FF, Trespalacios- Rangél AA. 2021. Draft genome sequences of 29 *Helicobacter pylori* strains isolated from Colombia. *Microbiol Resour Announc* 10:e00218-21. <https://doi.org/10.1128/MRA.00218-21>

Se seleccionaron 29 cepas de *H. pylori* que se encontraban depositadas en la colección de Microorganismos de la Pontificia Universidad Javeriana y que habían sido aisladas por el grupo de investigación en Enfermedades Infecciosas, entre los años 2009 y 2010, a partir de pacientes residentes en Bogotá, Colombia. Las biopsias a partir de las cuales se obtuvieron las cepas fueron recolectadas posterior a la firma de consentimiento informado por parte de

los pacientes. La selección de las cepas se realizó teniendo en cuenta un muestreo por conveniencia, se consideraron características previamente conocidas, tales como caracterización fenotípica de la susceptibilidad antibiótica, presencia de genes codificantes de genes de virulencia y presencia de genes predictores de profagos.

Los aislamientos bacterianos se recuperaron a partir del banco de cepas del grupo de investigación y se cultivaron en Agar Brucella BBLTM (Becton Dickinson), suplementado con sangre de caballo al 7%, IsoVitalexTM al 0,4% (BD, EE. UU.) y suplemento selectivo DENT al 0,2% (OxoidTM, Reino Unido). Los cultivos fueron incubados en condiciones microaerófilas (11% CO₂) a 37° C durante 4 a 7 días. A partir del crecimiento bacteriano, se extrajo el ADN total utilizando un kit DNeasy Blood & Tissue (QIAGEN, Hilden, Alemania) y siguiendo las instrucciones del fabricante. La cuantificación del ADN se realizó por ensayo fluorimétrico utilizando un fluorómetro Qubit 2.0 y el kit Qubit dsDNA HS (Life Technologies, Carlsbad, California, EE. UU).

El ADN genómico se secuenció usando la plataforma Illumina MiSeq (Illumina, San Diego, CA); Las bibliotecas de ADN se prepararon utilizando el kit ADN Nextera XT (Illumina), seguido de corridas de 2 × 300 paired-end con una cobertura de ~80X. Las secuencias de baja calidad se eliminaron con el paquete de software Trimmomatic versión 0.39 (Bolger et al., 2014). Las lecturas fueron usadas para realizar un ensamblaje *de novo* utilizando SPAdes v13.3 (Bankevich et al., 2012). Las secuencias se anotaron utilizando el canal de anotación automática de genomas procarióticos del NCBI (PGAAP) (Tatusova et al., 2016). Se utilizaron los parámetros predeterminados para todas las herramientas de software. Los genomas fueron depositados en la base de datos GenBank del National Center for Biotechnology Information (NCBI) (Tabla 1).

Además de los genomas secuenciados, se realizó una búsqueda de los genomas colombianos depositados en la base de datos PATRIC (Pathosystems Resource Integration Center) 3.6.5 (Wattam et al., 2014). Para el momento de la búsqueda (Julio de 2020) se encontraron registrados 192 aislamientos (Anexo 4). Estos aislamientos fueron descargados de la base de datos whole-genome shotgun (WGS) del NCBI.

Caracterización de factores de virulencia y genes relacionados con resistencia a los antimicrobianos en genomas colombianos de H. pylori

Tabla 1.

Genomas colombianos secuenciados en este estudio.

Strain ID	GenBank Accession N°	CDS	Genome size (bp)	%GC	Contigs
COL 1-PUJ	JAFCHS000000000	1,582	1,679,429	38.8%	72
COL 2-PUJ	JACSDV000000000	1,632	1,607,581	39%	46
COL 5-PUJ	JACSDU000000000	1,667	1,624,361	39.1%	133
COL 6-PUJ	JACSDT000000000	1,673	1,625,175	39.1%	133
COL 8-PUJ	JAFCHT000000000	1,664	1,661,424	38.9%	43
COL 9-PUJ	JAFCHU000000000	1,653	1,613,788	39%	47
COL 10-PUJ	JAFCHV000000000	1,680	1,642,843	39.4%	58
COL 11-PUJ	JAFCHW000000000	1,663	1,626,191	39.4%	58
COL 12-PUJ	JAFCHX000000000	1,633	1,637,127	38.9%	43
COL 13-PUJ	JAFCHY000000000	1,714	1,658,899	39.3%	58
COL 14-PUJ	JAFCHZ000000000	1,701	1,673,807	38.9%	54
COL 15-PUJ	JAFCIA000000000	1,563	1,546,556	39.2%	46
COL 16-PUJ	JACSDS000000000	1,665	1,634,541	39.3%	64
COL 18-PUJ	JAF CIB000000000	1,662	1,658,604	38.9%	37
COL 19-PUJ	JAF CIC000000000	1,654	1,656,342	38.9%	42
COL 20-PUJ	JAF CID000000000	1,674	1,619,537	39%	56
COL 21-PUJ	JAF CIE000000000	1,670	1,618,229	39%	55
COL 23-PUJ	JACSDR000000000	1,661	1,614,131	39.1%	91
COL 24-PUJ	JAF CIF000000000	1,674	1,622,244	39%	119
COL 25-PUJ	JAF CIG000000000	1,681	1,662,282	38.9%	88
COL 26-PUJ	JAF CIH000000000	1,631	1,619,895	39%	52
COL 27-PUJ	JAF CII000000000	1,656	1,653,551	38.9%	36
COL 28-PUJ	JAF CIJ000000000	1,671	1,653,809	38.9%	23
COL 29-PUJ	JAF CIK000000000	1,585	1,586,826	39.2%	31
COL 30-PUJ	JAF CIL000000000	1,596	1,587,263	39.1%	35
COL 31-PUJ	JAF CIM000000000	1,795	1,711,739	39%	14
COL 49-PUJ	JAF CIN000000000	1,786	1,735,252	39%	96
COL 50-PUJ	JAF CIO000000000	1,726	1,711,312	38.9%	57
COL 51-PUJ	JAF CIP000000000	1,656	1,666,802	38.9%	39

4.2.2 Determinación *in silico* de variantes genéticas en los factores de virulencia

i. Gen cagA asociado a la citotoxina (cagA)

Se determinó ausencia/presencia del gen *cagA* en las secuencias seleccionadas para el estudio, utilizando una estrategia basada en el uso de cebadores previamente reportados (Rodríguez Gómez et al., 2020) (Tabla 2) y la herramienta BLASTn (Altschul et al., 1990). Para determinar el tamaño de los productos se realizó un script en Python realizando la resta de inicio a final del alineamiento. Para aquellas cepas positivas se confirmó el motivo EPIYA (Tabla 3) mediante BLASTp (Altschul et al., 1990) de los motivos contra cada una de las cepas positivas utilizando como *query* los motivos descritos (Tabla 3).

Tabla 2.

Cebadores usados para la determinación *in silico* de factores de virulencia.

Región	Cebador	Secuencia (5' - 3')	Producto	Referencia
<i>vacA</i> <i>m1</i>	Forward	GGTCAAAATGCGGTCATGG	290 pb (2741-3030)	(Atherton, 1998)
	Reverse	CCATTGGTACCTGRAGAAAC		
<i>vacA</i> <i>m2</i>	Forward	GGAFCCCCAGGAAACATTG	352pb (976 -1327)	(Atherton, 1998)
	Reverse	CATAACTAGCGCCTTGAC		
<i>vacA s1</i>	Forward	ATGGAAATACAACAAAACACAC	259pb (797 -1055)	(Atherton, 1998)
	Reverse	CTGCTTGAATGCGCAAAC		
<i>vacA s2</i>	Forward	ATGGAAATACAACAAAACACAC	286pb (284 -569)	(Atherton, 1998)
	Reverse	CTGCTTGAATGCGCAAAC		
<i>vacA</i> <i>s1a</i>	Forward	GTCAGCATCACACCGCAAC	190pb (866-1055)	(Erzin et al., 2006)
	Reverse	CTGCTTGAATGCGCAAAC		
<i>vacA</i> <i>s1b</i>	Forward	AGCGCCATACCGCAAGAG	187pb	(Erzin et al., 2006)
	Reverse	CTGCTTGAATGCGCAAAC		
	Forward	TTAGTTTCTCTCGCTTTAGTRGGGYT	220pb	

Caracterización de factores de virulencia y genes relacionados con resistencia a los antimicrobianos en genomas colombianos de H. pylori

<i>vacA s1c</i>	Reverse	CTGCTTGAATGCGCCAAAC		(Erzin et al., 2006)
<i>vacA i1</i>	Forward	GTTGGGATTGGGGGAATGCCG	426pb	(Rhead et al., 2007)
	Reverse	TTAATTTAACGCTGTTTGAAG		
<i>vacA i2</i>	Forward	GTTGGGATTGGGGGAATGCCG	432pb	(Rhead et al., 2007)
	Reverse	GATCAACGCTCTGATTTGA		
<i>iceA1</i>	Forward	GTG TTT TTA ACC AAA GTA TC	247pb	(Chomvarin et al., 2008)
	Reverse	CTA TAG CCA STY TCT TTG CA		
<i>iceA2</i>	Forward	GTT GGG TAT ATC ACA ATT TAT	229 pb	(Chomvarin et al., 2008)
	Reverse	TTT CCC TAT TTT CTA GTA GGT		
<i>cagA</i>	Forward	ACCCTAGTCGGTAATGGG	400 a 700 pb	(Rodríguez Gómez et al., 2020)
	Reverse	GCTTTAGCTTCTGAYACYGC		

Tabla 3.

Secuencia de aminoácidos de los motivos EPIYA de la proteína CagA

Motivo	Secuencia Aminoácidos	Referencia
EPIYA-A	EPIYAKVNKKKAGQ	(Rodríguez Gómez et al., 2020)
EPIYA-B	EPIY(A/T)QVAKKVNNAKI	
EPIYA-C	EPIYATIDDLGGP	
EPIYA-D	EPIYATIDFDEANQAG	

ii. *Citotoxina vacuolizante A (vacA)*

Se determinaron los alelos de las regiones *m*, *s* e *i* utilizando BLASTn (Altschul et al., 1990) con cebadores previamente reportados para cada una de las regiones (Atherton, 1998; Erzin et al., 2006; Rhead et al., 2007) (Tabla 2).

iii. *Gen inducido por contacto con el epitelio iceA*

Mediante BLASTn (Altschul et al., 1990) y los cebadores previamente reportados (Chomvarin et al., 2008; van Doorn et al., 1998) (Tabla 2) se determinaron los alelos *iceA1* e *iceA2*.

iv. *Adhesina fijadora de antígenos del grupo sanguíneo babA2*

Se realizó un alineamiento múltiple (ClustalW) mediante el software Bioedit 7.2 (Hall, 1999) utilizando la secuencia del alelo *babA2* del gen *babA*, reportada como referencia en Virulence Factor Database (VFDB) (Altschul et al., 1990). Se realizó una comparación de las 30 primeras bases con cada una de las cepas del estudio para determinar aquellas cepas positivas para el alelo *babA2*.

v. *Búsqueda de otros genes asociados virulencia*

Para investigar la ocurrencia de otros genes de virulencia en los genomas de *H. pylori* en estudio, se utilizó la base de datos de factores de virulencia de bacterias patógenas (VFDB) (<http://www.mgc.ac.cn/VFs/main.htm>). En esta base de datos, los 221 genomas fueron analizados para evaluar la presencia o ausencia de otros genes de virulencia, diferentes a los que se analizaron por otros métodos. Dentro de estos genes, se incluyeron 26 genes de la isla de patogenicidad *cag* (*cagPAI*), 7 genes relacionados con adherencia (*alpB/hopB*, *babA/hopS*, *babB/hopT*, *hpaA*, *hopZ*, *sabA/hopP*, *sabB/hopO* y *alpA/hopC*) y 2 genes relacionados con modulación de la respuesta inmune (*napA* y *oipA/hopH*).

4.2.3 Detección *in silico* de mutaciones en genes relacionados con resistencia a los antimicrobianos

Para determinar las mutaciones asociadas a resistencia a los antimicrobianos (RAM) en los genomas de *H. pylori* incluidos en el estudio, se evaluaron en algunos casos cambios a nivel de nucleótidos y en otros, a nivel de aminoácidos. En todos los casos, se usaron como plantilla los genes de la cepa de referencia *H. pylori* 26695 (GenBank: NC_000915.1). Los procedimientos específicos se describen a continuación.

i. Evaluación de cambios mutacionales relacionados con la resistencia a Claritromicina, Tetraciclina y Nitroimidazoles.

Utilizando el paquete de alineación de secuencias MUMmer (Kurtz et al., 2004) se evaluaron las mutaciones puntuales reportadas para los genes *ARNr23s*, *ARNr16s*, *rdxA* y *frxA* (Tabla 4). Las secuencias de referencia se alinearon a cada genoma estudiado utilizando el algoritmo NUCmer. Posteriormente, se utilizó *delta-filter* para determinar el mejor alineamiento de cada genoma y se buscaron los polimorfismos de nucleótido único (SNPs), para lo cual se usó el archivo de salida de *delta-filter* en la utilidad *show-snps*. También, se determinaron los polimorfismos que no se encuentran en regiones repetitivas para evitar SNPs con ambigüedad. Finalmente, se determinó la frecuencia de las mutaciones, inserciones y deleciones asociadas a RAM, las cuales habían sido previamente reportadas (Tabla 4).

Tabla 4.

Mutaciones asociadas a mecanismos de resistencia a los antimicrobianos.

Antibiótico	Gen	Mutación	Referencia
CLA	<i>23S rARN</i>	A2142G, A2143G, A2143C y A2141C	(Saranathan et al., 2020)
TET	<i>16S rARN</i>	A926G	(Saranathan et al., 2020)
LEV	<i>gyrA</i>	N87I, N87K, N87Y, D91G, D91N, D91Y, N97I, N97K	(Saranathan et al., 2020; Trespalcios-Rangél et al., 2016)
	<i>gyrB</i>	R484K, S479G	

Caracterización de factores de virulencia y genes relacionados con resistencia a los antimicrobianos en genomas colombianos de H. pylori

AMX	<i>pbp1</i>	N107R, A201V, V250I, S543T	(Saranathan et al., 2020)
	<i>pbp2</i>	I259T	
	<i>pbp3</i>	D2N, A50S, F490Y, A541T y V374I	
MTZ y otros imidazoles	<i>rdxA</i>	Aparición de secuencias parada, cambios en el marco de lectura, inserciones y deleciones. Mutaciones D59N, R131K, R90K, A118T, I160F, H97T.	(Acosta et al., 2017; Saranathan et al., 2020)
	<i>frxA</i>	Aparición de secuencias parada, cambios en el marco de lectura, inserciones y deleciones.	

CLA, claritromicina; TET, tetraciclina; LEV, levofloxacina; AMX, amoxicilina; MTZ, metronidazol.

ii. Evaluación de cambios mutacionales relacionados con la resistencia a Levofloxacina y Amoxicilina

Se identificaron cambios en las secuencias de aminoácidos codificada por los genes *gyrA*, *gyrB*, *pbp1*, *pbp2* y *pbp3*. Las secuencias de aminoácidos de la cepa de referencia se alinearon con las secuencias traducidas de los genes de cada cepa estudiada. Se utilizó la herramienta *promer* del paquete MUMmer (Kurtz et al., 2004) para realizar el alineamiento de las secuencias de aminoácidos, la herramienta *delta-filter* para determinar el mejor alineamiento y la utilidad *show-snps* para determinar las variantes en la secuencia proteica. Finalmente, se determinó la frecuencia de las mutaciones reportadas (Tabla 4).

4.2.4 Análisis estadístico

Los datos se ingresaron y analizaron en el software IBM Statistical Package for the Social Sciences (SPSS) versión 24.0. Se analizaron y determinaron las frecuencias de los genotipos de virulencia y las mutaciones ligadas a RAM presentes en los genomas analizados.

4.3 Resultados

4.3.1 Características de los genomas y datos demográficos de los pacientes a partir de los cuales se obtuvieron los aislamientos de H. pylori.

Un total de 221 genomas colombianos de *H. pylori* fueron obtenidos y analizados (29 nuevas secuencias obtenidas por este estudio y 192 genomas recuperados de bases de datos públicas).

A través del análisis de datos demográficos de los pacientes de los cuales se aislaron las 29 cepas procesadas en este estudio, se logró conocer que el promedio de edad para este grupo de pacientes fue de 48 años con un rango de 18 a 79 años; todos los pacientes de los cuales se obtuvieron estos aislamientos eran residentes en la ciudad de Bogotá o municipios cercanos; la mayoría de los pacientes fueron mujeres (20/29) y de acuerdo con el resultado histológico de las biopsias, el 65,5% de los pacientes fueron diagnosticados con gastritis crónica no atrófica y el 34,5% con gastritis crónica atrófica.

Respecto a los genomas disponibles en bases de datos públicas, para la fecha de búsqueda (Julio de 2020) se encontraron 192 registros de genomas de cepas colombianas de *H. pylori*. A partir de la información suministrada en las bases de datos para estos genomas se logró conocer que fueron obtenidos a partir de aislamientos recolectados entre los años 2000 y 2018; la mayoría (133/192) de estos aislamientos, es decir el 69,2% fueron recuperados de pacientes residentes en Bogotá, los demás fueron obtenidos de zonas geográficas como: Cartagena (2), Risaralda (1), Tolima (3), Tunja (10), Valle (1) y Nariño (29); no se encontró información relacionada con la zona geográfica en 13 genomas.

En cuanto a la patología asociada con los genomas obtenidos de bases de datos, no se encontró información o se presentaba solamente descrita como “enfermedad gástrica” en 122 genomas (63%). Para los 70 genomas restantes, se encontró que 34 genomas (17,7% del total) provenían de pacientes diagnosticados con gastritis crónica no atrófica, 4 genomas (2% del total) de pacientes con gastritis crónica atrófica, 27 genomas (14% del total) provenían de pacientes diagnosticados con lesiones premalignas como displasia (4 pacientes) o metaplasia intestinal (23 pacientes) y finalmente, 5 genomas (2,6% del total) provenían de pacientes quienes padecían cáncer gástrico.

Respecto a datos demográficos como edad o género, la mayoría de los genomas registrados en bases de datos (176/192), no contaban con datos relacionados. Para los 16 genomas con datos, se encontró que la edad de los pacientes estaba en un rango de 37 a 61 años y la mayoría (15/16) de estos pacientes fueron hombres.

La mayoría (111/221) de estos genomas provenían de aislamientos recolectados en el año 2000, para algunos años (2001-2004, 2008 y 2013-2016) no se reportó ningún genoma y para los años 2012 y 2018 solo se encontró un genoma reportado. De 8 genomas no se encontró información relacionada (Tabla 5).

En cuanto a las estadísticas de ensamblaje de los genomas, para los 29 genomas obtenidos en este estudio se encontró que el tamaño promedio de estos genomas fue de 1.641.708 pb, con un porcentaje promedio de Guanina – Citocina (GC%) de 39,03% y un número promedio de secuencias codificantes (CDS) de 1664. Las estadísticas completas de ensamblaje para estos genomas se proporcionan en la Tabla 1. En el caso de los 192 genomas colombianos obtenidos de bases de datos, se logró determinar que el promedio del tamaño del genoma fue de 1.662.383 pb, con un promedio de GC% de 38,9% y un promedio de CDS de 1624.

Tabla 5.

Año de obtención de los aislamientos de los cuales se obtuvieron los genomas.

Año	# Genomas reportados	%
2000	111	50,2
2005	12	5,4
2006	26	11,8
2007	7	3,2
2009	36	16,3
2010	6	2,7
2011	11	5,0
2012	1	0,5
2017	2	0,9
2018	1	0,5
ND	8	3,6
Total	221	100

ND, No hay datos disponibles.

4.3.2 Distribución de genotipos de virulencia en aislamientos colombianos de *H. pylori*

i. Determinación de la isla de patogenicidad *cag* (*cag*-PAI) y del gen *cagA*.

En la muestra analizada se encontró presencia del gen *cagA* en el 75,1% (166/221) de los genomas. En estos genomas *cagA* positivos, se encontró que el motivo EPIYA, correspondía al de tipo occidental (EPIYA-C) y al analizar las repeticiones de C, se encontró que el motivo más predominante en estos aislamientos fue ABC, el cual estaba presente en el 72,3% (120/166) de los aislamientos *cagA* positivos, seguido por ABCC que estaba presente en el 25,3% (42/166) de los genomas. No se encontró presencia del motivo ABCCC (Tabla 6). En la figura 1 se pueden apreciar los algunos de los genomas analizados con los diferentes motivos hallados.



Figura 1. Alineamiento de la secuencia de aminoácidos de la región 3' de *cagA* para cuatro genomas colombianos de *H. pylori*. Motivos EPIYA: A (azul), B (amarillo), C (violeta), dos C (rosa). La cepa de referencia de *H. pylori* 26695 (motivo EPIYA-ABC); genomas Col 9-PUJ (AB), 1081 y Col 11-PUJ (ABC) y 3120 (ABCC).

De los 42 genomas portadores del motivo EPIYA-ABCC solo se logró recopilar información clínica de solamente 12 de estos genomas, encontrando que 5 provenían de pacientes diagnosticados con gastritis, 6 con displasia o metaplasia y solo un genoma provenía de un paciente diagnosticado con cáncer, mientras que para los 120 genomas portadores del motivo EPIYA-ABC se encontró información clínica de solo 51 genomas, de los cuales, tres

provenían de pacientes con cáncer, 30 de pacientes con gastritis y 18 con displasia o metaplasia (Tabla 6).

Tabla 6.

Análisis de Motivos EPIYA en genomas *cagA* positivos

Motivo EPIYA	# Genomas	% Genomas	Patología asociada				
			GC-NA	GCA	D-M	CAG	ND
AB	4	2,4	3	1	0	0	0
ABC	120	72,3	26	4	18	3	69
ABCC	42	25,3	3	2	6	1	30

GC-NA, Gastritis crónica no atrófica; GCA, Gastritis Crónica Atrófica; D-M, Displasia o Metaplasia; CAG, Cáncer gástrico; ND, No Dato.

Los análisis realizados para los genes de *cagPAI* mostraron que de las cepas *cagA* positivas, solamente el 1,2% (2/166) tenían la isla de patogenicidad *cagPAI* intacta, es decir, con presencia de los 26 genes disponibles para análisis en la base de datos VFDB. Estos dos genomas, correspondían a genomas descargados desde bases de datos y en ningún caso se encontró información acerca de la patología asociada en los pacientes.

En los demás aislamientos *cagA* positivos, se encontró *cagPAI* pero con ausencia de 2 a 4 genes. De acuerdo con el número y tipo de genes presentes en cada genoma, el *cagPAI* se clasificó como intacto, semi-intacto, parcialmente deletado y negativo (Tabla 7). Los genes que más frecuentemente faltaron fueron *cagP*, *cagQ* y *cagI*, mientras que el gen *cag3* estuvo presente en todos los aislamientos positivos para *cagA* (Tabla 8).

Tabla 7.

Estructura del islote de patogenicidad *cagPAI* en los genomas analizados.

Estado de <i>cagPAI</i>	# de genomas	% genomas n=166
Islote de patogenicidad Intacto (26 genes disponibles para análisis VFDB)	2	1,2
Islote de patogenicidad Semi – Intacto (<26 genes con presencia de 17 genes esenciales)	41	24,7
Islote de patogenicidad Parcialmente deletado (25-14 genes, pero con ausencia de al menos 1 de los genes esenciales)	109	65,7
Islote <i>cagPAI</i> negativo (0 – 5 genes)	14	8,4
Total	166	100

Al analizar los 17 genes presentes en el islote de patogenicidad *cagPAI* que se han descrito como indispensables para la traslocación de *cagA*, se encontró que 41 genomas eran portadores de estos genes. De estos, 11 genomas contaban con información clínica de los pacientes de los cuales se aislaron, identificando que 8 pacientes padecían gastritis y 3 padecían metaplasia intestinal.

Tabla 8.

Positividad de genes componentes de *cagPAI*

Gen	# Genomas positivos	% Genomas positivos
<i>cag1</i>	47	21,3
<i>cag2</i>	132	59,7
* <i>cag3</i>	172	77,8
* <i>cag4</i>	128	57,9
* <i>cag5</i>	171	77,4
* <i>cagC</i>	98	44,3
<i>cagD</i>	165	74,7
* <i>cagE</i>	168	76,0
<i>cagF</i>	166	75,1
* <i>cagG</i>	128	57,9
<i>cagH</i>	168	76,0
* <i>cagI</i>	169	76,5
* <i>cagL</i>	166	75,1
* <i>cagM</i>	164	74,2
* <i>cagN</i>	169	76,5
<i>cagP</i>	9	4,1
<i>cagQ</i>	16	7,2
<i>cagS</i>	167	75,6
* <i>cagT</i>	166	75,1
<i>cagU</i>	167	75,6
* <i>cagV</i>	166	75,1
* <i>cagW</i>	169	76,5
* <i>cagX</i>	168	76,0
* <i>cagY</i>	93	42,1
* <i>cagZ</i>	165	74,7
* <i>virB11</i>	169	76,5
<i>cagA</i>	166	75,1

*Genes esenciales para la translocación de la proteína *cagA*

ii. Caracterización del gen *vacA*

Para la genotipificación del gen *vacA* se analizaron las regiones *s*, *m*, *i*, evaluando los alelos *s1* (subtipos *s1a*, *s1b*, *s1c*), *s2*; *m1* y *m2*; e *i1* e *i2*. Se encontraron 17 diferentes combinaciones, siendo *vacAm1i1s1b* y *vacAm1i1s1c*, las combinaciones más frecuentes, presentes en el 40,7% y el 23,9% de los genomas, respectivamente (Tabla 9).

Tabla 9.

Combinaciones alélicas del gen *vacA* en los genomas analizados.

Genotipo	# Genomas	% Genomas n=221
<i>m1 i1 s1</i>	2	0,90
<i>m1 i1 s1a</i>	7	3,17
<i>m1 i1 s1b</i>	90	40,72
<i>m1 i1 s1c</i>	53	23,98
<i>m1 i1 s2</i>	8	3,62
<i>m1 i2 s1</i>	1	0,45
<i>m1 i2 s1b</i>	4	1,81
<i>m1 i2 s1c</i>	2	0,90
<i>m2 i1 s1</i>	2	0,90
<i>m2 i1 s2</i>	2	0,90
<i>m2 i1 s1a</i>	1	0,45
<i>m2 i1 s1b</i>	1	0,45
<i>m2 i1 s1c</i>	7	3,17
<i>m2 i2 s1</i>	7	3,17
<i>m2 i2 s1b</i>	4	1,81
<i>m2 i2 s1c</i>	24	10,86
<i>m2 i2 s2</i>	6	2,71
Total	221	100,00

Al analizar la presencia de cada uno de los alelos, para la región *s* se encontró que la mayoría (92,8%) de los genomas portaba de los aislamientos analizados portaban el alelo *s1*; en la región *m* el alelo más frecuente fue *m1*, presente en un 75,6% de los aislamientos analizados; y para la región *i*, se encontró que un 76,9% portaban el alelo *i1* (Tabla 10).

Tabla 10.

Frecuencia de alelos en las regiones analizadas del gen *vacA*

Región	Alelo	# Genomas	% Genomas
s	<i>s1</i>	12	5,4
	<i>s1a</i>	8	3,6
	<i>s1b</i>	99	44,8
	<i>s1c</i>	86	38,9
	<i>s2</i>	16	7,2
i	<i>i1</i>	170	76,9
	<i>i2</i>	51	23,1
m	<i>m1</i>	167	75,6
	<i>m2</i>	54	24,4

iii. *Caracterización de los genes de adherencia.*

El principal gen relacionado con adherencia, que se caracterizó en los genomas analizados fue *babA*, específicamente su alelo *babA2*. Se realizó caracterización manual mediante alineamiento y comparación de secuencias y también mediante el análisis de las secuencias de los genomas en la base de datos VFDB, encontrando resultados similares por ambas metodologías. El alelo *babA2* se encontró en 154 genomas, es decir, el 69,7% de los genomas analizados.

Referente a los demás genes involucrados en la adherencia y OMP involucradas en adherencia y modulación de la respuesta inmune, se analizaron los genomas mediante la base de datos VFDB para determinar la presencia de *alpB/hopB*, *babB/hopT*, *hpaA*, *hopZ*, *sabA/hopP*, *sabB/hopO*, *alpA/hopC*, *napA* y *oipA/hopH*. Todos los genes analizados estuvieron presentes en los genomas caracterizados (Tabla 11). Los genes más frecuentemente encontrados en los genomas analizados fueron *hpaA* y *hopZ*, ambos presentes en el 95% de los genomas (210/221), mientras que los de menos frecuencia, fueron *babB* y *sabB*, presentes en ambos casos en el 38,5% de los genomas (85/221).

Tabla 11.

Frecuencia de genes de adherencia en los genomas analizados

Genes OMP / Adherencia	# Genomas positivos	% Genomas - n=221
Todos los genes analizados	11	5,0
<i>alpB/hopB</i>	205	92,8
<i>babA/hopS</i>	154	69,7
<i>babB/hopT</i>	85	38,5
<i>babA + babB</i>	68	30,8
<i>hpaA</i>	210	95,0
<i>hopZ</i>	210	95,0
<i>sabA/hopP</i>	129	58,4
<i>sabB/hopO</i>	85	38,5
<i>sabA + sabB</i>	51	23,1
<i>alpA/hopC</i>	203	91,9
<i>napA</i>	169	76,5
<i>oipA/hopH</i>	192	86,9

iv. *Presencia del gen iceA*

Para estudiar la presencia del gen *iceA* se evaluaron los alelos *iceA1* e *iceA2*, encontrando que el alelo más frecuente fue *iceA2* en el 52% (115/221) de los genomas analizados. En cuanto al alelo *iceA1*, se observó que fue positivo en el 48% (106/221).

v. *Análisis de combinaciones alélicas de factores de virulencia.*

Para el análisis de los resultados también se consideraron diferentes combinaciones alélicas. En primer lugar, se tuvieron en cuenta los principales factores de virulencia previamente descritos, es decir, *vacA*, *babA*, *cagPAI*, *cagA* EPIYA e *iceA*; se decidió no incluir en este análisis los resultados de los demás genes de adherencia para minimizar la cantidad de combinaciones posibles.

Se buscó el genotipo con el perfil de mayor virulencia - *vacAs1/i1/m1- babA2 - cagPAI intacto - EPIYA ABCC – iceA1* – sin embargo, este perfil no estuvo presente en ninguno de los genomas analizados. Al buscar este genotipo, pero con presencia de *cagPAI* semi-intacto o parcialmente deletado, se encontraron 9 genomas.

El genotipo más frecuente en la muestra analizada fue *-vacAs1b/i1/m1 – iceA2 – cagPAI* parcialmente deletado – EPIYA ABC – *babA2*. Este genotipo estuvo presente en 16 genomas (7,2%).

Teniendo en cuenta que el correcto funcionamiento de *cagPAI* está ligada a una buena adhesión de *H. pylori* a la célula gástrica, se analizaron las combinaciones de presencia de los genes de adherencia con la presencia de *cagPAI* intacto o semi-intacto, encontrando que solo 11 genomas tenían presencia de todos los genes de adherencia analizados y de estos 11 genomas, ninguno tenía el *cagPAI* intacto y solo 4 genomas lo tenían semi-intacto. Pero los 2 genomas que tienen el *cagPAI* intacto portaban adicionalmente la mayoría de los genes de adherencia, excepto el gen *babB* y en un caso el *babA2*. (Tabla 12).

Tabla 12.

Análisis de genes de adherencia frente a la presencia de *cagPAI*

Genes OMP / Adherencia	# Genomas positivos	Vs. <i>cagPAI</i> intacto	Vs. <i>cagPAI</i> semi-intacto
Todos los genes analizados	11	0	4
<i>alpB/hopB</i>	205	2	41
<i>babA/hopS</i>	154	1	25
<i>babB/hopT</i>	85	0	15
<i>babA + babB</i>	68	0	10
<i>hpaA</i>	210	2	40
<i>hopZ</i>	210	2	40
<i>sabA/hopP</i>	129	2	22
<i>sabB/hopO</i>	85	2	22
<i>sabA + sabB</i>	51	2	13
<i>alpA/hopC</i>	203	2	41
<i>napA</i>	169	2	41
<i>oipA/hopH</i>	192	2	38

4.3.3 Determinación *in silico* de las mutaciones asociadas con resistencia a los antimicrobianos.

El análisis *in silico* de las mutaciones asociadas a resistencia antimicrobiana reveló la presencia de diferentes mutaciones en los genomas analizados, el resumen de las frecuencias de mutaciones asociadas a cada antimicrobiano se puede visualizar en la Figura 2 y en la Tabla 13.

Tabla 13.

Frecuencia de mutaciones asociados a RAM en los genomas analizados

CLA		TET		LEV		AMX		MTZ	
Mutación	%	Mutación	%	Mutación	%	Mutación	%	Mutación	%
<i>ARNr23s</i> A2142G	0,9	<i>ARNr 16s</i> A926G	3,61	GyrA D91G	0,45	PBP3 A50S	0,45	Stop <i>frxA</i>	0,45
<i>ARNr23s</i> A2143G	2,7	Total	3,61	GyrA D91Y	0,45	PBP3 F490Y	6,76	<i>rdxA</i> H97T	0,45
Total	3,60			GyrB R484K	0,45	Total	7,21	<i>rdxA</i> D90N	1,40
				GyrA N87K	0,90			Stop <i>rdxA</i>	4,95
				GyrA N87I	3,15			<i>rdxA</i> A118T	8,10
				GyrA D91N	4,95			<i>rdxA</i> R131K	18,50
				Total	10,35			<i>rdxA</i> R90K	41,90
								Total	75,75

En el caso de mutaciones asociadas a la resistencia a CLA, se encontró que un 3,61% (8/221) de los genomas colombianos estudiados portaban alguna mutación en el gen *ARNr23S*. De estos genomas, seis (2,7%) con presencia de la mutación A2143G y dos genomas (0,9%) con presencia de la mutación A2142G.

En cuanto a las mutaciones ligadas con resistencia a TET, se analizó la mutación A926G del gen *ARNr16s* en los genomas colombianos, encontrando la presencia de esta mutación en 8 genomas, es decir el 3,61%.

Para la búsqueda de mutaciones asociadas con la resistencia a AMX, se analizaron cambios en las proteínas PBP1, PBP2 Y PBP3 (Tabla 14). Solamente se encontraron mutaciones en PBP3, identificando la mutación F490Y en 15 (6,76%) de los genomas analizados y la mutación A50S en uno de los genomas (0,45%). Obteniendo un total de 16 genomas (7,23%) con presencia de mutaciones ligadas a resistencia a AMX.

En el caso de las mutaciones asociadas a resistencia a LEV, se analizaron las mutaciones N87I, N87K, N87Y, D91G, D91N, D91Y, N97I y N97K de la proteína GyrA y las mutaciones R484K y S279G en la proteína GyrB. En los genomas analizados se encontró que el 10.26% (23/221) de estos genomas presentaban alguna de estas mutaciones. La frecuencia de las mutaciones se distribuyó así, en GyrA N87I (7/221, 3,15%), N87K (2/221,

0,90%), D91G (1/221, 0,45%), D91N (11/221, 4,95%) y D91Y (1/221, 0,45%). En GyrB solo se encontró la mutación R484K y solo en uno de los genomas analizados (0,45%).

Al analizar las mutaciones asociadas con resistencia de *H. pylori* a MTZ, se tuvieron en cuenta 2 estrategias: en primer lugar, se determinó la presencia de cambios (secuencias de parada, inserciones o cambios en el marco de lectura) en los genes *rdxA* y *frxA*. Utilizando esta estrategia se encontraron 12 genomas (5,40%) con cambios asociados, siendo más frecuentes estos cambios en el gen *rdxA* (4,95%, 11/221) que en el gen *frxA* (0,45%, 1/221). Como segunda estrategia, se analizaron las mutaciones R90K, D90N, R131K, A118T, I160F y H97T en las secuencias de la proteína codificada por el gen *rdxA*. alguna de estas mutaciones estuvo presente en 160 de los 221 genomas analizados (72,1%), la distribución de las mutaciones encontradas fue, 41,9% (93/221) a la mutación R90K, 1,4% (3/221) a la mutación D59N, 18,5% (41/221) a la mutación R131K y 8,1% (18/221) atribuido a la mutación A118T; no se encontraron genomas con las mutaciones I160F y H97T. Al considerar las dos estrategias utilizadas, se obtuvo una frecuencia de mutaciones relacionadas con resistencia a MTZ en el 75,7% (168/221) de los genomas analizados (Tabla 13).

Al analizar todas las mutaciones en conjunto se pudo evidenciar que un 14,4% de los genomas (32/221) presentaban perfiles asociados con resistencia a más de una familia de antibióticos. En estos casos, se encontró que 28 (12,6%) genomas presentaban mutaciones asociadas a 2 familias de antibióticos, en donde la presentación más común fue LEV/MTZ, presente en 12 genomas (5,43%); adicionalmente 3 (1,3%) genomas presentaron mutaciones asociadas a la resistencia a tres familias de antibióticos y en uno (0,45%) de los genomas analizados (NQ4228 - Genbank acceso número: AKNT000000000), se detectaron mutaciones relacionadas con la resistencia a todos los antimicrobianos, presentando las mutaciones A2143G en el gen *ARNr23s*, A926G en el gen *ARNr16s*, D91N en GyrA, las mutaciones A50S y F490Y en PBP3, y las mutaciones R90K y R131K en RdxA.

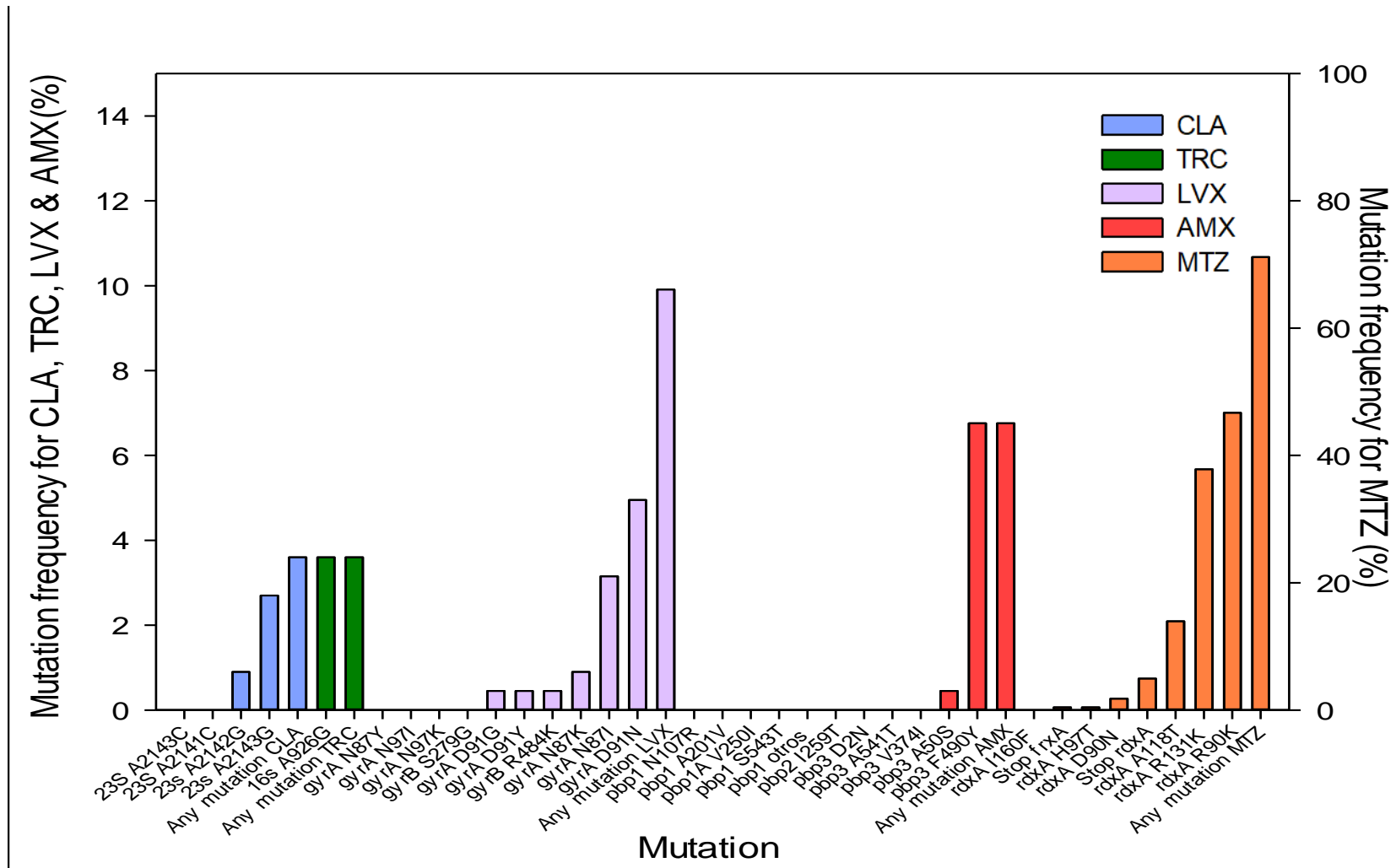


Figura 2. Frecuencia de mutaciones asociadas a RAM

Caracterización de factores de virulencia y genes relacionados con resistencia a los antimicrobianos en genomas colombianos de H. pylori

También se analizaron los resultados de las mutaciones en función del tiempo, es decir, teniendo en cuenta el año de aislamiento de la cepa bacteriana de la que se obtuvo cada genoma. Se encontró una tendencia al aumento de la presencia de mutaciones a través de los años, principalmente para las mutaciones relacionadas con AMX y MTZ (Figura 3 y Tabla 14). De los años 2012, 2017 y 2018 solo se obtuvieron 1 o 2 genomas razón por la cual los datos obtenidos para estos años no reflejan mucho respecto al resto de datos. Para 8 genomas no se encontró información relacionada con el año de aislamiento.

Tabla 14.

Distribución de RAM en el tiempo en los genomas colombianos analizados.

Año	Genomas reportados por año	CLA #	CLA %	TET #	TET %	LEV #	LEV %	AMX #	AMX %	MTZ #	MTZ %
2000	111	2	1,8	4	3,6	14	12,6	7	6,3	75	67,6
2005	12	0	0,0	0	0,0	2	16,7	0	0,0	8	66,7
2006	26	0	0,0	2	7,7	2	7,7	3	11,5	21	80,8
2007	7	0	0,0	0	0,0	1	14,3	1	14,3	7	100,0
2009	36	5	13,9	1	2,8	1	2,8	0	0,0	26	72,2
2010	6	0	0,0	0	0,0	1	16,7	1	16,7	6	100,0
2011	11	1	9,1	2	18,2	1	9,1	3	27,3	8	72,7
2012	1	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0
2017	2	0	0,0	0	0,0	0	0,0	0	0,0	1	50,0
2018	1	0	0,0	0	0,0	0	0,0	0	0,0	1	100,0
ND	8	0	0,0	0	0,0	0	0,0	0	0,0	5	62,5
Totales	221	8	3,6	9	4,1	22	10,0	15	6,8	159	71,9

ND, No Dato; CLA, Claritromicina; TET, Tetraciclina; LEV, Levofloxacina; AMX, Amoxicilina; MTZ, Metronidazol.

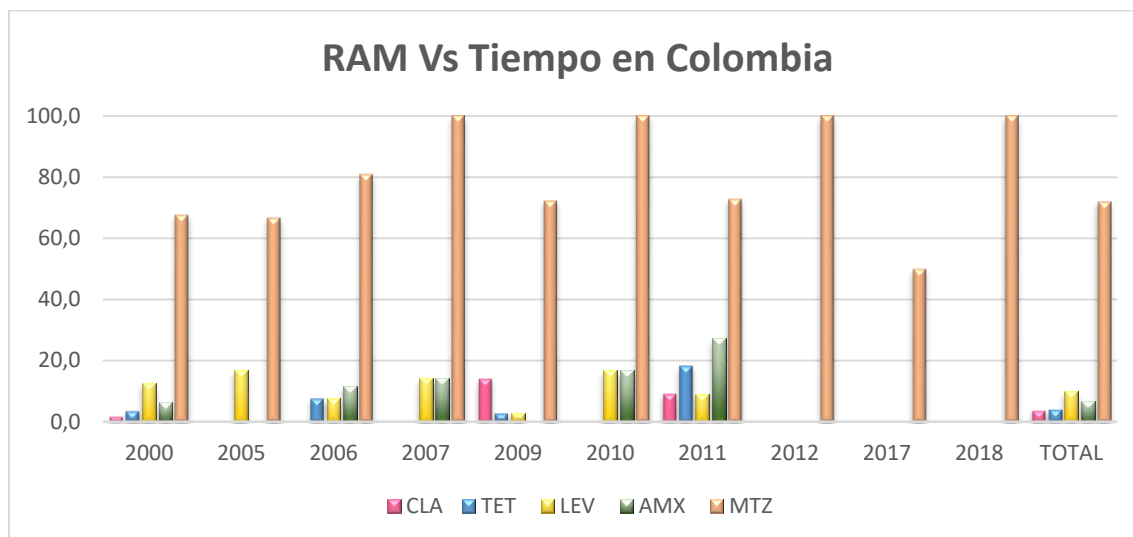


Figura 3. Distribución de RAM en el tiempo en los genomas colombianos analizados.

4.4 DISCUSIÓN

Aunque se han publicado cientos de genomas de *H. pylori*, pocos estudios se han centrado en analizar la diversidad existente en un mismo país. En Colombia y las Américas esto es especialmente importante teniendo en cuenta las altas tasas de cáncer gástrico reportadas en la región (Torres et al., 2013). Este estudio buscó caracterizar la presencia de factores de virulencia y de mutaciones asociadas a resistencia a los antimicrobianos a partir del análisis de los genomas colombianos de *H. pylori* disponibles en bases de datos públicas, junto con otros recientemente secuenciados.

Uno de los principales factores que se ha asociado con virulencia y con un mayor riesgo al desarrollo de enfermedades graves es *cagPAI* y el gen *cagA* (Nilsson et al., 2003). Estudios han reportado que la presencia del gen *cagA* se asocia con 2.1 a 2.4 veces más riesgo de desarrollar cáncer gástrico; así mismo, estudios de cohorte a largo plazo han demostrado que la infección por una cepa *cagA* positiva se asocia con un mayor riesgo de progresión de la lesión precancerosa gástrica (OR = 2,28) (Chang et al., 2018). En este estudio, se encontró que el gen *cagA* estuvo presente en el 75,1% de los genomas, una frecuencia que se encuentra en consonancia con lo previamente detectado en Colombia en donde las prevalencias de *cagA* se encuentran entre el 50% y el 73% (Arévalo-Galvis et al., 2012; Trujillo et al., 2014; Uribe

Echeverry et al., 2018; Watada et al., 2011), este dato coincide también con los reportes previos para Latinoamérica, en donde por ejemplo, para Brasil se ha reportado la presencia del gen *cagA* en el 82.8% de los aislamientos (Cavalcante et al., 2012).

Aquí no se realizaron análisis estadísticos de asociación entre los factores de virulencia como *cagPAI* y el desarrollo de patología gastroduodenal. Futuros estudios deberán ser planteados para contestar específicamente esta pregunta, y deberán contar con un diseño cuidadoso para asegurar que se controlen diferentes factores que pueden influir en la asociación. El tamaño de la muestra y la selección adecuada de los casos que ingresen al estudio serán de vital importancia. Para un país como Colombia otro factor importante para tener en cuenta será la inclusión de una muestra homogénea entre la población de zonas de bajo y alto riesgo de desarrollo de cáncer.

Hasta ahora los reportes que analizan la asociación del *cagPAI* intacto con el desenlace clínico no son concluyentes. Algunos autores reportan que existe una asociación significativa entre la integridad del islote y el riesgo de desarrollar resultados graves de la infección (Ahmadzadeh et al., 2015; Markovska et al., 2018). Sin embargo, en algunos estudios, no se han reportado diferencias en la prevalencia de *cagPAI* intacto entre pacientes con úlcera duodenal y gastritis (Patra et al., 2011). Dada la relevancia que tiene *cagPAI* en las actividades biológicas de *H. pylori* que conducen al daño tisular (Censini et al., 1996), los polimorfismos en los genes que codifican las proteínas que lo forman probablemente juegan un papel importante como factores de virulencia. La falta de evidencia concluyente resalta la importancia de que estudios futuros evalúen los diferentes genes de *cagPAI* y sus polimorfismos en pacientes colombianos con lesiones preneoplásicas para examinar si estas variantes también presentan diferencias significativas en la frecuencia entre las lesiones premalignas y la gastritis, para identificar individuos infectados con mayor riesgo de desarrollar cáncer gástrico y que serían buenos candidatos a erradicación.

Los resultados obtenidos en este estudio revelaron la presencia de *cagPAI* intacto solamente en dos de los genomas colombianos analizados; aunque es una baja frecuencia es importante tener en cuenta que la diversidad dentro de *cagPAI* se ha observado a través de las diferentes poblaciones de *H. pylori* y las diferentes regiones geográficas. De acuerdo con los reportes, solamente entre la mitad y dos tercios de las cepas de *H. pylori* occidental llevan el *cagPAI*

y se ha identificado una delección parcial de *cagPAI* en el 4% al 88% de las cepas de *H. pylori* (Hanafiah & Lopes, 2020). Se ha descrito que 17 genes de *cagPAI* son esenciales para la translocación de CagA y 14 genes que son indispensables para la inducción completa de IL-8. Dentro de los genes que son esenciales para la traslocación de CagA pero no para la producción de IL-8, se han descrito por ejemplo, *cag5/cagβ*, *cagZ*, *cagI* y *cagα/virB11* (Nguyen et al., 2010). El análisis de *cagPAI* en los genomas colombianos reveló que 43 de los 221 genomas analizados (20%) portan los 17 genes esenciales para la translocación de CagA.

Algunos análisis han mostrado que la prevalencia de *cagPAI* intacto varía mucho entre regiones geográficas (57,1% en Japón, 18,6% en Perú y 12% en la India) (Kausser et al., 2004). Respecto a Colombia, no se encuentran muchos reportes relacionados. En 2012 Bravo et al, reportaron que a partir de la amplificación por PCR de los genes *CagA*, *CagT*, *CagE* y *CagI0/cagV* se pudo concluir que el 77% de los aislamientos portaban un *CagPAI* íntegro (Bravo et al.). Dos análisis recientemente realizados sobre población Latinoamericana, incluyendo cepas procedentes de Colombia, (Canzian et al., 2020; Rizzato et al., 2020), buscaron establecer asociación entre la presencia y los polimorfismos de genes de *cagPAI* y el riesgo de desarrollar lesiones premalignas gástricas avanzadas y cáncer gástrico. Estos estudios demostraron que varios de los genes *cagPAI* de las cepas de América Latina contienen una serie de variantes no sinónimas en frecuencias relativamente altas que podrían influir en el resultado clínico. En ninguno de los 2 estudios se encontró una asociación que permaneciera estadísticamente significativa después del ajuste para comparación múltiple; sin embargo, las puntuaciones obtenidas para los polimorfismos individuales se asociaron significativamente con el riesgo de lesiones premalignas gástricas avanzadas y cáncer gástrico. Estos resultados resaltan la necesidad de realizar más estudios para establecer marcadores de riesgo en la población general, con el fin de orientar la erradicación de *H. pylori* en grupos de alto riesgo (Canzian et al., 2020).

Los análisis de los genes individuales de *cagPAI* en este estudio mostraron que el gen más conservado en los genomas colombianos era *cag3*, uno de los genes del sistema TSS4 que carece de homólogos en otras especies bacterianas (Fischer et al., 2020); por otra parte, los genes con menor frecuencia en los genomas colombianos fueron *cagI*, *cagP* y *cagQ*, genes

que no son considerados esenciales ni para la traslocación de *cagA* ni para la inducción de IL-8 (Nguyen et al., 2010). Probablemente, estos genes podrían ser considerados como marcadores de presencia de la integridad de *cagPAI* entre aislamientos colombianos.

En cuanto a los motivos EPIYA, se ha reportado que las múltiples repeticiones en C causan un aumento en el número de fosforilaciones en la proteína CagA aumentando la virulencia bacteriana. En reportes colombianos previos se ha descrito la presencia del motivo ABCCC en un 21% de los aislamientos estudiados (Rodríguez Gómez et al., 2020); sin embargo, en este estudio no se encontró este motivo en ninguno de los genomas analizados, aunque, si se encontró el motivo EPIYA-ABCC (25,3%) y el motivo EPIYA-ABC (72%) que fue el más prevalente entre la población analizada. Estudios han mostrado que en los países occidentales, las cepas que albergan múltiples motivos EPIYA-C (ABCC o ABCCC) tienen mayor capacidad de fosforilación y confieren un mayor riesgo de cáncer gástrico (OR = 3,28) en comparación con solo 1 motivo EPIYA-C (Li et al., 2017). Los análisis realizados aquí revelaron frecuencias coincidentes con estudios previos en la región de Latinoamérica, en países como México y Brasil han mostrado que el motivo EPIYA-ABC es el más frecuente en pacientes con gastritis crónica, mientras que los motivos EPIYA-ABCC y EPIYA-ABCCC se observan con mayor frecuencia en pacientes con úlcera péptica y cáncer gástrico (Atrisco-Morales et al., 2018; Beltrán-Anaya et al., 2014; Vianna et al., 2015; Vilar E Silva et al., 2014).

Otro factor de virulencia analizado en este estudio fue la citotoxina vacuolizante VacA, la cual está involucrada en la formación de poros en la célula infectada (Baj et al., 2020). El gen *vacA* que codifica para la producción de esta proteína está presente en todos los aislamientos de *H. pylori*; sin embargo, se ha comprobado que las diferentes combinaciones alélicas de las regiones *s*, *m*, *i* serán determinantes para analizar la virulencia de una cepa (McClain et al., 2017). Un estudio de metaanálisis mostró que en la población occidental, las personas infectadas con *H. pylori* que sea *vacA s1* o *m1* tienen un mayor riesgo de desarrollar cáncer gástrico (OR = 2,50–5,32) (Matos et al., 2013). En el presente estudio, 17 diferentes combinaciones alélicas fueron encontradas en los genomas colombianos de *H. pylori*, las más frecuentes fueron *vacAmlis1b* y *vacAmlis1c*, encontradas en el 40,7% y el 24% de los genomas analizados. Estas combinaciones alélicas

se consideran altamente virulentas y diferentes estudios han comprobado que la combinación *vacAs1m1* tiene mayor capacidad vacuolizante; sin embargo, se sabe que dentro del alelo *s1* se considera de mayor virulencia el subtipo *s1a*, seguido del *s1b* y del *s1c* (Amieva & El-Omar, 2008; Atherton et al., 1995)

El genotipo *vacAs1* es uno de los más abundantes en todos los aislamientos de *H. pylori* (Baj et al., 2020), algunos autores han reportado asociación de este genotipo y la presencia de gastritis (Atrisco-Morales et al., 2018), otros con úlcera péptica (Keikha et al., 2020) y otros con la generación de cáncer gástrico (Román-Román et al., 2017). En Colombia, se ha reportado que el genotipo predominante es *vacAs1m1* (Carlosama-Rosero et al., 2021); no obstante, los estudios disponibles en el país tampoco aportan una conclusión sobre asociación entre el genotipo *vacA* y el desenlace clínico. Desde 1999 se han tratado de abordar estas temáticas en el país, Yamaoka et al. (Yamaoka et al., 1999) compararon la distribución de los genotipos *vacAs1m1* en cuatro países y demostraron que el subtipo *s1m1* predomina en Colombia, pero no se encontraron asociaciones estadísticamente significativas con el riesgo de desarrollo de cáncer. En el 2002, Citty et al. (Citty et al., 2002) reportaron que el genotipo *s1m1* estuvo presente en el 81% de los pacientes con cáncer y en cerca del 40% de los pacientes con gastritis; sin embargo no reportan un análisis de asociación. En una reciente publicación realizada en pacientes del Cauca, se ha reportado el genotipo *vacAs1m1* fue el más prevalente tanto en pacientes con cáncer como en pacientes con gastritis no atrófica; pero por análisis de regresión logística lograron demostrar asociación entre este genotipo y cáncer gástrico, aunque solo en pacientes mayores de 50 años (OR: 6,18; IC: 1,25-30,51; p = 0,025) (Carlosama-Rosero et al., 2021). Los hallazgos obtenidos en el presente estudio muestran que la mayoría de los genomas analizados portan el genotipo *vacAs1m1*.

La región intermedia del gen *vacA* al parecer desempeña un papel funcional en la formación de vacuolas, siendo las cepas *vacA*il** las que tienen la capacidad vacuolizante (Keikha, 2020). Pocos estudios han analizado esta región en población colombiana, los datos hasta ahora disponibles muestran mayor frecuencia del alelo *il* (Melo-Narváez et al.). Los resultados obtenidos aquí muestran el 75% (166/221) de los genomas analizados son portadores del alelo *il*. Es importante tener en cuenta que numerosos estudios han demostrado que la frecuencia de este alelo es comúnmente alta en los genotipos *vacAs1m1*

(Basso et al., 2008; Ogiwara et al., 2008; Ogiwara et al., 2009; Rhead et al., 2007), el cual también fue el genotipo más frecuente dentro de los genomas analizados.

El papel que pueda desempeñar la región intermedia en el desarrollo de patologías ligadas a la infección por *H. pylori* aún es controvertido; mientras algunos autores describen asociación del genotipo *il* con el desarrollo de cáncer gástrico y/o úlcera péptica (Karlsson et al., 2012; Keikha, 2020), algunos otros muestran que no existen asociaciones significativas (Memon et al., 2014; Pormohammad et al., 2018). En este estudio, se encontró que solo 71 genomas de los 166 positivos para *il*, contaban con información clínica y en estos genomas no se observó una tendencia que permitiera buscar una relación de causalidad.

Otro factor de virulencia putativo que parece estar presente en prácticamente todas las cepas de *H. pylori* es el codificado por el gen *iceA*, que se presenta en dos formas alélicas, *iceA1* e *iceA2*. La expresión de *iceA1* se regula positivamente tras el contacto de *H. pylori* con las células epiteliales gástricas humanas y se ha asociado con la enfermedad de úlcera péptica; por el contrario, *iceA2* parece no estar involucrado en la patología gastrointestinal (van Doorn et al., 1998). En este estudio se encontró una frecuencia mayor del alelo *iceA2* (52%) que del alelo *iceA1* (48%), sin embargo, las cifras son muy cercanas, lo cual resulta poco concluyente.

En relación con los factores de adherencia, se sabe que son vitales en el inicio del proceso de infección de *H. pylori*, específicamente para lograr la adhesión a la mucosa gástrica y mediar en la respuesta inflamatoria. Se han identificado más de 30 OMP en *H. pylori*, y la mayoría de ellas están implicadas en la adherencia bacteriana (Tomb et al., 1997); sin embargo, solo algunas de las OMP han sido descritas como potencialmente determinantes en la patogenicidad de *H. pylori*, dentro de ellas, SabA, SabB, BabA, BabB, OipA, HopZ, AlpB, AlpB y HpaA. La expresión de estas OMP puede contribuir a la enfermedad gastrointestinal al mediar en la adhesión de *H. pylori* y mejorar la actividad de *cagPAI* (Cover, 2016).

Al analizar la presencia de los genes que codifican para estas OMP en los genomas colombianos se encontró que los más frecuentes fueron *hpaA* y *hopZ*, presentes en el 95% de los genomas. Respecto a *hpaA*, se sabe que la proteína que codifica, HpaA, puede unirse a

varios componentes de glicosilación en la superficie de las células epiteliales gástricas, garantiza la mediación de la unión al ácido siálico. Debido a su importante función, a que está presente en la mayoría de las cepas de *H. pylori* y a que se han encontrado anticuerpos frente a HpaA en pacientes infectados, esta proteína se ha propuesto recientemente como un buen antígeno candidato a vacuna (Banga Ndzouboukou et al., 2021). Del gen *hopZ*, se sabe que puede desempeñar un papel fundamental en la adhesión, pero sus receptores específicos aún no se conocen y la expresión del gen depende del mecanismo de emparejamiento de las hebras en las repeticiones de dinucleótidos CT en la región codificante del péptido señal. También depende del interruptor de encendido/apagado mediado por la variación de fase durante la colonización temprana de *H. pylori*, lo que indica que *hopZ* tiene una fuerte selectividad in vivo, juega un papel importante en la adaptación al entorno del huésped y tiene un papel esencial en la colonización de la mucosa gástrica durante la infección temprana (Xu et al., 2020). Los estudios disponibles hasta ahora no permiten dejar claro si *hopZ* está relacionado con otros factores de virulencia o está relacionado con enfermedades clínicas (Servetas et al., 2018).

Respecto al alelo *babA2* al gen *babA* asociado con adhesión a la célula epitelial, se encontró una frecuencia del 69,7%, una cifra similar a lo previamente reportado para Colombia y Latinoamérica, en donde se han reportado frecuencias de este gen que van desde el 46% hasta el 82,3% (Arévalo-Galvis et al., 2012; Gatti et al., 2006; Paniagua et al., 2009; Umit et al., 2009). Es importante tener en cuenta que detectar el alelo *babA2* no necesariamente refleja el estado funcional del gen. Según el nivel de producción de proteína BabA, los aislamientos de *H. pylori* se pueden dividir en altos productores de BabA (*babA2* detectable y buena actividad de unión a Le^b), productores bajos de BabA (*babA2* detectable pero carecen de actividad de unión a Le^b) y no productores de BabA (*babA2* no detectable y actividad nula de unión a Le^b) (Fujimoto et al., 2007). Es por esta razón que probablemente los estudios que intentan asociar la presencia de *babA2* con desarrollo de patologías, tienen resultados contradictorios. Adicionalmente, algunos resultados han mostrado que buscar el gen utilizando la técnica de PCR, no es lo más conveniente, ya que dependiendo del tipo de cebadores que se use, los resultados varían (Šterbenc et al., 2020). Estos datos sugieren que los métodos de detección de *babA2* deben seleccionarse cuidadosamente. De acuerdo

con algunos autores la mejor estrategia sería optar por metodologías basadas en la evaluación de proteínas o la expresión de estas. Así mismo, es importante tener en cuenta que existen cepas de *H. pylori* que son consideradas como “especializadas” por su capacidad de unión exclusivamente a los antígenos ALe^b, las cuales están predominantemente en países suramericanos (Yamaoka, 2008).

Al considerar las combinaciones alélicas de algunos genes evaluados, el genotipo más común fue *vacAs1b il m1 / babA2 / iceA2 / cagPAI* parcialmente deletado con *cagA*⁺ (EPIYA-ABC) encontrada en un 7,2% de los genomas evaluados. Otros reportes de genotipos en Colombia han determinado una frecuencia mayor en el fenotipo *vacAs1a m1 / babA2 / iceA1 / cagA*⁺ (Arevalo, 2010) evidenciando diferencias en el subalelo de las región *s* del gen *vacA* y en *iceA*, probablemente debido a los diferentes tamaños y tipos de muestra analizados. Otros estudios similares en el mundo han mostrado que existe un alto nivel de heterogeneidad genética entre diferentes cepas de *H. pylori*. Esta diversidad se ha estructurado principalmente por los tipos de pacientes, esto relacionado específicamente a la clínica, la localización geográfica y la etnia (Rhead et al., 2007). Es importante tener en cuenta que ninguno de los marcadores de virulencia puede considerarse un factor independiente para el resultado de la enfermedad y que los estudios han mostrado que cuando están presentes múltiples factores de virulencia, el riesgo de un resultado clínico grave es mayor (Kusters et al., 2006). Los datos analizados en el presente estudio no permitieron realizar un análisis de asociación para determinar si la presencia de los factores de virulencia representaba un mayor riesgo de lesiones gástricas graves en la población colombiana. Contar con más información demográfica y clínica de los pacientes a partir de los cuales se obtuvieron los genomas analizados hubiera permitido llegar a conclusiones más específicas sobre la realidad de la infección por *H. pylori* en el país y a través del tiempo.

En los 221 genomas incluidos en el estudio también se analizaron las mutaciones relacionadas con resistencia a los antimicrobianos. Diversos estudios han analizado esta problemática en el país; sin embargo, los análisis disponibles no logran aún reflejar la verdadera problemática de la resistencia en Colombia. En general los datos son diversos, muestran una situación relacionada con un grupo específico de pacientes y reflejan una imagen del momento específico en el cual se obtuvieron los aislamientos bacterianos.

Recopilar los genomas colombianos de *H. pylori* desde bases de datos y analizar en ellos la presencia de mutaciones asociadas a RAM pretende mostrar un panorama más amplio, que incluya datos de más regiones del país, con un espectro de tiempo de al menos 10 años y que permita dar cuenta de las características genómicas ligadas a RAM circulantes en Colombia. En este estudio se encontraron mutaciones asociadas a resistencia a CLA en el 3,6% de los genomas analizados, a TET en el 3,61%, a AMX en el 7,23%, a LEV en el 10,26% y a MTZ en el 75,7%.

En el caso de CLA, se buscaron las mutaciones A2142G, A2143G, A2143C y A2141C en el gen *rARN23s*, las cuales han demostrado tener buena correlación con el perfil fenotípico de resistencia (Trespacios et al., 2013). Los estudios disponibles de resistencia a este antibiótico en Colombia habían mostrado datos muy diversos con prevalencias de resistencia que van desde 0% hasta 46% (Arévalo-Jaimes et al., 2019; Bustamante-Rengifo et al., 2013; Mannion et al., 2021; Matta et al., 2018; Trespacios et al., 2013); estos márgenes tan amplios pueden estar relacionados entre otras causas, con los factores de inclusión utilizados en cada estudio, además porque cada estudio reportado analizó aislamientos provenientes de una región geográfica del país, al parecer la prevalencia de resistencia varía entre regiones en Colombia.

En este estudio se encontraron mutaciones asociadas con resistencia a CLA en el 3,6% de los genomas analizados, esto sumado a los estudios de prevalencia de resistencia en el país muestran la importancia de desarrollar estudios vigilancia de la resistencia mediante estudios genotípicos y fenotípicos. Esto permitirá obtener en tiempo real un dato preciso que refleje la realidad del país frente a la resistencia que expresa *H. pylori* a CLA. Información importante teniendo en cuenta que de acuerdo con las más recientes recomendaciones del consenso de Maastricht V / Florence para el tratamiento empírico, cuando en un país la resistencia a claritromicina es menor de 15%, se recomienda utilizar este antibiótico en una terapia triple de 14 días, pero si es mayor a 15%, se debería reemplazar este antibiótico por LEV en caso que la resistencia a este sea menor a 15% (Malfertheiner et al., 2017).

En cuanto a las mutaciones más comúnmente descritas en Colombia, la mutación A2143G ha sido la más frecuente, presente en el 90,5% de los aislamientos resistentes, seguida por la

transición A2142G (7,1 %) y la A2142C (2,4 %), los hallazgos obtenidos aquí fueron coherentes con estos datos.

Respecto a los hallazgos obtenidos para mutaciones ligadas con resistencia a TET, es importante mencionar que son pocos los reportes en el país y los datos difieren bastante entre algunos estudios. Un trabajo del año 2012 reporta una prevalencia de resistencia menor al 1% (Rodríguez - Gómez, 2012), otro reporta datos de resistencia a TET del 85,7% (Yepes et al., 2008) y un reciente estudio encontró una resistencia a TET en el 1,7% de las cepas analizadas (Mannion et al., 2021). Es importante mencionar que el estudio de Yepes et al., el cual encontró cifras muy altas de resistencia, usó la técnica de difusión con disco, la cual es poco recomendada para el estudio de resistencias en *H. pylori*. A nivel global, se ha reportado una resistencia a TET entre el 0% y el 7,3% siendo uno de los antibióticos propuestos como ideales por su baja resistencia (Hu et al., 2017).

En relación con el antibiótico AMX, uno de los de primera elección para el tratamiento de *H. pylori* en Colombia (Atehortua Rendon et al.), se han documentado tasas de resistencia en el país de alrededor del 7%, siendo una de las más altas de la región latinoamericana (Camargo et al., 2014). Aunque los datos encontrados en el presente estudio muestran la presencia de mutaciones en el 7,21% de los genomas analizados y esta cifra es concordante, en este estudio se encontraron mutaciones solamente en la proteína PBP3, la cual no había sido previamente analizada en Colombia. Probablemente los análisis anteriores han subestimado la resistencia a AMX en Colombia.

Por otro lado, la búsqueda de mutaciones asociadas con la resistencia a LEV permitió el hallazgo de mutaciones relacionadas en el 10,26% de los genomas analizados. Este antibiótico es una fluoroquinolona empleada en Colombia generalmente como terapia de rescate en tratamientos no exitosos donde se ha usado CLA (Atehortua Rendon et al.). Un estudio previo en Colombia realizó un seguimiento a la resistencia a LEV en pacientes de Bogotá entre los años 2009 y 2014, encontrando una tasa de resistencia para el año 2009 del 11,8 % y del 27,3% para el año 2014 (Trespacios-Rangél et al., 2016). Los autores consideraron que el aumento en la resistencia se debió a un aumento en el uso de LEV en tratamientos de infecciones del tracto respiratorio y urinario. Recientemente un estudio publicó una tasa de resistencia a LEV de 20,3%, descrita en dos poblaciones del sur del país

que se caracterizan por tener bajo y alto riesgo de desarrollar cáncer gástrico (Mannion et al., 2021). En cuanto a las mutaciones asociadas a la resistencia a LEV, las que se encontraron aquí fueron similares a las previamente reportadas, siendo N87I una de las mutaciones más frecuentes; no obstante, se debe destacar que se encontró la mutación D91N en el 4,98% de los genomas analizados. Esta mutación no había sido muy relevante en estudios anteriores (Rimbara et al., 2012). Los resultados obtenidos aquí brindan una información más amplia sobre las mutaciones que pueden estar involucradas en la resistencia a LEV en Colombia; sin embargo, es importante realizar un análisis de correlación de estas mutaciones frente a la determinación de la Concentración Mínima Inhibitoria (CMI).

Respecto a las mutaciones asociadas con resistencia de *H. pylori* a MTZ, en este estudio se encontró que el 75,68% de los genomas analizados contaban con algún cambio relacionado. En estudios previos en Latinoamérica, la resistencia a MTZ es alta (>30%). Sin embargo, en Colombia el promedio de las tasas de resistencia es del 83% (Camargo et al., 2014). A pesar de estas altas tasas de resistencia reportadas por los estudios, en Colombia, MTZ sigue siendo uno de los antimicrobianos más prescritos por los médicos para el tratamiento de *H. pylori* (Gómez et al., 2007). Los estudios que han descrito esta problemática en el país han utilizado métodos microbiológicos, con resultados de entre el 72 y el 97,6% (Trespacios et al., 2010; Yepes et al., 2008; Álvarez et al., 2009); pero también métodos genotípicos que han evaluado la frecuencia de mutaciones en el gen *RdxA* reportando cifras de 78,2% (Acosta et al., 2017). En el presente estudio se tuvieron en cuenta cambios no solo en este gen sino también en el gen *frxA*. Los hallazgos obtenidos en el presente estudio presentan un panorama más amplio de los cambios que pueden estar relacionados con resistencia a MTZ en Colombia y brinda una posible alternativa a estudiar para la detección molecular de resistencia a MTZ.

Al analizar los resultados de las mutaciones asociadas a RAM en todos los genomas estudiados, se encontraron 32 (14,4%) genomas con mutaciones ligadas a resistencia frente a más de un antibiótico. Este fenómeno de Multi-Drogo-Resistencia (MDR) o resistencia simultánea, se observa cada vez más dentro de la emergencia global de RAM en *H. pylori*. Algunos estudios han reportado la aparición de cepas MDR hasta en el 40% de las infecciones en algunas regiones del mundo, lo cual constituye una grave amenaza para las opciones terapéuticas disponibles (Boyanova et al., 2019). En Colombia algunos estudios

han demostrado la presencia de cepas de *H. pylori* MDR. Un estudio realizado sobre aislamientos obtenidos de pacientes de Tumaco en 2012, detectó la presencia de cepas resistentes a CLA y AMX (Figuerola et al., 2012), y otro reportado en 2019, describió la presencia de cepas resistentes a dos o más antibióticos (AMX, CLA, LEV y MTZ) y que habían sido aisladas de pacientes con tres o cuatro tratamientos fallidos (Arévalo et al., 2019).

En Colombia el fenómeno de MDR cobra especial importancia ya que un esquema terapéutico que aún se sigue usando en el país es la terapia cuádruple, en la cual se combinan un IBP, bismuto y dos antibióticos, generalmente CLA y MTZ. En el país se ha demostrado que después de tres tratamientos fallidos, los resultados de los cultivos evidencian la presencia de cepas MDR y los esquemas tradicionales no son una buena opción, sino que se deben usar terapias de rescate (Otero R et al., 2018). Ampliar la búsqueda de los aislamientos MDR en el país y conocer la prevalencia actual de estas cepas permitirá la generación de recomendaciones encaminadas a mejorar las altas tasas de falla terapéutica.

Es bien conocido que el fenómeno de MDR en *H. pylori* emerge principalmente por la aparición de diversas mutaciones que conducen simultáneamente a la resistencia frente a diferentes familias de fármacos (Tuan et al., 2019); sin embargo, algunos autores han descrito mecanismos MDR adicionales, aunque con una relevancia clínica limitada. Dentro de estos mecanismos se reconocen, por ejemplo, la sobreexpresión de los sistemas de bombas de eflujo, la disminución en la expresión de OMP que lleva a una absorción restringida de los fármacos, la formación de biopelículas in vivo y el cambio morfológico de *H. pylori* a forma cocoide (Tshibangu-Kabamba & Yamaoka, 2021). La presencia de estos mecanismos no se logra detectar en todos los casos mediante la búsqueda de mutaciones y en algunos casos tampoco podría ser detectada mediante pruebas de susceptibilidad fenotípica. Teniendo en cuenta esta problemática, es cada vez más evidente el reto que se presenta de encontrar pruebas que permitan un diagnóstico más preciso sobre la resistencia de *H. pylori* a los antimicrobianos y probablemente sea necesario un enfoque que combine las técnicas tradicionales de cultivo y las ciencias ómicas.

Los datos obtenidos aquí respecto a la presencia de mutaciones ligadas a RAM en general fueron coherentes con los estudios previos de resistencia en el país. Sin embargo, es

importante mencionar que la mayoría de los estudios previos utilizaron técnicas microbiológicas o detección de mutaciones por PCR, y pocos estudios han utilizado herramientas de genómica para describir esta problemática. Aquí se usaron por primera vez, los genomas colombianos de *H. pylori* depositados en bases de datos públicas, los cuales en su mayoría fueron secuenciados usando la plataforma Illumina para WGS. El principal problema de este método es que genera lecturas cortas lo cual normalmente produce genomas ensamblados de forma incompleta que se fragmentan en secuencias no continuas (contigs) (Kaas et al., 2014). Entonces, aunque las tecnologías proporcionen una cobertura alta para la detección de mutaciones, los contigs producidos a veces pueden resultar en genes faltantes o fragmentados, lo cual puede explicar una ligera disminución en algunas de las cifras encontradas. No obstante, para el caso de TET y AMX, los datos obtenidos aquí mostraron mayor capacidad de detección de mutaciones, respecto a los análisis previamente reportados en el país.

Nuestros resultados muestran que el análisis WGS para caracterizar virulencia y RAM en *H. pylori* es un método rentable y fácil de usar. Con la masificación de las técnicas de secuenciación cada vez será más fácil poder acceder a ellas y utilizarlas podría brindar una solución para la caracterización rápida y completa de *H. pylori*, ya que además de permitir la detección de mutaciones específicas, o de caracterizar los factores de virulencia, también brinda herramientas para analizar todo el genoma, analizar la filogenia, y realizar la búsqueda de otros potenciales blancos de resistencia o reordenamientos cromosómicos que pueden impactar en el fenotipo bacteriano.

Los resultados obtenidos aquí también permitieron tener una mirada más completa de las cepas de *H. pylori* que han venido circulando en el país. Aunque no se contó con información demográfica completa de los aislamientos, se lograron analizar genomas que provenían de aislamientos obtenidos desde el año 2000 hasta el 2018, de pacientes con diferentes patologías gástricas y provenientes de al menos siete regiones diferentes del país. Este estudio permite tener una aproximación a la problemática real de la infección por *H. pylori* en el país. La información obtenida aquí aportará conocimiento sobre el genotipo, la cual pueda ser usada para el desarrollo de aplicaciones que permitan un diagnóstico rápido de la virulencia y la resistencia antimicrobiana expresada por *H. pylori*.

4.5 CONCLUSIONES

Los análisis genómicos para la identificación de virulencia y resistencia antimicrobiana en genomas colombianos de *H. pylori*, mostraron que esta puede ser una forma rápida y confiable para la caracterización completa de aislamientos.

Los resultados permitieron visualizar que los genotipos de virulencia circulantes en el país son diversos y no están agrupados en una región geográfica específica o para un grupo de pacientes, sin embargo, el genotipo más frecuentemente encontrado fue *vacAs1b il ml / babA2 / iceA2 / cagPAI* parcialmente deletado con *cagA+* (EPIYA-ABC).

En el caso de RAM, se encontraron mutaciones asociadas a resistencia a CLA en el 3,6% de los genomas analizados, a TET en el 3,61%, a AMX en el 7,23%, a LEV en el 10,26% y a MTZ en el 75,7%. Es importante tener en cuenta que los datos obtenidos en este estudio corresponden a una visión global de los genomas colombianos de *H. pylori* en un tiempo determinado por lo tanto no reflejan la situación actual de resistencia en el país.

5. CAPITULO 3.

ANÁLISIS DE DIVERSIDAD Y FILOGENIA DE GENOMAS COLOMBIANOS DE *H. pylori*.

5.1 INTRODUCCIÓN

Helicobacter pylori ha evolucionado juntamente con los seres humanos durante más de 60000 años (Linz et al., 2007). Esta asociación simbiótica íntima y prolongada de *H. pylori* con los humanos y la acumulación de cambios adaptativos específicos en el hospedero han dado lugar a la aparición de diferentes genotipos de *H. pylori* (Yamaoka, 2009), los cuales se caracterizan por tener una distribución específica a través de las diferentes regiones geográficas del mundo (Falush, Wirth, et al., 2003).

Los primeros estudios de tipificación filogenética de *H. pylori* realizados por análisis de secuencias de múltiples locus (MLST) han demostrado que las cepas modernas de *H. pylori* se agrupan en diferentes poblaciones regionales según su origen geográfico, dentro de estas: hpEurope, hpNEAfrica, hpAfrica1, hpAfrica2, hpAsia2, hpSahul y hpEastAsia; hpEastAsia se divide en tres subpoblaciones, hspEAsia, hspMaori y hspAmerind (Achtman et al., 1999; Falush, Wirth, et al., 2003; Moodley et al., 2009; Yamaoka, 2009). La subpoblación hspAmerind es un reflejo de la migración humana que ocurrió hace 12.000 años desde Asia hacia las Américas a través del Estrecho de Bering (Kersulyte et al., 2010). Sin embargo, hace 500 años ocurrió en América uno de los eventos más importantes en la evolución humana. La colonización española, la cual conectó no solo a los recién llegados europeos, sino también a los africanos con los pueblos Indígenas Americanos (O'Rourke & Raff, 2010).

Así comenzó uno de los eventos de mezcla natural más grandes entre las poblaciones humanas modernas, que dio origen a la gran diversidad y multiculturalismo que se observa hoy en las Américas (Ruiz-Linares, 2014). Con la llegada de los nuevos pobladores, también llegaron nuevos patógenos, incluidas nuevas cepas de *H. pylori*, lo cual generó la desaparición casi completa de las cepas nativas amerindias en los países latinoamericanos en

las décadas posteriores (Parrish et al., 2008). Los estudios de MLST en aislamientos colombianos de *H. pylori* han demostrado que las cepas son principalmente del tipo hpEurope y, en menor proporción, del tipo hpAfrica; este último tipo se ha encontrado principalmente en la población afroamericana que vive a lo largo de la costa (de Sablet et al., 2011).

H. pylori es una bacteria capaz de realizar procesos adaptativos locales muy rápidos mediante mutación y recombinación homóloga con otras cepas (Furuta et al., 2015). En Colombia, estas características son especialmente evidentes. Se ha descrito la existencia de un componente poblacional específico en cepas aisladas de la población mestiza colombiana, lo que sugiere que, a pesar de su origen europeo o africano, estas cepas se encuentran claramente diferenciadas dentro de esta población (Muñoz-Ramírez et al., 2017). Estudios recientes basados en secuencias de genoma completo han demostrado que, en Colombia, las cepas de *H. pylori* han seguido vías evolutivas únicas; y que, en Colombia, las poblaciones bacterianas evolucionan rápidamente y han formado nuevas subpoblaciones de origen europeo (Gutiérrez-Escobar et al., 2017; Muñoz-Ramírez et al., 2020; Muñoz-Ramírez et al., 2017; Thorell, Yahara, et al., 2017). Estos hallazgos destacan la necesidad de realizar más estudios de caracterización filogenética para describir la estructura poblacional de los aislamientos colombianos de *H. pylori*.

En este estudio se caracterizaron un total de 221 genomas de aislamientos colombianos de *H. pylori*. Se realizaron análisis filogenéticos por medio del análisis de MLST y mediante la búsqueda de polimorfismos de nucleótido simple (SNP) en el genoma completo de los aislamientos.

5.2 METODOLOGÍA

Para realizar los análisis filogenéticos se utilizaron 221 genomas colombianos de *H. pylori*. 29 genomas secuenciados, anotados y publicados en este estudio, los cuales están descritos en la Tabla 1 del capítulo 2; y 192 genomas recuperados de la base de datos WGS del NCBI, los cuales están descritos en el anexo 4.

Algunos genomas de referencia se utilizaron para realizar las comparaciones filogeográficas, estos serán descritos en cada método.

5.2.1 Análisis filogenético por MLST

Para la asignación de poblaciones por este método se utilizaron las secuencias de siete genes “house-keeping” de *H. pylori* (*atpA*, *efp*, *trpC*, *ppa*, *mutY*, *yphC* y *ureI*). Se obtuvieron las secuencias de estos genes de 741 genomas de *H. pylori* disponibles en la base de datos PubMLST (<http://pubmlst.org/helicobacter/>) (Jolley et al., 2018), las cuales habían sido previamente descritas por Falush *et al.* (Falush, Wirth, et al., 2003) y Linz *et al.* (Linz et al., 2007), así como la secuencia de estos genes de los 221 genomas colombianos en estudio.

Se creó un archivo multi-FASTA con las secuencias concatenadas de los siete genes para cada genoma. Las secuencias se alinearon utilizando MAFFT versión 7 (Kato & Standley, 2013) y con el alineamiento se construyó un árbol filogenético en el software MEGA 7.0 (Molecular Evolutionary Genetics Analysis) (Kumar et al., 2016), usando el método de Neighbor-joining y el modelo de sustitución de dos parámetros de Kimura (Kimura, 1980). El número de repeticiones bootstrap se estableció en 1000.

Además, para cada gen “house-keeping” se creó un archivo multi-FASTA. Los alelos de estos 7 genes se alinearon usando FASTA y los resultados de los alineamientos se fusionaron en un archivo, el cual fue convertido al formato xmfA utilizando la herramienta xmfA2struct (Xavier Didelot, Universidad de Warwick, Coventry, Reino Unido). Este archivo fue analizado en el software Structure 2.3.4 (Falush, Stephens, et al., 2003; Falush et al., 2007; Pritchard et al., 2000) utilizando el modelo “Admixture” y realizando corridas en duplicado. En cada corrida, el método de MCMC (Markov Chain Monte Carlo) con 10,000 iteraciones

fue seleccionado. El valor medio más alto de probabilidad se comparó para múltiples corridas de $5 \leq K \leq 7$.

5.2.2 Análisis filogenético por búsqueda de polimorfismos de nucleótido simple.

Para el análisis de estructura poblacional mediante SNP (polimorfismos de nucleótido simple) se utilizó el servidor en línea del “Center for Genomic Epidemiology” (<https://cge.cbs.dtu.dk/services/CSIPhylogeny/>) y aquí se seleccionó el método “CSI phylogeny 1.4” (Kaas et al., 2014), utilizando por defecto el modelo evolutivo GTR. Se usaron los parámetros determinados por defecto, posteriormente, se incrementó la profundidad de las lecturas alineadas al polimorfismo se incrementó el z-score y la calidad de la lectura mapeada con el fin de comparar los distintos resultados. Se utilizó la cepa *H. pylori* 26695 (GenBank: NC_000915.1) para realizar las comparaciones y adicionalmente se incluyeron en los análisis otras 45 cepas de referencia de *H. pylori* previamente clasificadas en los grupos poblacionales mayores y se incluyeron también 24 cepas de Nicaragua y 32 de Perú. Se decidió incluir estas últimas cepas ya que, de acuerdo con estudios recientes, en estos países se ha descrito la generación de grupos poblacionales exclusivos que derivaron de la población europea (Gutiérrez-Escobar et al., 2020; Thorell, Yahara, et al., 2017).

Los árboles filogenéticos que se generaron, se editaron con la herramienta iTol v4 (Letunic & Bork, 2019).

5.2.3 Estimación del Pangenoma

Para estos análisis se utilizaron en total 259 genomas de *H. pylori*. El conjunto de datos total contenía 211 genomas colombianos (se excluyeron 10 genomas que no contaban con anotación en el NCBI), además 48 cepas de referencia de *H. pylori* con caracterización filogenética previa (6 hpWAfrica, 3 hpAfrica2, 11 hpEastAsia, 9 hspAmerind, 11 hpEurope y 2 hpSouthIndia). Se utilizó la anotación obtenida del canal de anotación automática de genomas procarióticos del NCBI (PGAAP) (Tatusova et al., 2016). Posteriormente se usó Roary versión 3.6.8 (Page et al., 2015) para evaluar el núcleo y el pangenoma, se utilizó un umbral de identidad del 90% y el parámetro -s para no separar parálogos en este umbral de identidad. Para la salida de la

matriz de presencia / ausencia de genes se definieron 4 grupos diferentes de genomas con respecto al porcentaje de cepas que presenta cada gen: core (>99%), soft core (>95% - 99%), shell (>15 - 95%) y accesorio (0,1% - <15%).

Los datos obtenidos de Roary para los genes ortólogos presentes en cada genoma se agruparon por las poblaciones filogenéticas obtenidas en los análisis de SNP's, utilizando esta información, se construyó un diagrama de Venn para analizar la relación de los genomas clasificados en la población hpEurope, hpAfrica, y las posibles nuevas poblaciones colombianas. Este diagrama fue construido usando la herramienta "VENN DIAGRAMS" disponible en la plataforma en línea "Bioinformatics and Evolutionary Genomics" (<http://bioinformatics.psb.ugent.be/webtools/Venn/>).

5.3 RESULTADOS

Para describir las características filogenéticas de los genomas colombianos de *H. pylori*, se analizaron 221 genomas colombianos. Los genomas analizados mostraron una media de 84,5 contigs, 38,9% de %GC, 1644 CDS y 1.652.045 pb de tamaño. Las estadísticas completas para estos genomas se proporcionan en la Tabla 1 del capítulo 2.

5.3.1 Análisis filogenético por MLST

El análisis MLST que se realizó usando el software Structure (Figura 4) mostró que $K = 6$ era el mejor valor para definir el número de poblaciones. Las poblaciones definidas fueron consistentes con la asignación previa de poblaciones principales, y estas se utilizaron para clasificar los nuevos genomas analizados, es decir los que provenían de aislamientos colombianos. Para estos genomas colombianos, el análisis MLST reveló que la mayoría de las cepas de Colombia (90.5%, 200/221) pertenecían a la población hpEurope, mientras que solo el 9,5% (21/221) fue clasificado en la población HpAfrica1. El árbol filogenético, construido usando las secuencias de los siete genes housekeeping (Figura 5), fue consistente con los hallazgos obtenidos en Structure. Ninguno de los genomas colombianos fue agrupado en las poblaciones hpAsia o hspAmerindn.

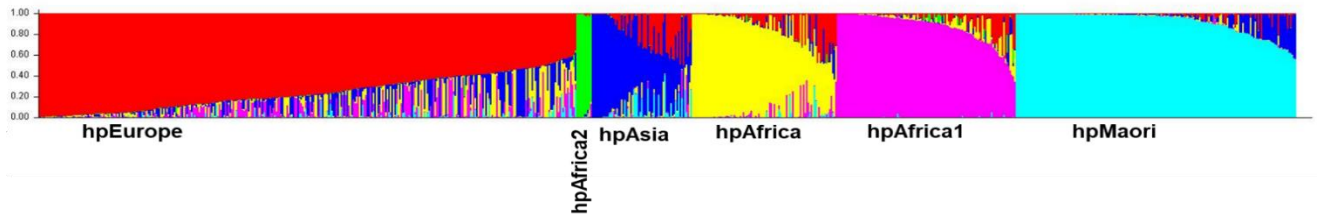


Figura 4. Poblaciones de *H. pylori* por análisis de MLST en el software Structure.

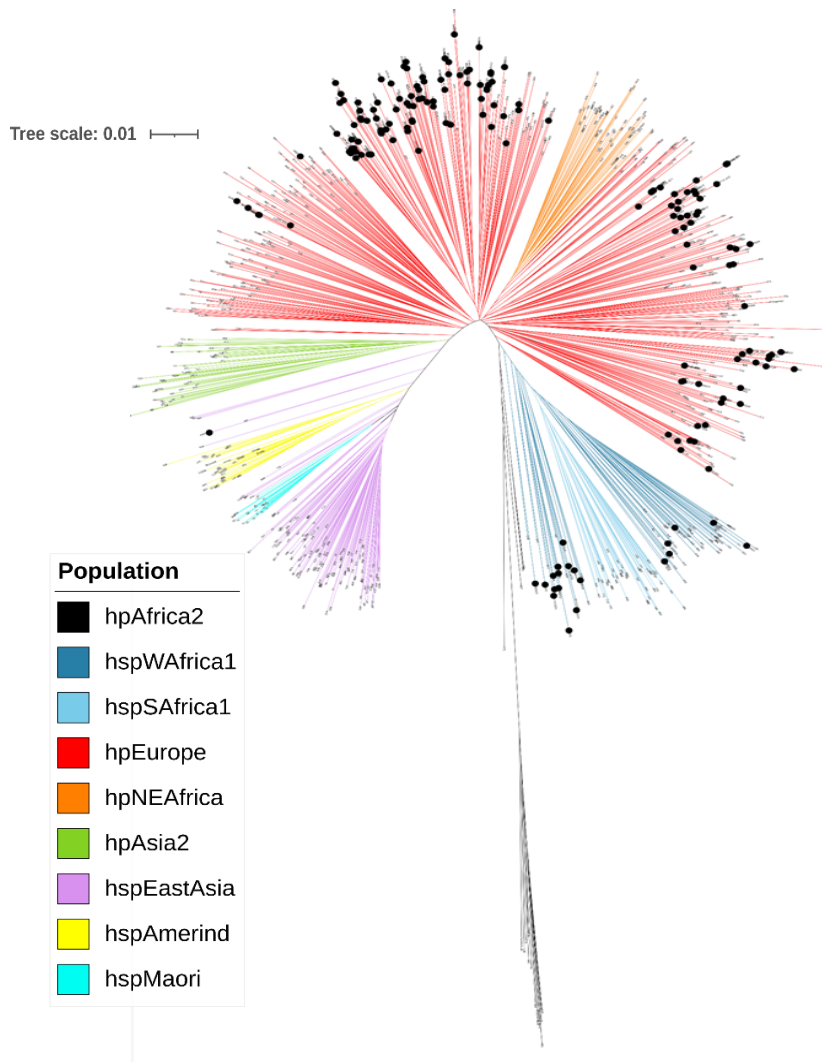


Figura 5. Análisis filogenético por MLST de los genomas colombianos y 741 genomas de referencia. Los puntos negros indican los 221 genomas colombianos analizados. El árbol se construyó usando el método de neighbor-joining y el modelo de sustitución de Kimura dos parámetros y posteriormente fue editado en la herramienta iTOL.

Análisis filogenético por SNP's

El árbol filogenético creado por la búsqueda de SNP's de los 221 genomas colombianos y 111 genomas de otras regiones geográficas (incluyendo 45 genomas de referencia y 55 genomas de Nicaragua y Perú los cuales mostraron separación filogenética en estudios previos) se puede observar en la Figura 6 donde se muestra que, de los 221 genomas colombianos, 36 se clasificaron como hpAfrica1, 46 se clasificaron como hpEurope y los 139 genomas colombianos restantes formaron dos grupos independientes, de estos 131 fueron clasificados en el grupo Colombia 1 y 8 en el grupo Colombia 2. Ninguno de los genomas colombianos se agrupó con las poblaciones hspAmerind o hpAsia. Curiosamente, el grupo Colombia 2 se ubicó en el medio de los genomas peruanos y nicaragüenses, los cuales habían sido clasificados previamente como nuevos grupos poblacionales derivados de la población hpEurope.

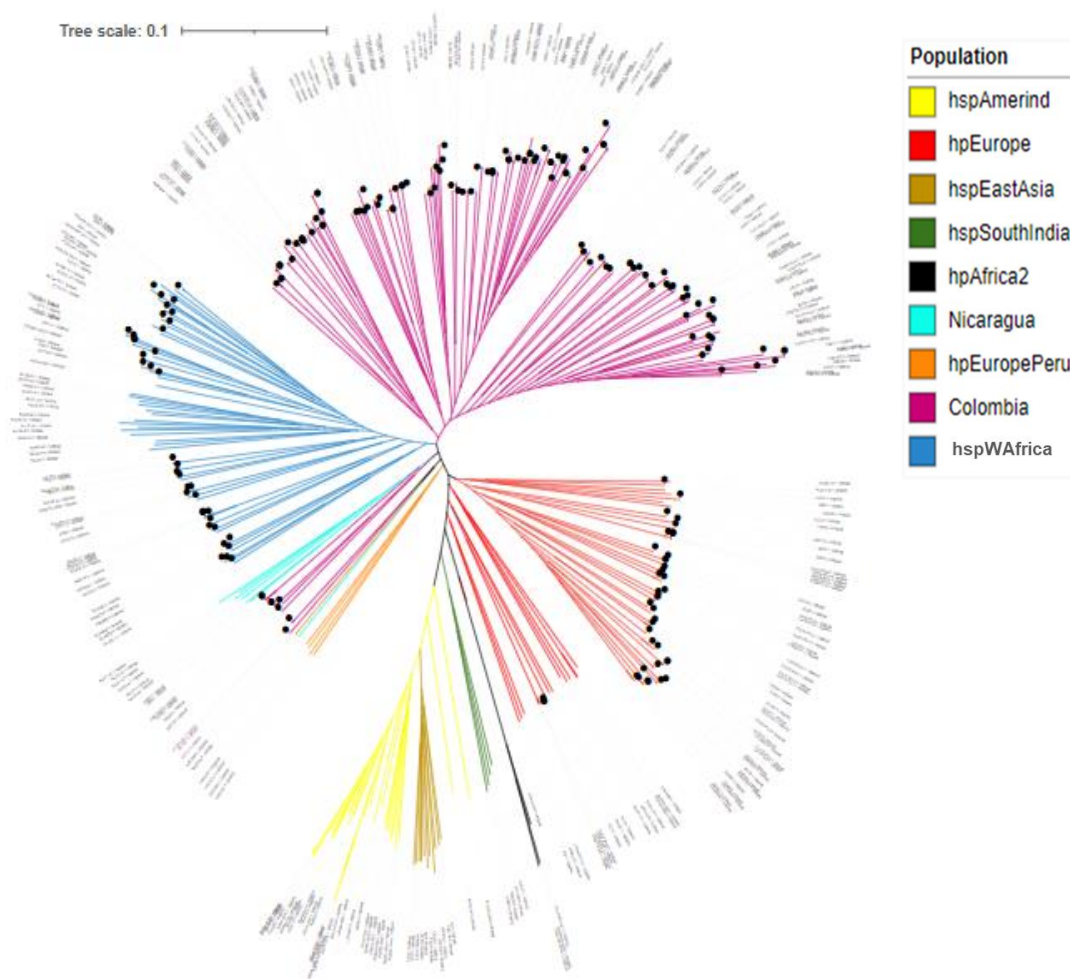


Figura 6. Caracterización filogenética por SNP de los genomas colombianos.

Los genomas colombianos se señalan con un punto negro. El árbol se construyó y editó en la herramienta iTOL.

Caracterización del pangenoma

El tamaño del pangenoma para los genomas analizados se estimó en 12776 genes, de los cuales, 446 genes correspondían al core, 307 al soft core, 1213 genes al grupo Shell y 10810 genes estuvieron en el grupo de genes accesorios (Figura 7)

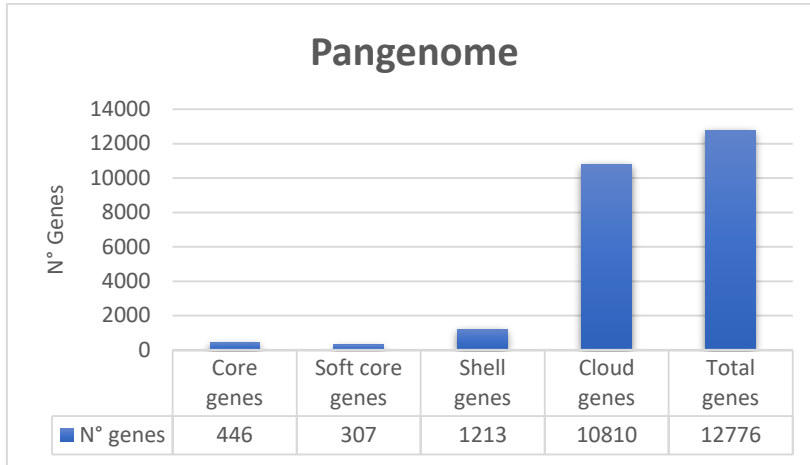


Figura 7. Caracterización del pangenoma colombiano de *H. pylori*

Los análisis de Roary además permitieron verificar que el pangenoma de *H. pylori*, en este caso para genomas colombianos, es un genoma abierto (Figura 8). Esto es evidente principalmente porque el número de genes siempre es creciente en la medida en que se van agregando genomas al análisis.

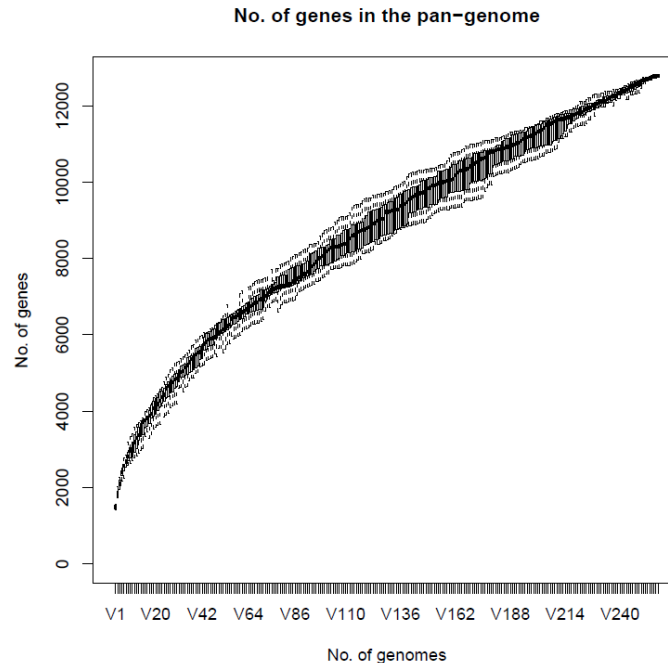


Figura 8. Número de genes que hacen parte del pangenoma de acuerdo con el número de genomas incluidos en el análisis.

A pesar de ser pangenoma abierto, el crecimiento en el número de genes siempre se encuentra a expensas de genes únicos, es decir genes clasificados en las categorías Shell o Cloud (Figura 9).

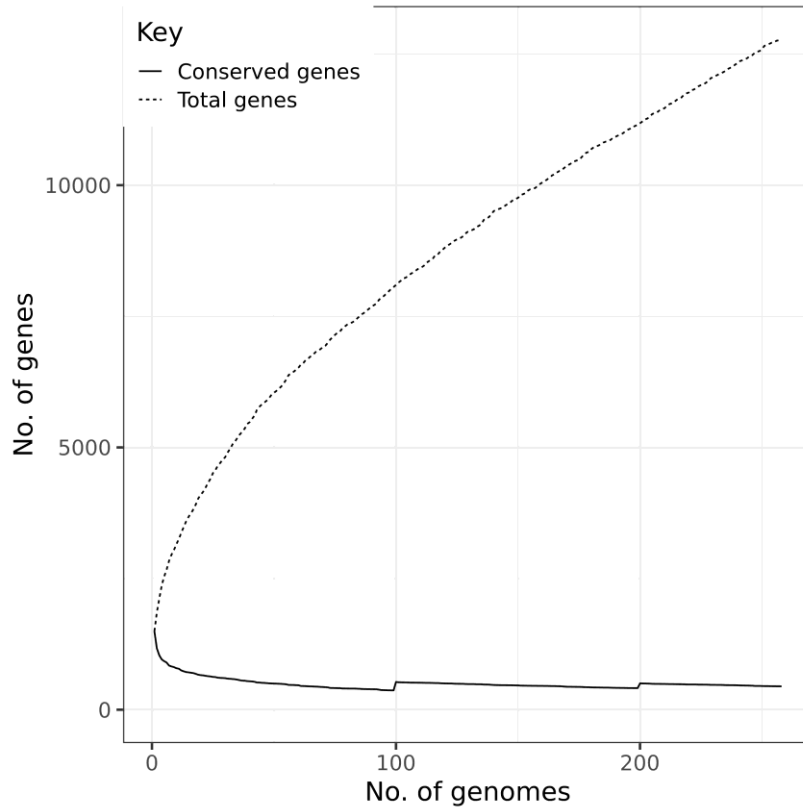


Figura 9. Número total de genes respecto a número de genes conservados en función del número de genomas analizados.

Al agrupar los resultados de Roary por poblaciones filogenéticas se encontraron diferentes números de genes entre las poblaciones, diferencia que se debe principalmente al número de genomas agrupados en cada población. (Tabla 15).

Tabla 15.

Análisis de pangenoma por grupo poblacionales

List names	number of genomes	number of orthogroups
Colombia1	124	7709
Colombia2	8	2661
HpEurope	60	6256
HpWAfrica	38	5022

HpAfrica2	3	2514
HspAmerind	9	2691
HpEastAsia	11	2843
HpEuropePeru	6	2200

Para la construcción del diagrama de Venn se seleccionaron los grupos de genes presentes solo en los genomas colombianos, es decir, los grupos denominados Colombia 1, Colombia 2, hpEurope, y hpWAfrica. Figura 10.

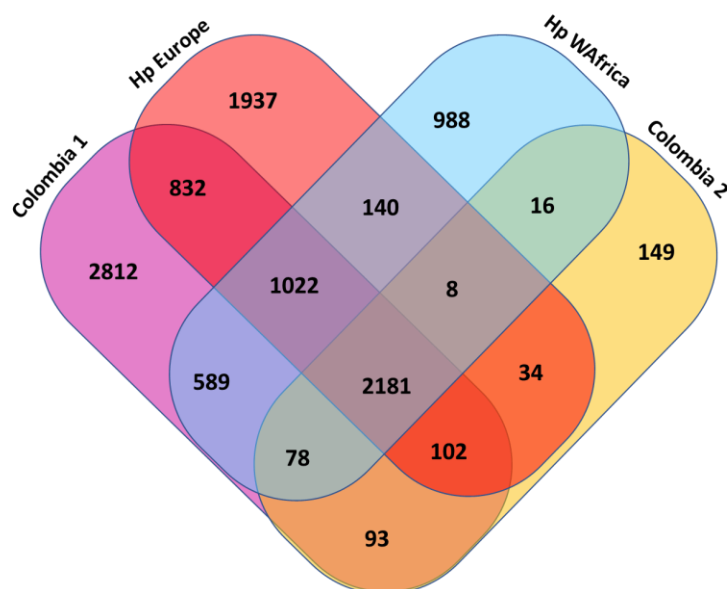


Figura 10. Diagrama de Venn para el análisis de pangenoma en los diferentes grupos poblacionales de los genomas colombianos analizados.

Esta comparación obtenida por el diagrama de Venn identificó 10981 grupos de genes ortólogos, donde 2181 fueron compartidos entre todos los genomas. En el grupo Colombia 1 se encontraron 2812 grupos de genes exclusivos de esta población, mientras que para el grupo Colombia 2 se encontraron 149. Al analizar estos grupos de genes presentes solamente en los nuevos grupos poblacionales colombianos, en el grupo Colombia 1, se encontraron 540 genes que solo estaban presentes en un genoma, los demás estaban presentes en bajas proporciones (<20% de los genomas del grupo); en grupo Colombia 2, los 149 genes exclusivos de esta población estuvieron presentes solo en un genoma. Al revisar la anotación de estos genes

exclusivos de estos 2 grupos poblacionales se encontró que la mayoría de corresponden a genes que codifican para proteínas hipotéticas.

5.4 DISCUSIÓN

Los estudios de caracterización filogenética de *H. pylori* hecha por análisis de MLST han permitido demostrar que existen siete poblaciones diferentes de *H. pylori* bien establecidas: hpEurope, hpNEAfrica, hpAfrica1, hpAfrica2, hpAsia2, hpSahul y hpEastAsia. Estos análisis adicionalmente han permitido verificar la coevolución existente entre *H. pylori* y el hombre (Achtman et al., 1999; Falush, Wirth, et al., 2003; Linz et al., 2007; Moodley et al., 2009). Se sabe adicionalmente que la población de hpEastAsia se divide en tres subpoblaciones, hspEAsia, hspMaori y hspAmerind (Achtman et al., 1999; Devi et al., 2007; Falush, Wirth, et al., 2003; Linz et al., 2007; Moodley et al., 2009; Yamaoka, 2009). La subpoblación hspAmerind refleja la migración de Asia a América a través del Estrecho de Bering que comenzó hace 12.000 años (Kersulyte et al., 2010). Sin embargo, este grupo poblacional fue reemplazado en la población mestiza americana debido a la entrada de nuevos patógenos, entre ellos *H. pylori*, hace 500 años, durante el proceso migratorio de colonización española que también trajo consigo, la llegada de población africana (de Sablet et al., 2011). Los resultados obtenidos en el presente estudio concuerdan con esta teoría, ninguno de los genomas colombianos analizado fue clasificado dentro de la población hpAmerind. De hecho, los resultados obtenidos por MLST mostraron que las cepas analizadas pertenecen principalmente a poblaciones de hpEurope y, en menor medida, de hpAfrica. Resultados similares a lo obtenido previamente por otros estudios que analizaron la filogenia en aislamientos colombianos (de Sablet et al., 2011; Shiota et al., 2014).

Estudios más recientes y que han utilizado estrategias de análisis del genoma completo, han demostrado la existencia de un componente poblacional específico en genomas de cepas aisladas de la población mestiza colombiana (Gutiérrez-Escobar et al., 2017; Muñoz-Ramírez et al., 2020; Muñoz-Ramírez et al., 2017; Thorell, Yahara, et al., 2017), lo que sugiere que las cepas colombianas de *H. pylori* han seguido vías evolutivas únicas (Muñoz-Ramírez et al., 2017) y han formado nuevas subpoblaciones a partir de una fuente europea, al parecer localizada en la Península Ibérica (Thorell, Yahara, et al., 2017). Es interesante que esta

nueva subpoblación colombiana reportada en los estudios recientes mostró predominio de la subpoblación europea hspSWEurope (Muñoz-Ramírez et al., 2020). Curiosamente y contrario a este hecho, para los profagos presentes en los aislamientos colombianos de *H. pylori* se encontró un nuevo grupo filogenético (presentado a continuación en el capítulo 4), derivado de la subpoblación hpNEurope.

Los resultados obtenidos aquí mediante los análisis de SNP's muestran la presencia de dos grupos poblacionales compuestos por genomas de origen colombiano. El 62,8% (139/221) de los genomas colombianos analizados, fueron ubicados en alguno de estos nuevos grupos, predominantemente en el grupo Colombia 1, en el cual se encontraron 131 genomas. Este hallazgo parece estar en concordancia con los recientes reportes, que demuestran la presencia de una subpoblación colombiana. Los genomas colombianos que no hicieron parte de estos nuevos grupos fueron clasificados dentro de las poblaciones hpEurope (20,8%) y hspAfrica2 (16,2%), lo cual es concordante con los hallazgos del análisis de MLST.

Estos resultados no permiten hacer inferencias sobre las líneas evolutivas que han dado origen a la generación de los nuevos grupos poblacionales en Colombia, sin embargo, los reportes hechos por Muñoz-Ramírez en el 2020 mostraron mediante el análisis de estructura poblacional y midiendo la proporción de ascendencia de cada aislamiento, se podía afirmar que las subpoblaciones americanas han evolucionado a partir de poblaciones europeas (Muñoz-Ramírez et al., 2020). Los datos obtenidos en el presente estudio muestran que los genomas que no se clasificaron en los grupos colombianos, fueron agrupados en su mayoría en el grupo hpEurope, lo que respalda los hallazgos de otros autores que sugieren que hpColombia diverge de una población europea, cuya prevalencia es residual.

Se encontró que el grupo Colombia 2 se ubicó en medio de un grupo de genomas que habían sido previamente clasificados como hpEuropaPeru (Gutiérrez-Escobar et al., 2020) y un grupo de genomas nicaragüenses que también habían sido agrupados en un nuevo grupo poblacional en un estudio anterior (Muñoz-Ramírez et al., 2017), los cuales habían sido clasificados previamente dentro de un nuevo grupo poblacional derivado de la población hpEurope. Sería muy valioso poder realizar nuevos estudios que incluyan una muestra mucho más completa, cubriendo varios países de América y de Europa, y que incluya poblaciones mestizas e indígenas de la población Latinoamericana para poder asegurar con certeza

cuantas y cuáles son las nuevas subpoblaciones circulantes en este momento en el continente. Así mismo, incluir en los estudios un análisis de tipo poblacional para determinar relaciones evolutivas con el hospedero humano de cada región, podría ayudar en la búsqueda de nuevas formas de contrarrestar la infección por *H. pylori* y sus efectos deletéreos.

Por otra parte, se sabe que aunque el genoma de *H. pylori* sea relativamente pequeño (1,6 Mbp, con aproximadamente 1500 marcos de lectura abiertos), su diversidad es bastante amplia. Por esta razón, aquí se propuso caracterizar el pangenoma de los genomas colombianos de *H. pylori*. Se encontró el tamaño del pangenoma fue de 12776 genes, con un core de 446 genes. Estudios previos han estimado que el core genoma de *H. pylori* consta de solo 244 genes (van Tonder et al., 2014) y se ha sugerido que *H. pylori* tiene un pangenoma abierto (Fischer et al., 2010; Kawai et al., 2011). Los hallazgos obtenidos aquí también mostraron dicha característica en el pangenoma colombiano.

Estas características que muestran un pangenoma abierto, grande y diverso, y además con presencia de muchos genes únicos, se debe principalmente a que muchos genes de *H. pylori* están permanentemente sujetos a una fuerte presión de selección debido a su interacción con el hospedero, principalmente por la necesidad de buscar mecanismos de evasión del sistema inmunológico, lo cual favorece la diversidad génica y el crecimiento del pangenoma. Dentro de los genes más diversos se pueden encontrar por ejemplo, algunos factores de virulencia, principalmente las proteínas que se ubican en la membrana y participan en la adherencia y la patogénesis de la bacteria, las cuales pueden sufrir varios cambios como producto de su relación con el hospedero (Rubio & Pérez-Pulido, 2021). Otros factores de virulencia como la oncoproteína CagA y el sistema de secreción tipo IV también tienen un papel muy importante en la patogenicidad, pero no hacen parte de los genes indispensables de *H. pylori* (Turkina et al., 2015). Finalmente, existe otra razón de encontrar pangenomas tan grandes y diversos: muchas proteínas codificadas por el genoma de *H. pylori* se anotan como proteínas desconocidas con funciones no caracterizadas (Park et al., 2012).

Aquí se analizaron las relaciones de los genes que componen el pangenoma de *H. pylori* de origen colombiano. Este análisis se realizó en función de los diferentes grupos filogenéticos en los cuales fueron ubicados los genomas analizados. Se encontraron 2 nuevos grupos filogenéticos, pero al analizar el componente génico exclusivo de estos grupos se encontró

que en su mayoría correspondían a genes presentes en un único genoma; de acuerdo con otros hallazgos previos, estos genes podrían ser los genes que se encuentran sujetos a presión selectiva, que codifican principalmente proteínas de membrana que pueden interactuar con el hospedero. Sin embargo, también podrían ser genes falsos predichos erróneamente por buscadores de genes (Rubio & Pérez-Pulido, 2021). El análisis hecho en el presente estudio mostró que la mayoría de estos genes únicos no tenían caracterización funcional, es decir que su anotación correspondía con proteínas hipotéticas o de función desconocida.

Los resultados obtenidos aquí fueron producto del análisis de 221 genomas colombianos de *H. pylori*; sin embargo, la mayoría de los genomas correspondía a genomas incompletos, esto debido a que solo hay 3 genomas colombianos completos depositados en bases de datos. En el futuro, se recomienda repetir los análisis con un conjunto de datos más robusto, que incluya un mayor número de genomas completos, en el cual se eliminen los genomas muy fragmentados (>50 contigs) y realizando desde cero los procesos de anotación para prevenir posibles errores en la determinación del número de genes.

5.5 CONCLUSIONES

Se comprobó mediante análisis filogenético que en Colombia existe una nueva línea evolutiva de *H. pylori* la cual procede del grupo poblacional hpEurope.

El análisis de pangenoma de *H. pylori* de origen colombiano mostró ser un pangenoma, abierto y muy diverso. No se encontró en el análisis del pangenoma, la presencia de algún gen marcador de distribución geográfica.

6. CAPITULO 4.

CARACTERIZACIÓN DE PROFAGOS DE *H. pylori* EN AISLAMIENTOS COLOMBIANOS

Los resultados presentados en este capítulo también pueden ser revisados en la publicación (Anexo 3):

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>

6.1 INTRODUCCIÓN

Los estudios han demostrado que alrededor del 20% de las cepas de *H. pylori* portan secuencias de profago parciales o completas, y que analizar las secuencias de genes integrasa y holin permite la diferenciación geográfica de las cepas (Lehours et al., 2011; Secka et al., 2017; Vale et al., 2015). Así mismo, se ha demostrado que *H. pylori* exhibe una amplia diversidad intraespecie (Thorell, Lehours, et al., 2017) y que esta diversidad parece estar influenciada por la presencia de elementos genómicos móviles, como los profagos (Vale & Lehours, 2018).

El método Prophage Sequence Typing (PST) informado por Vale et al. en 2015 (Vale et al., 2015) se basa en usar la secuencia de los genes integrasa y holin para predecir la presencia de profagos en cepas de *H. pylori* y para caracterizar sus estructura filogenética. Utilizando este método hasta ahora se han descrito cuatro poblaciones de profagos de *H. pylori*: hpAfrica1, hpEastAsia, hpNEurope y hpSWEurope, que son comparables a las identificadas en bacterias mediante MLST (Vale et al., 2015).

Los estudios hasta ahora disponibles para la búsqueda y caracterización de profagos no habían incluido cepas de *H. pylori* del continente americano, situación relevante teniendo en cuenta que las revisiones recientes de los genomas de *H. pylori* de América Latina han mostrado una separación filogenética. Además, se han detectado eventos de recombinación frecuentes y recientes en grupos de cepas exclusivamente colombianas, mexicanas y

nicaragüenses (Muñoz-Ramírez et al., 2020; Muñoz-Ramírez et al., 2017). Los nuevos linajes de *H. pylori* presentes en el continente americano al parecer resultan de la deriva genética y la mezcla entre *H. pylori* de origen europeo y africano (Thorell, Yahara, et al., 2017).

En este estudio, se examinaron 213 aislamientos colombianos de *H. pylori* para determinar la prevalencia de los genes del profago integrasa y holin, así como para identificar la población filogenética de los profagos presentes en estos aislamientos. También se utilizaron genomas colombianos de *H. pylori* disponibles en bases de datos públicas para enriquecer nuestra colección. Estos análisis pretenden brindar información sobre la prevalencia de profagos en cepas colombianas, su estructura poblacional y origen ancestral.

6.2 METODOLOGÍA

6.2.1 Cepas de *H. pylori*

La presencia de genes de profago holin e integrasa fue detectada por PCR. Se utilizó un banco de muestras de ADN procedente de 213 cepas de *H. pylori* aisladas de pacientes diagnosticados de gastritis crónica y residentes en Bogotá, Cundinamarca Colombia. El banco de ADN se obtuvo de estudios previos realizados en el grupo de investigación de Enfermedades Infecciosas de la Pontificia Universidad Javeriana en Bogotá, Colombia. En el momento de la obtención de las cepas, el ADN total se extrajo usando DNAzol® (Invitrogen, Carlsbad, CA, USA) y siguiendo los protocolos estandarizados (Trespalcios et al., 2015). Las muestras se encontraban almacenadas a -20°C para su posterior uso.

Para la caracterización de los profagos, las cepas que resultaron positivas para los genes integrasa y/o holin fueron recuperadas por cultivo. Siguiendo los protocolos estandarizados (Trespalcios et al., 2013), se logró recuperar ocho cepas (Tabla 1) (Col-2PUJ, Col-5PUJ, Col-6PUJ, Col-7PUJ, Col-16PUJ, Col-17PUJ, Col-22PUJ, y Col 23-PUJ). A partir de estas cepas, se realizó nueva extracción de ADN total utilizando el kit DNeasy Blood & Tissue (QIAGEN, Hilden, Germany) y siguiendo las instrucciones del fabricante. Se realizó cuantificación del ADN extraído utilizando el fluorómetro Qubit 2.0 (Life Technologies, Carlsbad, California, US) y el kit Qubit dsDNA HS Assay Kit 0.2–100 ng/ μl , siguiendo las instrucciones del fabricante. Las muestras de ADN obtenidas fueron almacenadas a -20°C

para ser secuenciadas posteriormente mediante la tecnología de next-generation sequencing (NGS).

Para la secuenciación por NGS se usó la plataforma Illumina MiSeq (Illumina, San Diego, CA); las librerías de ADN fueron preparadas usando el kit Nextera XT DNA Library Preparation (Illumina), seguido de corridas de 2×300 paired-end con una cobertura de $\sim 80X$. Las lecturas fueron usadas para realizar un ensamblaje *de novo* utilizando SPAdes v13.3 (Bankevich et al., 2012).

Adicionalmente, cepas de *H. pylori* portadoras de los genes integrasa y holin fueron recuperadas de la base de datos del National Center for Biotechnology Information (NCBI) (Tabla 2). Para esto, un análisis de BLASTn fue realizado con las cepas de *H. pylori* disponibles en la base whole-genome shotgun (WGS) del NCBI. Para esto se tuvo en cuenta un umbral de $<1e^{-6}$.

6.2.2 Búsqueda de profagos de *H. pylori*

Las 213 muestras de ADN genómico fueron analizadas por la técnica de PCR para la búsqueda de los genes holin en integrasa. Se usó GoTaq® Green Master Mix kit (Promega, Madison, WI) y se usaron los iniciadores (Tabla 16) y los protocolos previamente descritos por Vale et al. (Vale et al., 2015). El ensayo de PCR se realizó con las siguientes condiciones: Denaturación inicial a 95 °C por 4 minutos; seguido de 35 ciclos de 95 °C por 30 segundos, 58 °C por 30 segundos y 72 °C por 1 minuto, y una extensión final a 72 °C por 7 minutos. Como control positivo se empleó una muestra de ADN de la cepa *H. pylori* B45 (primera cepa reportada como portadora de profago) (Lehours et al., 2011), la cual fue donada por el - Centre National de Référence des Campylobacters et Hélicobacters-, Bordeaux, Francia.

Tabla 16. Iniciadores utilizados para la detección de los genes holin e integrasa.

Gen	Iniciador	Secuencia 5' – 3'	Tamaño Producto amplificado
Integrasa	Forward	AAGYTTTTTATAGMGTTTTGYG	529 pb
	Reverse	CGCCCTGGCTTAGCATC	
Holin	Forward	CCATCCCGTATTTGTTGGTG	225 pb
	Reverse	ACCCAATGCCTCCACTAATC	

Los productos de las cepas que resultaron positivas por PCR para integrasa y/o holin fueron purificados y bidireccionalmente secuenciados usando un secuenciador ABI PRISM 3500® (Applied Biosystems, Foster City, CA).

Para identificar homología con profagos de *H. pylori* previamente descritos, los genomas fueron analizados usando MEGABLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), utilizando como referencia los fagos KHP30 (número de acceso: NC_019928.1) y phiHP33 (NC_016568.1). Los genomas de los profagos fueron puestos juntos manualmente por mapeo de contigs con los genomas de los fagos de referencia.

El genoma del fago phiHP33 y de otros 15 fagos previamente descritos por Vale *et al.* (Vale *et al.*, 2017) fueron usados como referencia para alinear y comparar los fagos encontrados en las cepas colombianas. El software MAUVE fue usado para realizar el múltiple alineamiento. Además, el fago phiHP33 y uno (1) fago de cada población filogenética previamente descrita (hpNEurope, hpSWEurope, hpEastAsia, y hpAfrica1) fueron seleccionados para construir un cuadro de comparación usando BLASTn y el software Easyfig.

6.2.3 Análisis de Prophage Sequence Typing (PST)

Para determinar la estructura poblacional de los profagos por medio de PST se incluyeron: la secuencia de los genes holin e integrasa de las cepas colombianas positivas, de las cepas seleccionadas desde el NCBI y de 95 cepas previamente caracterizadas (Secka *et al.*, 2017; Vale *et al.*, 2017) (secuencias de los genes de un total de 167 cepas).

Todas las secuencias de los genes integrasa y holin se alinearon usando MAFFT versión 7 (Kato & Standley, 2013). A partir del alineamiento se obtuvo un archivo Multi-fasta el cual fue convertido al formato de archivo de entrada de Structure usando el programa xmf2struct (<http://www.xavierdidelot.xtreemhost.com/clonalframe.htm>). El archivo convertido fue analizado en el programa Structure 2.3.4 (Falush, Stephens, et al., 2003; Falush et al., 2007; Pritchard et al., 2000) para identificar el número de poblaciones (k). En Structure se usó el modelo “admixture” y se realizaron corridas en duplicado. En cada corrida, el método de MCMC (Markov Chain Monte Carlo) con 10,000 iteraciones fue seleccionado. El valor medio más alto de probabilidad se comparó para múltiples corridas de $2 \leq K \leq 6$. Para evaluar los resultados del análisis PST y visualizar los valores de probabilidad en múltiples valores de K (detectar el número de poblaciones que mejor se ajustan a los datos), se utilizó el método Evanno de la herramienta Structure Harvester (Earl & vonHoldt, 2012).

Adicional al análisis poblacional realizado en Structure, también se construyó un árbol filogenético en el Software “Molecular Evolutionary Genetic Analysis” Versión 7.0 (MEGA 7) (Kumar et al., 2016). Para esto se utilizaron los alineamientos de las secuencias de los genes integrasa y holin concatenados, juntos o individualmente, y se usó el método de Neighbor-joining y el modelo de sustitución de dos parámetros de Kimura (Kimura, 1980). El número de repeticiones bootstrap se estableció en 1000.

También se construyó otro árbol filogenético, siguiendo las mismas pautas, pero utilizando el genoma completo de los profagos encontrados en cepas colombianas (obtenidas de este estudio y de bases de datos), así como los fagos de referencia, PhiHp33, KHP30, India7 (número de acceso CP002331), Cuz20 (CP002076), 1961P (NC_019512.1), KHP40 (NC_019931.1) y 20 profagos descritos previamente (Vale et al., 2017); en total se analizaron en este caso 40 genomas de profagos.

6.3 RESULTADOS

6.3.1 Identificación de profagos de *H. pylori*.

La presencia de profagos en cepas colombianas de *H. pylori* se determinó inicialmente mediante la detección de los genes integrasa y holin. De las 213 ADN de *H. pylori* analizados

por PCR, 27 cepas (12,6%) fueron positivas para integrasa y 9 (4,2%) fueron positivas para ambos genes. Para el análisis completo del genoma, un profago se consideró intacto si su tamaño era de al menos 20 kb, como lo describieron otros (Lehours et al., 2011). Se secuenciaron por NGS ocho cepas y encontramos profagos completos en cinco cepas (Col-5PUJ, Col-6PUJ, Col-7PUJ, Col-22PUJ y Col-23PUJ) y profagos incompletos en tres cepas (Col-2PUJ, Col-16PUJ, y Col-17PUJ). Algunas de las cepas analizadas (Col-6PUJ y Col-7PUJ, Col-22PUJ y Col-23PUJ, Col-16PUJ y Col-17PUJ) habían sido aisladas del mismo paciente y eran idénticas (porcentajes de identidad superiores al 99% en todos los casos). Por tanto, una cepa de cada par idéntico se retiró de los análisis posteriores. Los genomas de las cepas y profagos incluidos en el estudio se sometieron al NCBI bajo el Bioproyecto número [PRJNA656306](https://www.ncbi.nlm.nih.gov/bioproject/656306). Los números de acceso de los genomas bacterianos se pueden observar en la tabla 1 del capítulo 2, los números de acceso para los genomas de profagos se incluyen en la tabla 17.

Tabla 17.

Características de los profagos reportados al NCBI.

Profago	Genbank N° acceso	Tamaño (pb)	CDS
Col 2-PUJ	MW247143	19298	21
Col 5-PUJ	MW247144	26486	31
Col 6-PUJ	MW247145	26487	31
Col 16-PUJ	MW247146	19744	21
Col 23-PUJ	MW247147	28622	38

Los profagos intactos tenían un promedio de 33 genes predichos, 28,6 kb de longitud y 37% de GC. Los resultados de BLASTn y Easyfig se pueden observar en la Figura 11, allí se puede observar que se encontró sintenia entre los fagos descritos anteriormente y los profagos intactos hallados en cepas colombianas. Así mismo, los resultados de la alineación múltiple mostraron que los fagos colombianos tienen longitudes y organización génica similares a las

descritas para los profagos reportados previamente (Figura 11). En los profagos incompletos, es probable que se perdieran marcos de lectura abiertos específicos en el ensamblaje del genoma, lo que lleva a genomas de menos de 20 kb. En todos los casos, la parte faltante correspondió a la parte final del profago.

Los procedimientos de anotación y alineación utilizados permitieron la identificación de una secuencia de inserción (IS) en el fago Col 23-PUJ. Este IS codifica una enzima transposasa, se identificó como IS_{Cco1}, corresponde a un IS606 (tnpA) y es miembro de la familia de transposasas IS200 / IS605 (He et al., 2015; Kersulyte et al., 1998). Un análisis BLASTp reveló que esta IS se encuentra frecuentemente en diferentes cepas de *H. pylori* y sus profagos (Vale et al., 2017).

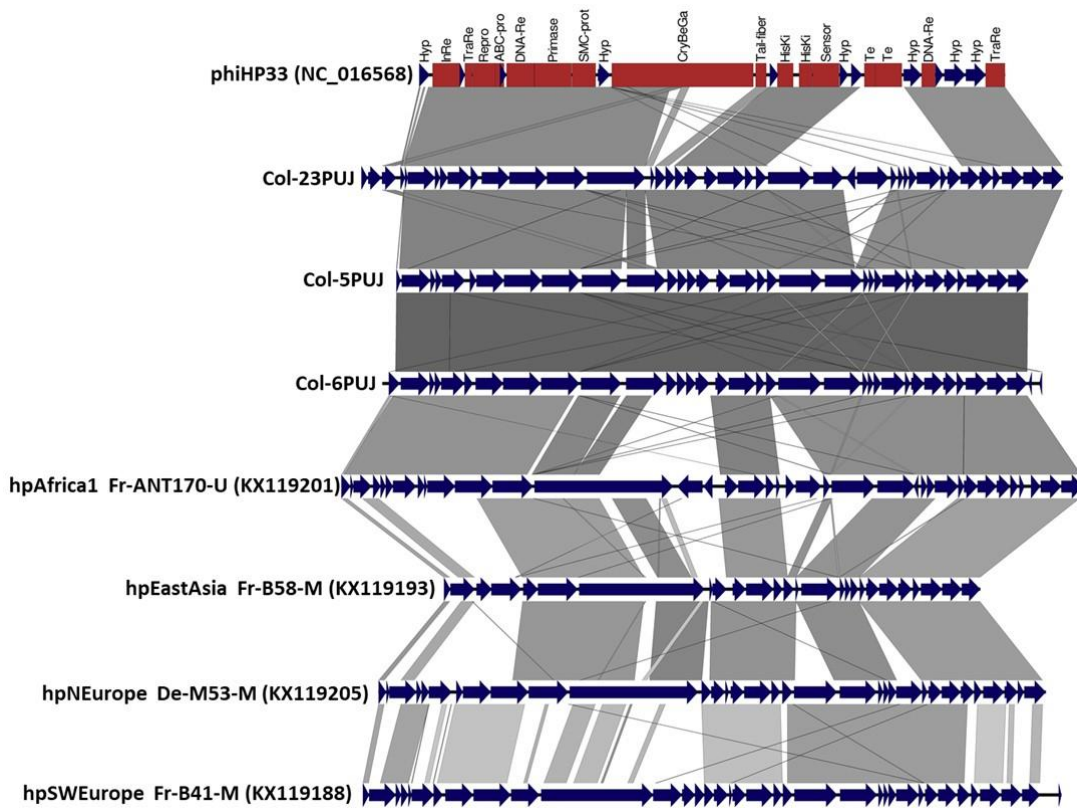


Figura 11. Alineamiento de los profagos colombianos de *H. pylori*. Se utilizaron como referencia, el fago phiHP33 y un fago para cada grupo poblacional previamente descrito. (Vale et al., 2017).

Caracterización de Profagos de H. pylori en Aislamientos Colombianos

Se determinó el sitio de inserción en el genoma bacteriano para los profagos de *H. pylori* colombianos (Tabla 18). En todos los profagos completos el sitio de inserción del profago solo se logró determinar solo en el extremo 5'. Esto debido a que, siempre al final del extremo 3', el contig siempre terminaba con secuencias de profagos. Para el profago incompleto Col 2-PUJ, no fue posible determinar el sitio de inserción en ninguno de los extremos. Los fagos completos, Col 5-PUJ y Col 6-PUJ se insertaron contiguos al gen que codifica una enzima de modificación del ADN de tipo II, el fago Col 23-PUJ se insertó contiguo al gen *RuvA* de la ADN helicasa dependiente de ATP de la unión Holliday, y el fago incompleto Col 16-PUJ se insertó adyacente al gen de fosfolípido fosfatasa asociado a la membrana. Para los profagos colombianos obtenidos de las cepas del NCBI, determinamos el sitio de inserción en el extremo 5' en cinco profagos. Cuatro de ellos (HP3076, CM22388, NQ4044 y HP3133) se insertaron junto al gen de la S-adenosilmetionina sintetasa; y el profago PZ5019-3A3 se localizó junto al gen que codifica el factor Trigger. Los sitios de inserción en el extremo 3' solo se determinaron en dos de estos profagos, Hp3076 y NQ4044; en ambos casos, el sitio de inserción estaba próximo al gen que codifica la UDP-3-O- [3-hidroximiristoil] glucosamina N-aciltransferasa.

Table 18. Sitios de Inserción para profagos colombianos.

Profago	PST	5'	3'	Fuente
Col 2-PUJ	hpColombia	ND	ND	Este estudio
Col 5-PUJ	hpAfrica1	Type II DNA modification	ND	Este estudio
Col 6-PUJ	hpAfrica1	Type II DNA modification	ND	Este estudio
Col 16-PUJ	hpColombia	Membrane-associated phospholipid phosphatase	ND	Este estudio
Col 23-PUJ	hpAfrica1	Holliday junction ATP-dependent DNA helicase RuvA	ND	Este estudio
HP3076	hpColombia	Methionine adenosyltransferase	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	NCBI
PZ5080	hpSWEurope	ND	membrane protein	NCBI
PZ5019_3A3	hpAfrica1	Trigger factor	ND	NCBI
CM22388	hpAfrica1	S-adenosylmethionine synthasa	ND	NCBI
NQ4044	hpAfrica1	Methionine adenosyltransferase	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	NCBI
3133	hpColombia	Methionine adenosyltransferase	ND	NCBI

ND = No determinado

6.3.2 Estructura poblacional de los profagos de *H. pylori*.

El análisis hecho en STRUCTURE 2.3.4 para PST, sorprendentemente, reveló que $K = 5$ tenía el mejor valor de probabilidad, lo que indica la presencia de cinco poblaciones. Las cinco poblaciones identificadas por PST fueron hpEastAsia, hpAfrica1, hpSWEurope, hpNEurope, y una quinta población nueva, que contiene solo cepas colombianas, la cual se denominó hpColombia (Figura 12, Tabla 19). La evaluación de estos resultados utilizando la herramienta Harvester Structure y el método Evanno (Earl & vonHoldt) confirmó que $K = 5$ era el mejor valor (Tabla 20).

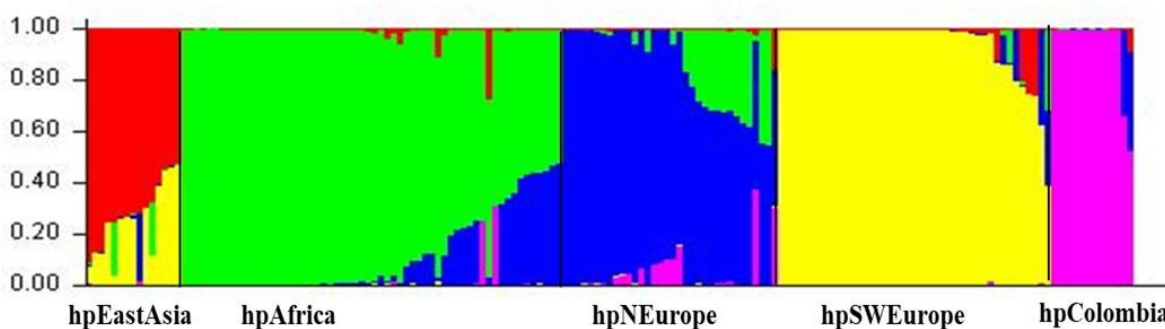


Figura 12. Poblaciones filogenéticas de los profagos de *H. pylori*. Análisis realizado usando el software Structure 2.3.4

Tabla 19. Resultados del análisis de PST en el software Structure

Strain	hpEastAsia	hpAfrica1	hpNEurope	hpSWEurope	hpColombia	PST
Col 2-PUJ	0.001	0.000	0.001	0.000	0.998	hpColombia
Col 3-PUJ	0.001	0.000	0.000	0.998	0.000	hpSWEurope
Col 4-PUJ	0.000	0.998	0.001	0.000	0.001	hpAfrica1
Col 5-PUJ	0.001	0.998	0.001	0.000	0.001	hpAfrica1
Col 6-PUJ	0.000	0.998	0.001	0.000	0.001	hpAfrica1
Col16-PUJ	0.002	0.001	0.001	0.000	0.997	hpColombia
Col 23-PUJ	0.001	0.871	0.126	0.001	0.001	hpAfrica1
Col 32-PUJ	0.000	0.001	0.001	0.000	0.998	hpColombia
Col 33-PUJ	0.001	0.744	0.001	0.000	0.254	hpAfrica1
Col 34-PUJ	0.000	0.001	0.001	0.000	0.998	hpColombia
Col 35-PUJ	0.001	0.000	0.000	0.998	0.000	hpSWEurope
Col 36-PUJ	0.001	0.000	0.000	0.998	0.000	hpSWEurope
Col 37-PUJ	0.000	0.000	0.001	0.000	0.998	hpColombia

Caracterización de Profagos de H. pylori en Aislamientos Colombianos

Col 38-PUJ	0.001	0.000	0.000	0.998	0.000	hpSWEurope
Col 39-PUJ	0.001	0.688	0.001	0.000	0.309	hpAfrica1
Col 40-PUJ	0.000	0.001	0.001	0.000	0.998	hpSWEurope
Col 41-PUJ	0.000	0.000	0.001	0.000	0.998	hpSWEurope
Col 42-PUJ	0.001	0.997	0.001	0.000	0.001	hpAfrica1
Col 43-PUJ	0.000	0.998	0.001	0.000	0.001	hpAfrica1
Col 44-PUJ	0.000	0.001	0.001	0.000	0.998	hpColombia
Col 45-PUJ	0.001	0.000	0.000	0.998	0.000	hpSWEurope
Col 46-PUJ	0.001	0.995	0.002	0.001	0.002	hpAfrica1
Col 47-PUJ	0.000	0.000	0.001	0.000	0.998	hpColombia
Col 48-PUJ	0.000	0.001	0.001	0.000	0.998	hpColombia
HP3076	0.086	0.001	0.388	0.002	0.523	hpColombia
PZ5080	0.003	0.002	0.001	0.993	0.002	hpSWEurope
PZ5019_3A3	0.004	0.896	0.093	0.001	0.005	hpAfrica1
CA22019	0.001	0.997	0.001	0.000	0.001	hpAfrica1
CA22362	0.002	0.000	0.001	0.996	0.000	hpSWEurope
CG22371	0.001	0.998	0.001	0.000	0.001	hpAfrica1
CM22388	0.001	0.636	0.361	0.001	0.001	hpAfrica1
3096	0.001	0.001	0.004	0.000	0.994	hpColombia
22386	0.127	0.001	0.001	0.871	0.000	hpSWEurope
NQ4044	0.001	0.558	0.433	0.002	0.005	hpAfrica1
NQ1712	0.003	0.324	0.665	0.000	0.007	hpNEurope
NQ315	0.002	0.317	0.667	0.001	0.013	hpNEurope
2036	0.001	0.998	0.001	0.000	0.000	hpAfrica1
22371	0.001	0.998	0.001	0.000	0.001	hpAfrica1
22341	0.001	0.997	0.001	0.000	0.001	hpAfrica1
3133	0.003	0.002	0.331	0.002	0.661	hpColombia
Fr-B41-M [‡]	0,001	0,001	0,002	0,996	0	hpSWEurope
Fr-B58-M [‡]	0,871	0,002	0,002	0,123	0,002	hpEastAsia
Uk-EN31-U [‡]	0,001	0,001	0,994	0,001	0,003	hpNEurope
Pt-4481-G [‡]	0,002	0,22	0,776	0,001	0,001	hpNEurope
Pt-212-U [‡]	0,001	0	0,001	0,997	0,001	hpSWEurope
Pt-1918-U [‡]	0,001	0,001	0	0,998	0	hpSWEurope
Pt-259_01-G [‡]	0,004	0	0,001	0,995	0,001	hpSWEurope
Pt-4497-U [‡]	0,001	0,002	0	0,996	0,001	hpSWEurope
Pe-Cuz20 [‡]	0,159	0,002	0,531	0,003	0,305	hpNEurope
Fr-ANT170-U [‡]	0,001	0,554	0,443	0,001	0,001	hpAfrica1

Caracterización de Profagos de H. pylori en Aislamientos Colombianos

Fr-MEG235-U [‡]	0,001	0,558	0,44	0,001	0,001	hpAfrica1
Pt-1846-U [‡]	0,001	0,278	0,712	0,008	0,001	hpNEurope
Pt-228_99-G [‡]	0,001	0,767	0,23	0,001	0,001	hpAfrica1
KHP30 [‡]	0,694	0	0,001	0,305	0,001	hpEastAsia

Solo se presentan los resultados para los profagos nuevos y los de referencia. La ancestría es inferida de acuerdo con un análisis Bayesiano de las secuencias de los genes integrasa y holin. La población más probable de cada profago es resaltada en gris.

[‡] -Profagos de Referencia.

Table 20. Resultados para el analisis en Structure Harvester[†].

K	REPS	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	Ln''(K)	Delta K
2	2	-26605,5	9,7581	-	-	-
3	2	-24054,9	46,3862	2550,6	1378	29,707108
4	2	-22882,3	40,7294	1172,6	487,5	11,969255
5	2	-22197,2	21,3546	685,1	648,9	30,38686
6	2	-22161	73,3977	36,2	-	-

[†] Structure Harvester es un programa que procesa y resume los datos obtenidos en Structure software. Este programa ejecuta el método “Evanno” y genera una tabla de los valores medios de verosimilitud por número de población (K). El resultado gris muestra el valor más grande en la columna Delta K. En este caso, indica que el número más probable de poblaciones es 5.

Los árboles filogenéticos neighbor-joining para los genes de integrasa y holin concatenados y para el genoma completo de los profagos también mostraron un nuevo agrupamiento al cual pertenecían exclusivamente profagos de cepas colombianas, lo que apoya la asignación de una nueva población (Figuras 13 y 14). En el árbol con los genomas completos (Figura 14), se puede observar que la mayoría de los profagos reunidos por el grupo filogeográfico se agrupan de acuerdo con su población asignada por el análisis PST. Sin embargo, se encontraron algunas excepciones, a saber, para Pt-4472-G, Fr-G12-G, Fr-CG43-A, Pt-21299R-U y Cuz20 que se agrupan con otra población de profagos.

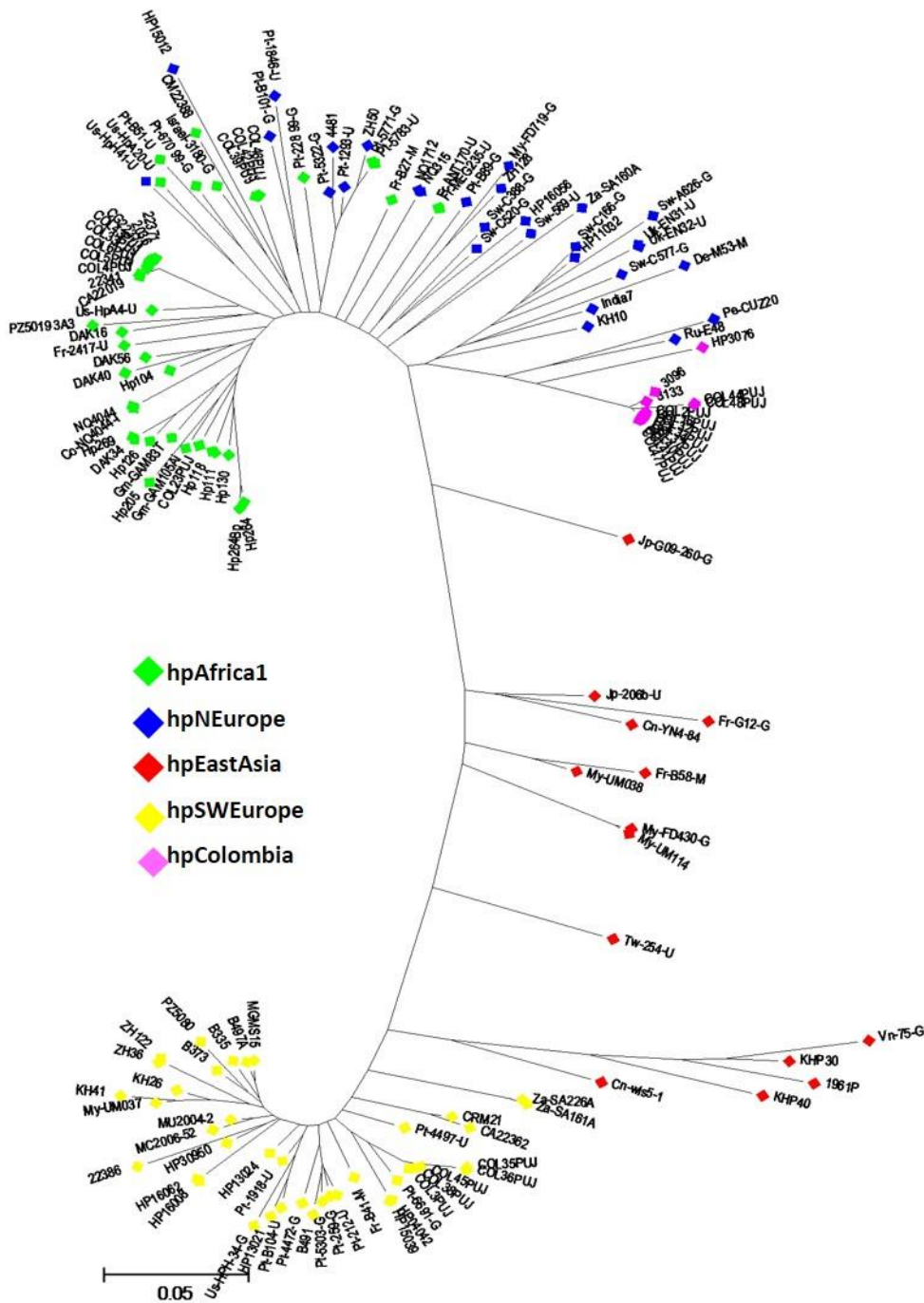


Figura. 13. Árbol filogenético de los genes de integrasa y holin. Se utilizaron 167 secuencias de profagos (27 secuencias colombianas obtenidas en este estudio, 45 secuencias de cepas seleccionadas por NCBI y 95 secuencias de profagos *H. pylori* previamente caracterizados).

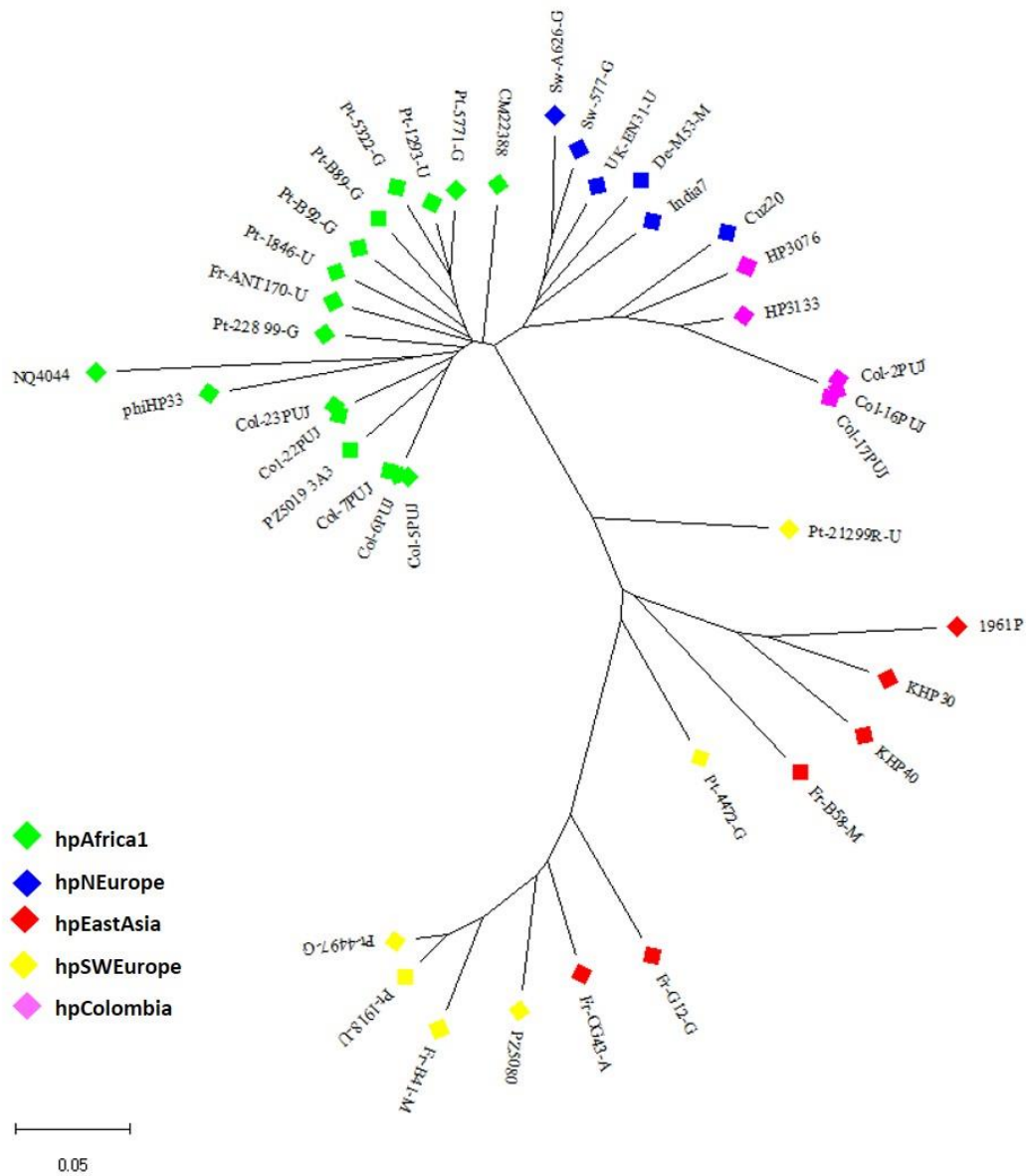


Figura 14. Árbol filogenético de genomas completos de 40 profagos. Se utilizaron ocho profagos colombianos encontrados en este estudio, seis profagos colombianos recuperados de bases de datos, 20 profagos descritos previamente y seis profagos de referencia (PhiHp33, KHP30, India7, Cuz20, 1961P, KHP40).

6.4 DISCUSIÓN

Este estudio describe por primera vez la identificación de profagos en cepas colombianas de *H. pylori*. Si bien un número creciente de estudios informa la presencia de estos profagos, esta es la primera descripción de estos elementos en aislamientos de origen americano. La prevalencia observada de genes integrasa y holin no coincidió con la previamente determinada. Los informes anteriores indican que los profagos están presentes en aproximadamente el 20% de las cepas de *H. pylori* (Yahara et al., 2019). Aquí se encontró que solo el 12,6% de las cepas colombianas fueron positivas para integrasa y solo el 4,2% fueron positivas para ambos genes. Esta situación fue similar a lo observado para los genomas colombianos de *H. pylori* recuperados de bases de datos públicas. Ya que de las 186 cepas colombianas analizadas de la base de datos Patric (Wattam et al., 2014), solo 17 (9,1%) resultaron positivas para el gen de la integrasa y 6 (3,2%) para ambos genes. Estos resultados sugieren una menor prevalencia de secuencias de profago en aislamientos de origen americano, lo que puede deberse a un cuello de botella de genomas de *H. pylori* pobres en secuencias de profago que llegaron a América desde Europa y África después del descubrimiento del continente. Alternativamente, pudo ser que las cepas de *H. pylori* que se adaptaron mejor a la población colombiana pueden haber sido aquellas sin profagos.

Los profagos identificados en las cepas colombianas de *H. pylori* son consistentes con los tamaños y características estimados para los fagos de *H. pylori* previamente reportados, lo cual puede sugerir un origen común. De hecho, los profagos colombianos mostraron sintenia genética entre ellos y con el fago phiHP33 (Lehours et al., 2011). Como se ha descrito para otros fagos de *H. pylori*, los ORF de los profagos se encontraron principalmente en la misma dirección, muchos ORF correspondían a proteínas hipotéticas o sin función conocida, y no se identificaron genes de virulencia. Todos los genes de “soft-core” de profagos (Yahara et al., 2019) estuvieron presentes en los profagos colombianos, aunque el gen holin presentó un porcentaje de identidad menor (72%). En conjunto, los resultados encontrados aquí apoyan la teoría de que los profagos colombianos tienen el mismo origen que el resto de los profagos de *H. pylori*.

Al igual que en reportes previos (Lehours et al., 2011; Vale et al., 2015), en este estudio, las cepas de *H. pylori* portadoras de profagos no mostraron asociación con patrones específicos

de enfermedad. A pesar de esto, varios autores han discutido que los profagos pueden desempeñar un papel fundamental en las poblaciones bacterianas. Muchos profagos contienen genes que contribuyen a la aptitud bacteriana y la capacidad de colonización en un nicho particular (Brüssow et al., 2004). Los genes presentes en los profagos pueden aumentar la aptitud bacteriana a través de una amplia gama de mecanismos y juegan un papel esencial en las enfermedades bacterianas (Taylor et al., 2019). En el caso de *H. pylori*, se han encontrado genes ortólogos de fagos en los genomas de cepas que también portan genes de virulencia *cagA* y / o *vacA* (Kyrillos et al., 2016). Los estudios han sugerido que los profagos de *H. pylori* podrían representar un elemento esencial para la adaptación de *H. pylori* a un entorno hostil, como el estómago humano (Vale & Lehours, 2018; Vale et al., 2017). Sin embargo, se requieren estudios adicionales para dilucidar las funciones específicas de los profagos de *H. pylori* y el papel de estos profagos en el establecimiento de la enfermedad relacionada con *H. pylori* (Vale et al., 2017).

Un informe anterior sobre profagos de *H. pylori* reveló que las IS están presentes en alrededor del 39% de los profagos (Vale et al., 2017). En los aislamientos colombianos, se identificó solamente una IS en solo un profago (Col 23-PUJ). IS606, la IS identificada aquí, también se ha descrito comúnmente en el genoma bacteriano de *H. pylori* y se encontró en un profago incompleto descrito previamente (Sw-C388-G) (Vale et al., 2017). Algunos autores han planteado la hipótesis de que los profagos pueden mediar en la transferencia de IS, contribuyendo aún más a la plasticidad del genoma de *H. pylori*. Además, se ha postulado que los IS presentes en los profagos desempeñan un papel en la inactivación e inmovilización de los fagos entrantes (Ooka et al., 2009; Vale et al., 2017). Más estudios al respecto son necesarios para definir las funciones relacionadas con la presencia de IS en los profagos de *H. pylori*.

Estudios anteriores relacionados con fagos de *H. pylori*, han demostrado una relación entre el sitio de inserción del profago y la clasificación filogenética por PST (Vale et al., 2017). En este estudio, se encontró que los sitios de inserción de los profagos colombianos estaban relacionados principalmente con los previamente descritos para las poblaciones hpNEurope y hpAfrica1, lo cual es consistente con las poblaciones descritas aquí. Sin embargo, en tres casos de los fagos reportados aquí, se encontraron nuevos sitios de inserción en el extremo 5'. Los profagos Col 5-PUJ y Col 6-PUJ (de la población hpAfrica1) se insertaron junto a un

gen que codifica la enzima de modificación del ADN de tipo II. Curiosamente, este producto enzimático corresponde a una endonucleasa de restricción presente en el sistema de modificación de restricción que protege a las bacterias contra la invasión de ADN extraño (Pingoud et al., 2005). El profago Col 23-PUJ (también de la población hpAfrica1) se insertó junto al gen *RuvA*, que codifica una helicasa dependiente de ATP que forma parte de la maquinaria de recombinación homóloga en bacterias (Chen et al., 2002); finalmente, el profago presente en la cepa PZ5019 (disponible de NCBI) (de la población hpAfrica1) se insertó junto a un gen que codifica un factor desencadenante, del cual no hay mucha información. En estos últimos 3 casos, el sitio de inserción podría corresponder a ubicaciones aleatorias. Se necesitarían estudios de expresión genética para evaluar el impacto del sitio de inserción del profago en las bacterias.

Utilizando la técnica de PST, hasta ahora se han identificado cuatro poblaciones de profagos, y se lograron discriminar dos poblaciones europeas (hpNEurope y hpSWEurope) (Vale et al., 2015); hallazgos que fueron posteriormente validados por estudios que caracterizaron el genoma completo de *H. pylori* y que también encontraron dos subpoblaciones de la población previamente denominada hpEurope (Thorell, Yahara, et al., 2017). El análisis de PST realizado aquí reveló la presencia de cinco grupos filogenéticos, uno más de lo informado anteriormente, esto confirma el nivel de alta resolución del análisis de PST en la discriminación de poblaciones de profagos de *H. pylori*.

En los profagos colombianos, el 25,0% (10) fueron asignados a la población hpSWEurope, el 5,0% (2) a hpNEurope, el 42,5% (17) a hpAfrica1 y el 27,5% (11) a una nueva población. La quinta y nueva población estuvo compuesta únicamente por profagos de origen colombiano por lo que recibió el nombre de hpColombia. Ninguno de los profagos colombianos se clasificó en la población hpEastAsia. La diversidad de las poblaciones de los profagos colombianos refleja la compleja historia de colonización de Colombia. Posibles eventos de recombinación e introgresión entre cepas pueden haber llevado a la generación de un nuevo linaje. La naturaleza híbrida de las poblaciones de profagos colombianos también es evidente en otra cepa de América Latina, Cuz20, que corresponde a un aislamiento de Perú. Estos hallazgos deberían estar relacionados con la historia reciente de las olas de migración humana a Colombia, por parte de europeos y africanos.

Los árboles filogenéticos realizados en este estudio colocan a la población de hpColombia más cerca de la población de hpNEurope. Al igual que para las nuevas poblaciones bacterianas, mencionadas en el capítulo 3, la población de profagos hpColombia puede haber divergido de los profagos hpNEurope que llegaron a América, debido a la estrecha agrupación filogenética (Figuras 14 y 15) entre los profagos colombianos y los pertenecientes a la población hpNEurope. La identificación de una nueva población de profagos de las Américas concuerda con los hallazgos filogenéticos descritos en el capítulo 3 y con otros estudios que señalan a la existencia de subpoblaciones de *H. pylori* en Colombia (Muñoz-Ramírez et al., 2020; Thorell, Yahara, et al., 2017). Los sitios de inserción de profagos de la población de hpColombia también coinciden con los que se encuentran con mayor frecuencia para las poblaciones de hpNEurope, lo que refuerza que estas dos poblaciones de profagos están estrechamente relacionadas. El árbol filogenético realizado con los genomas completos de los profagos (Figura 15) permitió evidenciar que algunos profagos mostraban una segregación filogeográfica discrepante de su clasificación PST, lo que sugiere la existencia de supuestos eventos de recombinación, como se había informado anteriormente (Vale et al., 2017). Este hallazgo además sugiere que estos profagos tienen un mosaico genómico compuesto por ambas poblaciones. Por ejemplo, el profago Cuz20 (genoma disponible en el NCBI) se asignó a la población hpNEurope, pero el resultado de PST indicó cierta introgresión de hpColombia y hpEastAsia (Tabla 20), lo cual es evidente también en el árbol filogenético, en donde el profago Cuz20 se agrupó junto con los profagos hpColombia (Figura 15). Cuz20 corresponde a un profago encontrado en un aislamiento peruano, lo cual respalda la teoría sobre la naturaleza híbrida de las cepas americanas y podría estar relacionado con los hallazgos recientes de nuevas subpoblaciones presentes en aislamientos peruanos (Gutiérrez-Escobar et al., 2020).

Como se discutió en el capítulo 3, se ha demostrado la existencia de un componente poblacional específico en cepas aisladas de la población mestiza colombiana (Gutiérrez-Escobar et al., 2017; Muñoz-Ramírez et al., 2020; Muñoz-Ramírez et al., 2017; Thorell, Yahara, et al., 2017), lo que sugiere que las cepas colombianas de *H. pylori* han seguido vías evolutivas únicas (Muñoz-Ramírez et al., 2017) y han formado nuevas subpoblaciones a partir de una fuente europea (Thorell, Yahara, et al., 2017), específicamente a partir de la población hspSWEurope, dando origen a la subpoblación colombiana denominada

hspSWEuropeColombia (Muñoz-Ramirez et al., 2020). Curiosamente, los hallazgos filogenéticos relacionados con el nuevo grupo de profagos de Colombia muestran que esta nueva población de profagos tiene mayor cercanía a la población hpNEurope y no a la hpSWEurope. En general, los hallazgos respaldan el hecho de que las cepas colombianas de *H. pylori* y sus profagos son definitivamente de origen europeo, pero han evolucionado de una forma independiente para lograr la adaptación al medio y al hospedero.

Nuestros resultados verifican que existe una coevolución entre *H. pylori* y sus profagos, llevando al desarrollo de nuevos linajes en el nuevo mundo.

6.5 CONCLUSIONES

Los resultados presentados en este capítulo permiten concluir que existe una nueva población filogenética de profagos de *H. pylori*, hpColombia, correspondiente a una población compuesta exclusivamente por cepas de origen colombiano. Además, estos resultados permitieron verificar que la técnica de PST puede diferenciar poblaciones individuales de profagos y que los profagos afectan la diversidad genética de *H. pylori*, la cual está fuertemente relacionada con la geografía. La realización de estudios adicionales con un mayor número de cepas colombianas y otras cepas de las Américas podría permitir una caracterización en profundidad de la estructura de la población de profagos a un mayor nivel.

7. CONCLUSIONES GENERALES

- La genómica comparativa es una herramienta útil en la evaluación de varios elementos que contribuyen a la diversidad entre los genomas de *H. pylori*.
- Los análisis basados en técnicas de secuenciación de nueva generación de *H. pylori* permiten la detección de mutaciones asociadas a generación de mecanismos de resistencia a CLA, LEV, TET, AMX y MTZ. Sin embargo, más investigación en este campo es necesaria para cubrir posibles mecanismos raros o que no han sido previamente descritos. Así mismo, sería recomendable usar genomas poco fragmentados.
- El genotipo de virulencia más frecuentemente encontrado fue *vacAs1b il m1 / babA2 / iceA2 / cagPAI* parcialmente deletado con *cagA+* (EPIYA-ABC). Los genotipos de virulencia circulantes en Colombia son diversos y no se logró establecer que alguno estuviera presente con mayor frecuencia en una región geográfica específica o para un grupo de pacientes.
- Se comprobó mediante análisis filogenético que en Colombia existe una nueva línea evolutiva de *H. pylori* la cual procede del grupo poblacional hpEurope; sin embargo, el análisis del pangenoma de *H. pylori* colombiano no mostró la presencia de algún gen que pudiera considerarse como un marcador de esta especial filogeografía. Estos resultados destacan la relación entre la diversidad genética de *H. pylori* y el seguimiento de las migraciones humanas, y verifican la coevolución entre *H. pylori* y sus profagos, lo que probablemente llevo al desarrollo de un nuevo linaje de profagos en el nuevo mundo.
- Se logró concluir que en Colombia existe una nueva población filogenética de profagos de *H. pylori*, denominada hpColombia. Estudios adicionales con un mayor número de cepas colombianas y otras cepas de las Américas podría permitir una caracterización en profundidad de la estructura de la población de profagos a un mayor nivel.

8. RECOMENDACIONES

H. pylori es una bacteria extremadamente diversa. En este estudio, fue evidente que la rápida evolución de la bacteria ha influido en la diferenciación de subpoblaciones de *H. pylori*. Sin embargo, aún es necesario comprender la propagación de ciertos genes entre subpoblaciones para comprender mejor la adaptación de la bacteria a su hospedero humano. Se recomienda plantear nuevos estudios en Colombia para determinar las relaciones filogenéticas que existen entre la bacteria y la población mestiza, indígena y afrodescendiente. Específicamente en relación con la virulencia y la capacidad que tiene la bacteria para desencadenar enfermedades clínicas en ciertas poblaciones.

La resistencia a los antimicrobianos en *H. pylori* es un fenómeno creciente que se ha convertido en una de las mayores amenazas para la salud humana. Estudiar los mecanismos genéticos subyacentes de este fenómeno es importante para la implementación de estrategias de manejo y vigilancia. Sin embargo, algunos mecanismos de resistencia relacionados con cambios fisiológicos de la bacteria (regulación de la expresión de bombas de eflujo o de OMP) o propiedades de adaptación (formación de biopelículas o formas cocoides) no logran ser evidenciados mediante las estrategias tradicionales de cultivo o técnicas moleculares que detectan mutaciones. Se recomienda incluir en futuros estudios nuevas estrategias de análisis en las cuales se combinen los enfoques de microbiología tradicional y los multiómicos para comprender mejor la resistencia a los antimicrobianos en *H. pylori*.

Los hallazgos obtenidos en este estudio sobre los profagos de *H. pylori* reflejan la existencia de relaciones coevolutivas de estos fagos y la bacteria. Para avanzar en el conocimiento de este tema, es importante investigar en detalle las características biológicas y genómicas de los fagos de *H. pylori*. Se recomienda plantear nuevos estudios para comprender el papel de los fagos en el genoma bacteriano, esto teniendo en cuenta que los profagos pueden actuar como reservorios de genes, lo que puede beneficiar a los patógenos en formas que apenas están comenzando a determinarse.

Recomendaciones

Para el diseño de nuevos estudios que pretendan dar cuenta de viruloma o resistoma de *H. pylori* en Colombia, se recomienda incluir una muestra más amplia que incluya aislamientos procedentes de pacientes con diversas patologías, que provengan de varias regiones geográficas del país y que cuenten con la información demográfica y clínica lo más completa posible. Así mismo, sería valioso poder incluir en los estudios cepas o genomas de otros continentes para poder comprobar a un nivel más profundo la teoría de la evolución que ha tenido *H. pylori* en el país y en la región de las Américas.

Los análisis realizados aquí se limitaron a las cepas disponibles públicamente y unas pocas que se secuenciaron recientemente. Se recomienda realizar un análisis a mayor escala que ayude a comprender la distribución geográfica y la asociación de la enfermedad con diferentes elementos genómicos de *H. pylori* en Colombia. Así mismo, se recomienda incluir en los análisis genomas completos o no muy fragmentados (< 50 contigs) para evitar errores en la caracterización genómica que puedan ser dados por los gaps existentes en ellos genomas.

BIBLIOGRAFÍA

- Achtman, M., Azuma, T., Berg, D. E., Ito, Y., Morelli, G., Pan, Z. J., . . . van Doorn, L. J. (1999). Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol Microbiol*, 32(3), 459-470. <https://doi.org/10.1046/j.1365-2958.1999.01382.x>
- Acosta, C. P., Hurtado, F. A., & Trespalacios, A. A. (2014). Determinación de mutaciones de un solo nucleótido en el gen 23S rRNA de *Helicobacter pylori* relacionadas con resistencia a claritromicina en una población del departamento del Cauca, Colombia. *Biomédica*, 34, 156-162.
- Acosta, C. P., Quiroga, A. J., Sierra-Torres, C. H., & Trespalacios, A. A. (2017). Frecuencia de mutaciones de la nitrorreductasa RdxA de *Helicobacter pylori* para la activación del metronidazol en una población del departamento del Cauca, Colombia. *Biomédica*, 37, 191-199.
- Acosta, N., Quiroga, A., Delgado, P., Bravo, M. M., & Jaramillo, C. (2010). *Helicobacter pylori* CagA protein polymorphisms and their lack of association with pathogenesis. *World J Gastroenterol*, 16(31), 3936-3943. <https://doi.org/10.3748/wjg.v16.i31.3936>
- Ahmadzadeh, A., Ghalehnoei, H., Farzi, N., Yadegar, A., Alebouyeh, M., Aghdaei, H. A., . . . Pour Hossein Gholi, M. A. (2015). Association of CagPAI integrity with severeness of *Helicobacter pylori* infection in patients with gastritis. *Pathol Biol (Paris)*, 63(6), 252-257. <https://doi.org/10.1016/j.patbio.2015.09.004>
- Ailloud, F., Didelot, X., Woltemate, S., Pfaffinger, G., Overmann, J., Bader, R. C., . . . Suerbaum, S. (2019). Within-host evolution of *Helicobacter pylori* shaped by niche-specific adaptation, intragastric migrations and selective sweeps. *Nat Commun*, 10(1), 2273. <https://doi.org/10.1038/s41467-019-10050-1>
- Alm, R. A., Ling, L. S., Moir, D. T., King, B. L., Brown, E. D., Doig, P. C., . . . Trust, T. J. (1999). Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature*, 397(6715), 176-180. <https://doi.org/10.1038/16495>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol*, 215(3), 403-410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Amieva, M. R., & El-Omar, E. M. (2008). Host-bacterial interactions in *Helicobacter pylori* infection. *Gastroenterology*, 134(1), 306-323. <https://doi.org/10.1053/j.gastro.2007.11.009>
- Ansari, S., & Yamaoka, Y. (2017). BabA in adaptation for gastric colonization. *World J Gastroenterol*, 23(23), 4158-4169. <https://doi.org/10.3748/wjg.v23.i23.4158>
- Ansari, S., & Yamaoka, Y. (2020). Virulence Factor Cytotoxin-Associated Gene A (CagA)-Mediated Gastric Pathogenicity. *Int J Mol Sci*, 21(19). <https://doi.org/10.3390/ijms21197430>
- Arama, S. S., Tiliscan, C., Negoita, C., Croitoru, A., Arama, V., Mihai, C. M., . . . Garg, A. (2016). Efficacy of 7-Day and 14-Day Triple Therapy Regimens for the Eradication of *Helicobacter pylori*: A Comparative Study in a Cohort of Romanian Patients. *Gastroenterol Res Pract*, 2016, 5061640. <https://doi.org/10.1155/2016/5061640>
- Arevalo, A. (2010). *PREVALENCIA DE LOS GENOTIPOS DE VIRULENCIA DE Helicobacter pylori cagA, vacA, baba2 e iceA EN PACIENTES COLOMBIANOS CON DISPEPSIA FUNCIONAL* Pontificia Universidad Javeriana]. Bogotá, Colombia. <https://repository.javeriana.edu.co/bitstream/handle/10554/818/cien20.pdf?sequence=1>
- Arévalo, A., Otero, W., & Trespalacios, A. A. (2019). *Helicobacter pylori*: resistencia múltiple en pacientes de Bogotá, Colombia. *Biomédica*, 39, 125-134.

- Arévalo, A., Trespalacios, A. A., & Otero, W. (2009). Importancia de la proteína CagA en infección por *Helicobacter pylori*. *Revista Colombiana de Gastroenterología*, *24*, 388-395.
- Arévalo-Galvis, A., Trespalacios-Rangell, A. A., Otero, W., Mercado-Reyes, M. M., & Poutou-Piñales, R. A. (2012). Prevalence of *cagA*, *vacA*, *babA2* and *iceA* genes in *H. pylori* strains isolated from Colombian patients with functional dyspepsia. *Pol J Microbiol*, *61*(1), 33-40.
- Arévalo-Jaimes, B. V., Rojas-Rengifo, D. F., Jaramillo, C. A., de Molano, B. M., Vera-Chamorro, J. F., & Del Pilar Delgado, M. (2019). Genotypic determination of resistance and heteroresistance to clarithromycin in *Helicobacter pylori* isolates from antrum and corpus of Colombian symptomatic patients. *BMC Infect Dis*, *19*(1), 546. <https://doi.org/10.1186/s12879-019-4178-x>
- Atehortua Rendon, J. D., Pérez Cala, T. L., & Martínez, A. Descripción de la resistencia de *Helicobacter pylori* a seis antibióticos de uso frecuente en Colombia.
- Atherton, J. C. (1998). *H. pylori* virulence factors. *Br Med Bull*, *54*(1), 105-120.
- Atherton, J. C., Cao, P., Peek, R. M., Tummuru, M. K., Blaser, M. J., & Cover, T. L. (1995). Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. Association of specific *vacA* types with cytotoxin production and peptic ulceration. *J Biol Chem*, *270*(30), 17771-17777.
- Atrisco-Morales, J., Martínez-Santos, V. I., Román-Román, A., Alarcón-Millán, J., De Sampedro-Reyes, J., Cruz-Del Carmen, I., . . . Fernández-Tilapa, G. (2018). *vacA* s1m1 genotype and *cagA* EPIYA-ABC pattern are predominant among *Helicobacter pylori* strains isolated from Mexican patients with chronic gastritis. *J Med Microbiol*, *67*(3), 314-324. <https://doi.org/10.1099/jmm.0.000660>
- Backert, S., & Selbach, M. (2008). Role of type IV secretion in *Helicobacter pylori* pathogenesis. *Cell Microbiol*, *10*(8), 1573-1581. <https://doi.org/10.1111/j.1462-5822.2008.01156.x>
- Baj, J., Forma, A., Sitarz, M., Portincasa, P., Garruti, G., Krasowska, D., & Maciejewski, R. (2020). Virulence Factors-Mechanisms of Bacterial Pathogenicity in the Gastric Microenvironment. *Cells*, *10*(1). <https://doi.org/10.3390/cells10010027>
- Banga Ndzouboukou, J. L., Lei, Q., Ullah, N., Zhang, Y., Hao, L., & Fan, X. (2021). *Helicobacter pylori* adhesins: HpaA a potential antigen in experimental vaccines for *H. pylori*. *Helicobacter*, *26*(1), e12758. <https://doi.org/10.1111/hel.12758>
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., . . . Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*, *19*(5), 455-477. <https://doi.org/10.1089/cmb.2012.0021>
- Basso, D., Zambon, C. F., Letley, D. P., Stranges, A., Marchet, A., Rhead, J. L., . . . Atherton, J. C. (2008). Clinical relevance of *Helicobacter pylori* *cagA* and *vacA* gene polymorphisms. *Gastroenterology*, *135*(1), 91-99. <https://doi.org/10.1053/j.gastro.2008.03.041>
- Beltrán-Anaya, F. O., Poblete, T. M., Román-Román, A., Reyes, S., de Sampedro, J., Peralta-Zaragoza, O., . . . Fernández-Tilapa, G. (2014). The EPIYA-ABCC motif pattern in *CagA* of *Helicobacter pylori* is associated with peptic ulcer and gastric cancer in Mexican population. *BMC Gastroenterol*, *14*, 223. <https://doi.org/10.1186/s12876-014-0223-9>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Boyanova, L., Hadzhiyski, P., Kandilarov, N., Markovska, R., & Mitov, I. (2019). Multidrug resistance in. *Expert Rev Clin Pharmacol*, *12*(9), 909-915. <https://doi.org/10.1080/17512433.2019.1654858>

- Boyanova, L., Yordanov, D., Gergova, G., Markovska, R., & Mitov, I. (2010). Association of iceA and babA genotypes in Helicobacter pylori strains with patient and strain characteristics. *Antonie Van Leeuwenhoek*, 98(3), 343-350. <https://doi.org/10.1007/s10482-010-9448-y>
- Bravo, M. M., Martínez, T., & Bravo, J. C. Relación entre la integridad del CagPAI y los polimorfismos en el gen CagA con la severidad de la gastritis en pacientes infectados con H. pylori CagA positivo.
- Bridge, D. R., & Merrell, D. S. (2013). Polymorphism in the Helicobacter pylori CagA and VacA toxins and disease. *Gut Microbes*, 4(2), 101-117. <https://doi.org/10.4161/gmic.23797>
- Brüssow, H., Canchaya, C., & Hardt, W. D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev*, 68(3), 560-602, table of contents. <https://doi.org/10.1128/MMBR.68.3.560-602.2004>
- Burucoa, C., & Axon, A. (2017). Epidemiology of Helicobacter pylori infection. *Helicobacter*, 22 Suppl 1. <https://doi.org/10.1111/hel.12403>
- Bustamante-Rengifo, J. A., Matta, A. J., Pazos, A., & Bravo, L. E. (2013). In vitro effect of amoxicillin and clarithromycin on the 3' region of cagA gene in Helicobacter pylori isolates. *World J Gastroenterol*, 19(36), 6044-6054. <https://doi.org/10.3748/wjg.v19.i36.6044>
- Camargo, M. C., Beltran, M., Conde-Glez, C. J., Harris, P. R., Michel, A., Waterboer, T., . . . Rabkin, C. S. (2015). Serological response to Helicobacter pylori infection among Latin American populations with contrasting risks of gastric cancer. *Int J Cancer*, 137(12), 3000-3005. <https://doi.org/10.1002/ijc.29678>
- Camargo, M. C., García, A., Riquelme, A., Otero, W., Camargo, C. A., Hernandez-García, T., . . . Rabkin, C. S. (2014). The problem of Helicobacter pylori resistance to antibiotics: a systematic review in Latin America. *Am J Gastroenterol*, 109(4), 485-495. <https://doi.org/10.1038/ajg.2014.24>
- Camilo, V., Sugiyama, T., & Touati, E. (2017). Pathogenesis of Helicobacter pylori infection. *Helicobacter*, 22 Suppl 1. <https://doi.org/10.1111/hel.12405>
- Camorlinga-Ponce, M., Perez-Perez, G., Gonzalez-Valencia, G., Mendoza, I., Peñaloza-Espinosa, R., Ramos, I., . . . Torres, J. (2011). Helicobacter pylori genotyping from American indigenous groups shows novel Amerindian vacA and cagA alleles and Asian, African and European admixture. *PLoS One*, 6(11), e27212. <https://doi.org/10.1371/journal.pone.0027212>
- Canzian, F., Rizzato, C., Obazee, O., Stein, A., Flores-Luna, L., Camorlinga-Ponce, M., . . . Kato, I. (2020). Genetic polymorphisms in the cag pathogenicity island of Helicobacter pylori and risk of stomach cancer and high-grade premalignant gastric lesions. *Int J Cancer*, 147(9), 2437-2445. <https://doi.org/10.1002/ijc.33032>
- Carlosama-Rosero, Y. H., Acosta-Astaiza, C. P., Sierra-Torres, C. H., & Bolaños-Bravo, H. J. (2021). Helicobacter pylori genotypes associated with gastric cancer and dysplasia in Colombian patients. *Rev Gastroenterol Mex*. <https://doi.org/10.1016/j.rgm.2021.01.005>
- Cavalcante, M. Q., Silva, C. I., Braga-Neto, M. B., Fialho, A. B., Nunes Fialho, A., Barbosa, A. M., . . . Braga, L. L. (2012). Helicobacter pylori vacA and cagA genotypes in patients from northeastern Brazil with upper gastrointestinal diseases. *Mem Inst Oswaldo Cruz*, 107(4), 561-563. <https://doi.org/10.1590/s0074-02762012000400021>
- Censini, S., Lange, C., Xiang, Z., Crabtree, J. E., Ghiara, P., Borodovsky, M., . . . Covacci, A. (1996). cag, a pathogenicity island of Helicobacter pylori, encodes type I-specific and disease-associated virulence factors. *Proc Natl Acad Sci U S A*, 93(25), 14648-14653. <https://doi.org/10.1073/pnas.93.25.14648>

- Chang, W. L., Yeh, Y. C., & Sheu, B. S. (2018). The impacts of *H. pylori* virulence factors on the development of gastroduodenal diseases. *J Biomed Sci*, 25(1), 68. <https://doi.org/10.1186/s12929-018-0466-9>
- Chauhan, N., Tay, A. C. Y., Marshall, B. J., & Jain, U. (2019). Helicobacter pylori VacA, a distinct toxin exerts diverse functionalities in numerous cells: An overview. *Helicobacter*, 24(1), e12544. <https://doi.org/10.1111/hel.12544>
- Chen, Y. J., Yu, X., & Egelman, E. H. (2002). The hexameric ring structure of the Escherichia coli RuvB branch migration protein. *J Mol Biol*, 319(3), 587-591. [https://doi.org/10.1016/S0022-2836\(02\)00353-4](https://doi.org/10.1016/S0022-2836(02)00353-4)
- Chey, W. D., Leontiadis, G. I., Howden, C. W., & Moss, S. F. (2017). ACG Clinical Guideline: Treatment of Helicobacter pylori Infection. *Am J Gastroenterol*, 112(2), 212-239. <https://doi.org/10.1038/ajg.2016.563>
- Chomvarin, C., Namwat, W., Chaicumpar, K., Mairiang, P., Sangchan, A., Sripa, B., . . . Vilaichone, R. K. (2008). Prevalence of Helicobacter pylori vacA, cagA, cagE, iceA and babA2 genotypes in Thai dyspeptic patients. *Int J Infect Dis*, 12(1), 30-36. <https://doi.org/10.1016/j.ijid.2007.03.012>
- Cittelly, D. M., Huertas, M. G., Martínez, J. D., Oliveros, R., Posso, H., Bravo, M. M., & Orozco, O. (2002). [Helicobacter pylori genotypes in non atrophic gastritis are different of the found in peptic ulcer, premalignant lesions and gastric cancer in Colombia]. *Rev Med Chil*, 130(2), 143-151.
- Conradi, J., Tegtmeyer, N., Woźna, M., Wissbrock, M., Michalek, C., Gagell, C., . . . Backert, S. (2012). An RGD helper sequence in CagL of Helicobacter pylori assists in interactions with integrins and injection of CagA. *Front Cell Infect Microbiol*, 2, 70. <https://doi.org/10.3389/fcimb.2012.00070>
- Correa, P., & Piazuelo, M. B. (2010). Cáncer gástrico: el enigma colombiano. *Revista Colombiana de Gastroenterología*, 25, 334-337.
- Cover, T. L. (2016). Helicobacter pylori Diversity and Gastric Cancer Risk. *MBio*, 7(1), e01869-01815. <https://doi.org/10.1128/mBio.01869-15>
- Cover, T. L., & Blanke, S. R. (2005). Helicobacter pylori VacA, a paradigm for toxin multifunctionality. *Nat Rev Microbiol*, 3(4), 320-332. <https://doi.org/10.1038/nrmicro1095>
- Cover, T. L., Lacy, D. B., & Ohi, M. D. (2020). The Helicobacter pylori Cag Type IV Secretion System. *Trends Microbiol*, 28(8), 682-695. <https://doi.org/10.1016/j.tim.2020.02.004>
- Dang, B. N., & Graham, D. Y. (2017). Helicobacter pylori infection and antibiotic resistance: a WHO high priority? *Nat Rev Gastroenterol Hepatol*. <https://doi.org/10.1038/nrgastro.2017.57>
- de Sablet, T., Piazuelo, M. B., Shaffer, C. L., Schneider, B. G., Asim, M., Chaturvedi, R., . . . Wilson, K. T. (2011). Phylogeographic origin of Helicobacter pylori is a determinant of gastric cancer risk. *Gut*, 60(9), 1189-1195. <https://doi.org/10.1136/gut.2010.234468>
- Denic, M., Touati, E., & De Reuse, H. (2020). Review: Pathogenesis of Helicobacter pylori infection. *Helicobacter*, 25 Suppl 1, e12736. <https://doi.org/10.1111/hel.12736>
- Devi, S. M., Ahmed, I., Francalacci, P., Hussain, M. A., Akhter, Y., Alvi, A., . . . Ahmed, N. (2007). Ancestral European roots of Helicobacter pylori in India. *BMC Genomics*, 8, 184. <https://doi.org/10.1186/1471-2164-8-184>
- Earl, D. A., & vonHoldt, B. M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method.
- Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method.

- Erzin, Y., Koksall, V., Altun, S., Dobrucali, A., Aslan, M., Erdamar, S., . . . Kocazeybek, B. (2006). Prevalence of *Helicobacter pylori* vacA, cagA, cagE, iceA, babA2 genotypes and correlation with clinical outcome in Turkish patients with dyspepsia. *Helicobacter*, *11*(6), 574-580. <https://doi.org/10.1111/j.1523-5378.2006.00461.x>
- Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, *164*(4), 1567-1587.
- Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes*, *7*(4), 574-578. <https://doi.org/10.1111/j.1471-8286.2007.01758.x>
- Falush, D., Wirth, T., Linz, B., Pritchard, J. K., Stephens, M., Kidd, M., . . . Suerbaum, S. (2003). Traces of human migrations in *Helicobacter pylori* populations. *Science*, *299*(5612), 1582-1585. <https://doi.org/10.1126/science.1080857>
- Feliciano, O., Gutierrez, O., Valdés, L., Fragoso, T., Calderin, A. M., Valdes, A. E., & Llanes, R. (2015). Prevalence of *Helicobacter pylori* vacA, cagA, and iceA Genotypes in Cuban Patients with Upper Gastrointestinal Diseases. *Biomed Res Int*, *2015*, 753710. <https://doi.org/10.1155/2015/753710>
- Ferreira, R. M., Machado, J. C., Letley, D., Atherton, J. C., Pardo, M. L., Gonzalez, C. A., . . . Figueiredo, C. (2012). A novel method for genotyping the *Helicobacter pylori* vacA intermediate region directly in gastric biopsy specimens. *J Clin Microbiol*, *50*(12), 3983-3989. <https://doi.org/10.1128/JCM.02087-12>
- Figueiredo, C., Quint, W. G., Sanna, R., Sablon, E., Donahue, J. P., Xu, Q., . . . van Doorn, L. J. (2000). Genetic organization and heterogeneity of the iceA locus of *Helicobacter pylori*. *Gene*, *246*(1-2), 59-68. [https://doi.org/10.1016/s0378-1119\(00\)00054-8](https://doi.org/10.1016/s0378-1119(00)00054-8)
- Figueroa, M., Cortés, A., Pazos, Á., & Bravo, L. E. (2012). Sensibilidad in vitro a amoxicilina y claritromicina de *Helicobacter pylori* obtenido de biopsias gástricas de pacientes en zona de bajo riesgo para cáncer gástrico. *Biomédica*, *32*, 32-42.
- Fischer, W., Tegtmeyer, N., Stingl, K., & Backert, S. (2020). Four Chromosomal Type IV Secretion Systems in. *Front Microbiol*, *11*, 1592. <https://doi.org/10.3389/fmicb.2020.01592>
- Fischer, W., Windhager, L., Rohrer, S., Zeiller, M., Karnholz, A., Hoffmann, R., . . . Haas, R. (2010). Strain-specific genes of *Helicobacter pylori*: genome evolution driven by a novel type IV secretion system and genomic island transfer. *Nucleic Acids Res*, *38*(18), 6089-6101. <https://doi.org/10.1093/nar/gkq378>
- FitzGerald, R., & Smith, S. M. (2021). An Overview of *Helicobacter pylori* Infection. *Methods Mol Biol*, *2283*, 1-14. https://doi.org/10.1007/978-1-0716-1302-3_1
- Fujimoto, S., Olaniyi Ojo, O., Arnqvist, A., Wu, J. Y., Odenbreit, S., Haas, R., . . . Yamaoka, Y. (2007). *Helicobacter pylori* BabA expression, gastric mucosal injury, and clinical outcome. *Clin Gastroenterol Hepatol*, *5*(1), 49-58. <https://doi.org/10.1016/j.cgh.2006.09.015>
- Furuta, Y., Konno, M., Osaki, T., Yonezawa, H., Ishige, T., Imai, M., . . . Kobayashi, I. (2015). Microevolution of Virulence-Related Genes in *Helicobacter pylori* Familial Infection. *PLoS One*, *10*(5), e0127197. <https://doi.org/10.1371/journal.pone.0127197>
- Gatti, L. L., Módena, J. L., Payão, S. L., Smith, M. e. A., Fukuhara, Y., de Oliveira, R. B., & Brocchi, M. (2006). Prevalence of *Helicobacter pylori* cagA, iceA and babA2 alleles in Brazilian patients with upper gastrointestinal diseases. *Acta Trop*, *100*(3), 232-240. <https://doi.org/10.1016/j.actatropica.2006.08.014>

- Gerrits, M. M., van Vliet, A. H., Kuipers, E. J., & Kusters, J. G. (2006). Helicobacter pylori and antimicrobial resistance: molecular mechanisms and clinical implications. *Lancet Infect Dis*, 6(11), 699-709. [https://doi.org/10.1016/S1473-3099\(06\)70627-2](https://doi.org/10.1016/S1473-3099(06)70627-2)
- Ghotaslou, R., Leylabadlo, H. E., & Asl, Y. M. (2015). Prevalence of antibiotic resistance in Helicobacter pylori: A recent literature review. *World J Methodol*, 5(3), 164-174. <https://doi.org/10.5662/wjm.v5.i3.164>
- Gong, Y., & Yuan, Y. (2018). Resistance mechanisms of Helicobacter pylori and its dual target precise therapy. *Crit Rev Microbiol*, 44(3), 371-392. <https://doi.org/10.1080/1040841X.2017.1418285>
- González, C. A., Figueiredo, C., Lic, C. B., Ferreira, R. M., Pardo, M. L., Ruiz Liso, J. M., . . . Sanz-Anquela, J. M. (2011). Helicobacter pylori cagA and vacA genotypes as predictors of progression of gastric preneoplastic lesions: a long-term follow-up in a high-risk area in Spain. *Am J Gastroenterol*, 106(5), 867-874. <https://doi.org/10.1038/ajg.2011.1>
- Goodwin, C. S., & Worsley, B. W. (1993). Microbiology of Helicobacter pylori. *Gastroenterol Clin North Am*, 22(1), 5-19.
- Graham, D. Y., Lee, Y. C., & Wu, M. S. (2014). Rational Helicobacter pylori therapy: evidence-based medicine rather than medicine-based evidence. *Clin Gastroenterol Hepatol*, 12(2), 177-186.e173; Discussion e112-173. <https://doi.org/10.1016/j.cgh.2013.05.028>
- Group, I. H. p. W. (2014). *Helicobacter pylori* Eradication as a Strategy for Preventing Gastric Cancer. . In. Lyon France: International Agency for Research on Cancer (IARC Working Group Reports, No. 8). .
- Gutiérrez-Escobar, A. J., Trujillo, E., Acevedo, O., & Bravo, M. M. (2017). Phylogenomics of Colombian. *Gut Pathog*, 9, 52. <https://doi.org/10.1186/s13099-017-0201-1>
- Gutiérrez-Escobar, A. J., Velapatiño, B., Borda, V., Rabkin, C. S., Tarazona-Santos, E., Cabrera, L., . . . Camargo, M. C. (2020). Identification of New. *Front Microbiol*, 11, 601839. <https://doi.org/10.3389/fmicb.2020.601839>
- Gómez, M., Otero, W., & Gutiérrez, Ó. (2007). Tratamiento de la infección por Helicobacter pylori. Encuesta en un grupo de médicos generales y especialistas en Colombia. *Revista Colombiana de Gastroenterología*, 22, 7-16.
- Gómez, M., Ruíz, O., Páramo Hernández, D., Albis, R., & Sabbagh, L. C. (2015). Erradicación del Helicobacter pylori: encuesta realizada por la Asociación Colombiana de Gastroenterología. *Revista Colombiana de Gastroenterología*, 30, 25-31.
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series*,
- Hanafiah, A., & Lopes, B. S. (2020). Genetic diversity and virulence characteristics of Helicobacter pylori isolates in different human ethnic groups. *Infect Genet Evol*, 78, 104135. <https://doi.org/10.1016/j.meegid.2019.104135>
- Hatakeyama, M., & Higashi, H. (2005). Helicobacter pylori CagA: a new paradigm for bacterial carcinogenesis. *Cancer Sci*, 96(12), 835-843. <https://doi.org/10.1111/j.1349-7006.2005.00130.x>
- He, S., Corneloup, A., Guynet, C., Lavatine, L., Caumont-Sarcos, A., Siguier, P., . . . Ton Hoang, B. (2015). The IS200/IS605 Family and "Peel and Paste" Single-strand Transposition Mechanism. *Microbiol Spectr*, 3(4). <https://doi.org/10.1128/microbiolspec.MDNA3-0039-2014>
- Heintschel von Heinegg, E., Nalik, H. P., & Schmid, E. N. (1993). Characterisation of a Helicobacter pylori phage (HP1). *J Med Microbiol*, 38(4), 245-249. <https://doi.org/10.1099/00222615-38-4-245>

- Hennig, E. E., Allen, J. M., & Cover, T. L. (2006). Multiple chromosomal loci for the babA gene in *Helicobacter pylori*. *Infect Immun*, 74(5), 3046-3051. <https://doi.org/10.1128/IAI.74.5.3046-3051.2006>
- Higashi, H., Tsutsumi, R., Fujita, A., Yamazaki, S., Asaka, M., Azuma, T., & Hatakeyama, M. (2002). 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*, 99(22), 14428-14433. <https://doi.org/10.1073/pnas.222375399>
- Hooi, J. K. Y., Lai, W. Y., Ng, W. K., Suen, M. M. Y., Underwood, F. E., Tanyingoh, D., . . . Ng, S. C. (2017). Global Prevalence of *Helicobacter pylori* Infection: Systematic Review and Meta-Analysis. *Gastroenterology*, 153(2), 420-429. <https://doi.org/10.1053/j.gastro.2017.04.022>
- Hu, Y., Zhu, Y., & Lu, N. H. (2017). Novel and Effective Therapeutic Regimens for. *Front Cell Infect Microbiol*, 7, 168. <https://doi.org/10.3389/fcimb.2017.00168>
- Humbert, O., Dorer, M. S., & Salama, N. R. (2011). Characterization of *Helicobacter pylori* factors that control transformation frequency and integration length during inter-strain DNA recombination. *Mol Microbiol*, 79(2), 387-401. <https://doi.org/10.1111/j.1365-2958.2010.07456.x>
- Hunt, R. H., Xiao, S. D., Megraud, F., Leon-Barua, R., Bazzoli, F., van der Merwe, S., . . . Organization, W. G. (2011). *Helicobacter pylori* in developing countries. World Gastroenterology Organisation Global Guideline. *J Gastrointestin Liver Dis*, 20(3), 299-304.
- IARC. (1994). Schistosomes, liver flukes and *Helicobacter pylori*. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7-14 June 1994. *IARC Monogr Eval Carcinog Risks Hum*, 61, 1-241.
- IARC. (2020). *Global Cancer Observatory: cancer today*. 170-colombia-fact-sheets. International Agency for Research on Cancer. Retrieved July 30, 2021 from <https://gco.iarc.fr/today/data/factsheets/populations/170-colombia-fact-sheets.pdf>
- Ilver, D., Arnqvist, A., Ogren, J., Frick, I. M., Kersulyte, D., Incecik, E. T., . . . Borén, T. (1998). *Helicobacter pylori* adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science*, 279(5349), 373-377. <https://doi.org/10.1126/science.279.5349.373>
- Ishaq, S., & Nunn, L. (2015). *Helicobacter pylori* and gastric cancer: a state of the art review. *Gastroenterol Hepatol Bed Bench*, 8(Suppl 1), S6-S14.
- Iwamoto, A., Tanahashi, T., Okada, R., Yoshida, Y., Kikuchi, K., Keida, Y., . . . Azuma, T. (2014). Whole-genome sequencing of clarithromycin resistant *Helicobacter pylori* characterizes unidentified variants of multidrug resistant efflux pump genes. *Gut Pathog*, 6, 27. <https://doi.org/10.1186/1757-4749-6-27>
- Jolley, K. A., Bray, J. E., & Maiden, M. C. J. (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.14826.1>
- Kaas, R. S., Leekitcharoenphon, P., Aarestrup, F. M., & Lund, O. (2014). Solving the problem of comparing whole bacterial genomes across different sequencing platforms. *PLoS One*, 9(8), e104984. <https://doi.org/10.1371/journal.pone.0104984>
- Kao, C. Y., Sheu, B. S., & Wu, J. J. (2016). *Helicobacter pylori* infection: An overview of bacterial virulence factors and pathogenesis. *Biomed J*, 39(1), 14-23. <https://doi.org/10.1016/j.bj.2015.06.002>
- Karlsson, A., Ryberg, A., Dehnoei, M. N., Borch, K., & Monstein, H. J. (2012). Association between cagA and vacA genotypes and pathogenesis in a *Helicobacter pylori* infected population from South-eastern Sweden. *BMC Microbiol*, 12, 129. <https://doi.org/10.1186/1471-2180-12-129>

- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*, 30(4), 772-780. <https://doi.org/10.1093/molbev/mst010>
- Kausser, F., Khan, A. A., Hussain, M. A., Carroll, I. M., Ahmad, N., Tiwari, S., . . . Ahmed, N. (2004). The cag pathogenicity island of *Helicobacter pylori* is disrupted in the majority of patient isolates from different human populations. *J Clin Microbiol*, 42(11), 5302-5308. <https://doi.org/10.1128/JCM.42.11.5302-5308.2004>
- Kawai, M., Furuta, Y., Yahara, K., Tsuru, T., Oshima, K., Handa, N., . . . Kobayashi, I. (2011). Evolution in an oncogenic bacterial species with extreme genome plasticity: *Helicobacter pylori* East Asian genomes. *BMC Microbiol*, 11, 104. <https://doi.org/10.1186/1471-2180-11-104>
- Keikha, M. (2020). Is there a relationship between. *New Microbes New Infect*, 36, 100726. <https://doi.org/10.1016/j.nmni.2020.100726>
- Keikha, M., Ali-Hassanzadeh, M., & Karbalaei, M. (2020). Association of *Helicobacter pylori vacA* genotypes and peptic ulcer in Iranian population: a systematic review and meta-analysis. *BMC Gastroenterol*, 20(1), 266. <https://doi.org/10.1186/s12876-020-01406-9>
- Kennemann, L., Didelot, X., Aebischer, T., Kuhn, S., Drescher, B., Droege, M., . . . Suerbaum, S. (2011). *Helicobacter pylori* genome evolution during human infection. *Proc Natl Acad Sci U S A*, 108(12), 5033-5038. <https://doi.org/10.1073/pnas.1018444108>
- Kersulyte, D., Akopyants, N. S., Clifton, S. W., Roe, B. A., & Berg, D. E. (1998). Novel sequence organization and insertion specificity of IS605 and IS606: chimaeric transposable elements of *Helicobacter pylori*. *Gene*, 223(1-2), 175-186. [https://doi.org/10.1016/s0378-1119\(98\)00164-4](https://doi.org/10.1016/s0378-1119(98)00164-4)
- Kersulyte, D., Kalia, A., Gilman, R. H., Mendez, M., Herrera, P., Cabrera, L., . . . Berg, D. E. (2010). *Helicobacter pylori* from Peruvian amerindians: traces of human migrations in strains from remote Amazon, and genome sequence of an Amerind strain. *PLoS One*, 5(11), e15076. <https://doi.org/10.1371/journal.pone.0015076>
- Kersulyte, D., Mukhopadhyay, A. K., Velapatiño, B., Su, W., Pan, Z., Garcia, C., . . . Berg, D. E. (2000). Differences in genotypes of *Helicobacter pylori* from different human populations. *J Bacteriol*, 182(11), 3210-3218. <https://doi.org/10.1128/jb.182.11.3210-3218.2000>
- Khaledi, M., Bagheri, N., Validi, M., Zamanzad, B., Afkhami, H., Fathi, J., . . . Gholipour, A. (2020). Determination of. *Heliyon*, 6(9), e04971. <https://doi.org/10.1016/j.heliyon.2020.e04971>
- Kim, I. J., & Blanke, S. R. (2012). Remodeling the host environment: modulation of the gastric epithelium by the *Helicobacter pylori* vacuolating toxin (VacA). *Front Cell Infect Microbiol*, 2, 37. <https://doi.org/10.3389/fcimb.2012.00037>
- Kim, K. S., Kang, J. O., Eun, C. S., Han, D. S., & Choi, T. Y. (2002). Mutations in the 23S rRNA gene of *Helicobacter pylori* associated with clarithromycin resistance. *J Korean Med Sci*, 17(5), 599-603. <https://doi.org/10.3346/jkms.2002.17.5.599>
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*, 16(2), 111-120.
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*, 33(7), 1870-1874. <https://doi.org/10.1093/molbev/msw054>
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biol*, 5(2), R12. <https://doi.org/10.1186/gb-2004-5-2-r12>

- Kusters, J. G., van Vliet, A. H., & Kuipers, E. J. (2006). Pathogenesis of *Helicobacter pylori* infection. *Clin Microbiol Rev*, *19*(3), 449-490. <https://doi.org/10.1128/CMR.00054-05>
- Kyrillos, A., Arora, G., Murray, B., & Rosenwald, A. G. (2016). The Presence of Phage Orthologous Genes in *Helicobacter pylori* Correlates with the Presence of the Virulence Factors CagA and VacA. *Helicobacter*, *21*(3), 226-233. <https://doi.org/10.1111/hel.12282>
- Ladeira, M. S., Bueno, R. C., Dos Santos, B. F., Pinto, C. L., Prado, R. P., Silveira, M. G., . . . Salvadori, D. M. (2008). Relationship among oxidative DNA damage, gastric mucosal density and the relevance of cagA, vacA and iceA genotypes of *Helicobacter pylori*. *Dig Dis Sci*, *53*(1), 248-255. <https://doi.org/10.1007/s10620-007-9850-0>
- Ladeira, M. S., Rodrigues, M. A., Salvadori, D. M., Neto, P. P., Achilles, P., Lerco, M. M., . . . Freire-Maia, D. V. (2004). Relationships between cagA, vacA, and iceA genotypes of *Helicobacter pylori* and DNA damage in the gastric mucosa. *Environ Mol Mutagen*, *44*(2), 91-98. <https://doi.org/10.1002/em.20045>
- Lee, J. W., Kim, N., Nam, R. H., Park, J. H., Kim, J. M., Jung, H. C., & Song, I. S. (2011). Mutations of *Helicobacter pylori* associated with fluoroquinolone resistance in Korea. *Helicobacter*, *16*(4), 301-310. <https://doi.org/10.1111/j.1523-5378.2011.00840.x>
- Lehours, P., Vale, F. F., Bjursell, M. K., Melefors, O., Advani, R., Glavas, S., . . . Andersson, A. F. (2011). Genome sequencing reveals a phage in *Helicobacter pylori*. *mBio*, *2*(6). <https://doi.org/10.1128/mBio.00239-11>
- Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*, *47*(W1), W256-W259. <https://doi.org/10.1093/nar/gkz239>
- Li, Q., Liu, J., Gong, Y., & Yuan, Y. (2017). Association of CagA EPIYA-D or EPIYA-C phosphorylation sites with peptic ulcer and gastric cancer risks: A meta-analysis. *Medicine (Baltimore)*, *96*(17), e6620. <https://doi.org/10.1097/MD.0000000000006620>
- Lim, S. G., Park, R. W., Shin, S. J., Yoon, D., Kang, J. K., Hwang, J. C., . . . Lee, K. M. (2016). The relationship between the failure to eradicate *Helicobacter pylori* and previous antibiotics use. *Dig Liver Dis*, *48*(4), 385-390. <https://doi.org/10.1016/j.dld.2015.12.001>
- Lin, H. J., Perng, C. L., Lo, W. C., Wu, C. W., Tseng, G. Y., Li, A. F., . . . Ou, Y. H. (2004). *Helicobacter pylori* cagA, iceA and vacA genotypes in patients with gastric cancer in Taiwan. *World J Gastroenterol*, *10*(17), 2493-2497. <https://doi.org/10.3748/wjg.v10.i17.2493>
- Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Roumagnac, P., . . . Achtman, M. (2007). An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature*, *445*(7130), 915-918. <https://doi.org/10.1038/nature05562>
- Linz, B., Vololonantenainab, C. R., Seck, A., Carod, J. F., Dia, D., Garin, B., . . . Breurec, S. (2014). Population genetic structure and isolation by distance of *Helicobacter pylori* in Senegal and Madagascar. *PLoS One*, *9*(1), e87355. <https://doi.org/10.1371/journal.pone.0087355>
- Luo, C. H., Chiou, P. Y., Yang, C. Y., & Lin, N. T. (2012). Genome, integration, and transduction of a novel temperate phage of *Helicobacter pylori*. *J Virol*, *86*(16), 8781-8792. <https://doi.org/10.1128/JVI.00446-12>
- Machlowska, J., Baj, J., Sitarz, M., Maciejewski, R., & Sitarz, R. (2020). Gastric Cancer: Epidemiology, Risk Factors, Classification, Genomic Characteristics and Treatment Strategies. *Int J Mol Sci*, *21*(11). <https://doi.org/10.3390/ijms21114012>

- Malfertheiner, P., Megraud, F., O'Morain, C. A., Atherton, J., Axon, A. T., Bazzoli, F., . . . Group, E. H. S. (2012). Management of *Helicobacter pylori* infection--the Maastricht IV/ Florence Consensus Report. *Gut*, *61*(5), 646-664. <https://doi.org/10.1136/gutjnl-2012-302084>
- Malfertheiner, P., Megraud, F., O'Morain, C. A., Gisbert, J. P., Kuipers, E. J., Axon, A. T., . . . panel, E. H. a. M. S. G. a. C. (2017). Management of *Helicobacter pylori* infection-the Maastricht V/Florence Consensus Report. *Gut*, *66*(1), 6-30. <https://doi.org/10.1136/gutjnl-2016-312288>
- Mannion, A., Dzik-Fox, J., Shen, Z., Piazuolo, M. B., Wilson, K. T., Correa, P., . . . Fox, J. G. (2021). *Helicobacter pylori* Antimicrobial Resistance and Gene Variants in High- and Low-Gastric-Cancer-Risk Populations. *J Clin Microbiol*, *59*(5). <https://doi.org/10.1128/JCM.03203-20>
- Markovska, R., Boyanova, L., Yordanov, D., Stankova, P., Gergova, G., & Mitov, I. (2018). Status of *Helicobacter pylori* *cag* pathogenicity island (*cagPAI*) integrity and significance of its individual genes. *Infect Genet Evol*, *59*, 167-171. <https://doi.org/10.1016/j.meegid.2018.02.009>
- Marques, A. T., Vítor, J. M. B., Santos, A., Oleastro, M., & Vale, F. F. (2020). Trends in. *Microb Genom*, *6*(3). <https://doi.org/10.1099/mgen.0.000344>
- Marshall, B. J., Armstrong, J. A., Francis, G. J., Nokes, N. T., & Wee, S. H. (1987). Antibacterial action of bismuth in relation to *Campylobacter pyloridis* colonization and gastritis. *Digestion*, *37 Suppl 2*, 16-30.
- Mascellino, M. T., Porowska, B., De Angelis, M., & Oliva, A. (2017). Antibiotic susceptibility, heteroresistance, and updated treatment strategies in *Helicobacter pylori* infection. *Drug Des Devel Ther*, *11*, 2209-2220. <https://doi.org/10.2147/DDDT.S136240>
- Matos, J. I., de Sousa, H. A., Marcos-Pinto, R., & Dinis-Ribeiro, M. (2013). *Helicobacter pylori* *CagA* and *VacA* genotypes and gastric phenotype: a meta-analysis. *Eur J Gastroenterol Hepatol*, *25*(12), 1431-1441. <https://doi.org/10.1097/MEG.0b013e328364b53e>
- Matta, A. J., Pazos, A. J., Bustamante-Rengifo, J. A., & Bravo, L. E. (2017). Genomic variability of *Helicobacter pylori* isolates of gastric regions from two Colombian populations. *World J Gastroenterol*, *23*(5), 800-809. <https://doi.org/10.3748/wjg.v23.i5.800>
- Matta, A. J., Zambrano, D. C., & Pazos, A. J. (2018). Punctual mutations in. *World J Gastroenterol*, *24*(14), 1531-1539. <https://doi.org/10.3748/wjg.v24.i14.1531>
- McClain, M. S., Beckett, A. C., & Cover, T. L. (2017). *Helicobacter pylori* Vacuolating Toxin and Gastric Cancer. *Toxins (Basel)*, *9*(10). <https://doi.org/10.3390/toxins9100316>
- Melo-Narváez, M. C., Rojas-Rengifo, D. F., Jimenez-Soto, L. F., Delgado Perafán, M. d. P., Mendoza de Molano, B. E., Vera-Chamorro, J. F., & Jaramillo, C. A. Genotipificación de *cagA* y de la región intermedia de *vacA* en cepas de *Helicobacter pylori* aisladas de pacientes adultos colombianos y asociación con enfermedades gástricas.
- Memon, A. A., Hussein, N. R., Miendje Deyi, V. Y., Burette, A., & Atherton, J. C. (2014). Vacuolating cytotoxin genotypes are strong markers of gastric cancer and duodenal ulcer-associated *Helicobacter pylori* strains: a matched case-control study. *J Clin Microbiol*, *52*(8), 2984-2989. <https://doi.org/10.1128/JCM.00551-14>
- Miftahussurur, M., Shrestha, P. K., Subsomwong, P., Sharma, R. P., & Yamaoka, Y. (2016). Emerging *Helicobacter pylori* levofloxacin resistance and novel genetic mutation in Nepal. *BMC Microbiol*, *16*(1), 256. <https://doi.org/10.1186/s12866-016-0873-6>
- Miftahussurur, M., Yamaoka, Y., & Graham, D. Y. (2017). *Helicobacter pylori* as an oncogenic pathogen, revisited. *Expert Rev Mol Med*, *19*, e4. <https://doi.org/10.1017/erm.2017.4>

- Moodley, Y., Linz, B., Yamaoka, Y., Windsor, H. M., Breurec, S., Wu, J. Y., . . . Achtman, M. (2009). The peopling of the Pacific from a bacterial perspective. *Science*, 323(5913), 527-530. <https://doi.org/10.1126/science.1166083>
- Murakami, K., Okimoto, T., Kodama, M., Tanahashi, J., Fujioka, T., Ikeda, F., . . . Kobayashi, I. (2009). Sitafloxacin activity against *Helicobacter pylori* isolates, including those with *gyrA* mutations. *Antimicrob Agents Chemother*, 53(7), 3097-3099. <https://doi.org/10.1128/AAC.01552-08>
- Muñoz, A. B., Stepanian, J., Trespacios, A. A., & Vale, F. F. (2020). Bacteriophages of *Helicobacter pylori*. *Front Microbiol*, 11, 549084. <https://doi.org/10.3389/fmicb.2020.549084>
- Muñoz-Ramirez, Z. Y., Pascoe, B., Mendez-Tenorio, A., Mourkas, E., Sandoval-Motta, S., Perez-Perez, G., . . . Torres, J. (2020). A 500-year tale of co-evolution, adaptation, and virulence: *Helicobacter pylori* in the Americas. *ISME J*. <https://doi.org/10.1038/s41396-020-00758-0>
- Muñoz-Ramírez, Z. Y., Mendez-Tenorio, A., Kato, I., Bravo, M. M., Rizzato, C., Thorell, K., . . . Torres, J. (2017). Whole Genome Sequence and Phylogenetic Analysis Show *Helicobacter pylori* Strains from Latin America Have Followed a Unique Evolution Pathway [Original Research]. *Frontiers in Cellular and Infection Microbiology*, 7(50). <https://doi.org/10.3389/fcimb.2017.00050>
- Mégraud, F. (2004). H pylori antibiotic resistance: prevalence, importance, and advances in testing. *Gut*, 53(9), 1374-1384. <https://doi.org/10.1136/gut.2003.022111>
- Mégraud, F., Bénéjat, L., Ontsira Ngoyi, E. N., & Lehours, P. (2015). Molecular Approaches to Identify *Helicobacter pylori* Antimicrobial Resistance. *Gastroenterol Clin North Am*, 44(3), 577-596. <https://doi.org/10.1016/j.gtc.2015.05.002>
- Mégraud, F., & Lehours, P. (2007). *Helicobacter pylori* detection and antimicrobial susceptibility testing. *Clin Microbiol Rev*, 20(2), 280-322. <https://doi.org/10.1128/CMR.00033-06>
- Nguyen, L. T., Uchida, T., Tsukamoto, Y., Trinh, T. D., Ta, L., Mai, H. B., . . . Moriyama, M. (2010). Clinical relevance of *cagPAI* intactness in *Helicobacter pylori* isolates from Vietnam. *Eur J Clin Microbiol Infect Dis*, 29(6), 651-660. <https://doi.org/10.1007/s10096-010-0909-z>
- Nilsson, C., Sillén, A., Eriksson, L., Strand, M. L., Enroth, H., Normark, S., . . . Engstrand, L. (2003). Correlation between *cag* pathogenicity island composition and *Helicobacter pylori*-associated gastroduodenal disease. *Infect Immun*, 71(11), 6573-6581. <https://doi.org/10.1128/iai.71.11.6573-6581.2003>
- O'Connor, J. P., Taneike, I., & O'Morain, C. (2009). Improving compliance with *helicobacter pylori* eradication therapy: when and how? *Therap Adv Gastroenterol*, 2(5), 273-279. <https://doi.org/10.1177/1756283X09337342>
- O'Rourke, D. H., & Raff, J. A. (2010). The human genetic history of the Americas: the final frontier. *Curr Biol*, 20(4), R202-207. <https://doi.org/10.1016/j.cub.2009.11.051>
- Odenbreit, S., Püls, J., Sedlmaier, B., Gerland, E., Fischer, W., & Haas, R. (2000). Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. *Science*, 287(5457), 1497-1500. <https://doi.org/10.1126/science.287.5457.1497>
- Ogiwara, H., Graham, D. Y., & Yamaoka, Y. (2008). *vacA* i-region subtyping. *Gastroenterology*, 134(4), 1267; author reply 1268. <https://doi.org/10.1053/j.gastro.2007.11.062>
- Ogiwara, H., Sugimoto, M., Ohno, T., Vilaichone, R. K., Mahachai, V., Graham, D. Y., & Yamaoka, Y. (2009). Role of deletion located between the intermediate and middle regions of the *Helicobacter pylori vacA* gene in cases of gastroduodenal diseases. *J Clin Microbiol*, 47(11), 3493-3500. <https://doi.org/10.1128/JCM.00887-09>

- Ooka, T., Ogura, Y., Asadulghani, M., Ohnishi, M., Nakayama, K., Terajima, J., . . . Hayashi, T. (2009). Inference of the impact of insertion sequence (IS) elements on bacterial genome diversification through analysis of small-size structural polymorphisms in *Escherichia coli* O157 genomes. *Genome Res*, *19*(10), 1809-1816. <https://doi.org/10.1101/gr.089615.108>
- Otero R, W., Gómez Z, M., Otero P, L., & Trespacios R, A. (2018). [Helicobacter pylori: How do we treat it in 2018?]. *Rev Gastroenterol Peru*, *38*(1), 54-63.
- Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., . . . Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, *31*(22), 3691-3693. <https://doi.org/10.1093/bioinformatics/btv421>
- Paniagua, G. L., Monroy, E., Rodríguez, R., Arroniz, S., Rodríguez, C., Cortés, J. L., . . . Vaca, S. (2009). Frequency of vacA, cagA and babA2 virulence markers in *Helicobacter pylori* strains isolated from Mexican patients with chronic gastritis. *Ann Clin Microbiol Antimicrob*, *8*, 14. <https://doi.org/10.1186/1476-0711-8-14>
- Park, J. Y., Forman, D., Waskito, L. A., Yamaoka, Y., & Crabtree, J. E. (2018). Epidemiology of *Helicobacter pylori* and CagA-Positive Infections and Global Variations in Gastric Cancer. *Toxins (Basel)*, *10*(4). <https://doi.org/10.3390/toxins10040163>
- Park, S. J., Son, W. S., & Lee, B. J. (2012). Structural analysis of hypothetical proteins from *Helicobacter pylori*: an approach to estimate functions of unknown or hypothetical proteins. *Int J Mol Sci*, *13*(6), 7109-7137. <https://doi.org/10.3390/ijms13067109>
- Parrish, C. R., Holmes, E. C., Morens, D. M., Park, E. C., Burke, D. S., Calisher, C. H., . . . Daszak, P. (2008). Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol Mol Biol Rev*, *72*(3), 457-470. <https://doi.org/10.1128/MMBR.00004-08>
- Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A. J., Thomson, N. R., . . . Brockhurst, M. A. (2010). Antagonistic coevolution accelerates molecular evolution. *Nature*, *464*(7286), 275-278. <https://doi.org/10.1038/nature08798>
- Patra, R., Chattopadhyay, S., De, R., Datta, S., Chowdhury, A., Ramamurthy, T., . . . Mukhopadhyay, A. K. (2011). Intact cag pathogenicity island of *Helicobacter pylori* without disease association in Kolkata, India. *Int J Med Microbiol*, *301*(4), 293-302. <https://doi.org/10.1016/j.ijmm.2010.10.003>
- Peek, R. M., & Blaser, M. J. (2002). *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. *Nat Rev Cancer*, *2*(1), 28-37. <https://doi.org/10.1038/nrc703>
- Pingoud, A., Fuxreiter, M., Pingoud, V., & Wende, W. (2005). Type II restriction endonucleases: structure and mechanism. *Cell Mol Life Sci*, *62*(6), 685-707. <https://doi.org/10.1007/s00018-004-4513-1>
- Pormohammad, A., Ghotaslou, R., Leylabadlo, H. E., Nasiri, M. J., Dabiri, H., & Hashemi, A. (2018). Risk of gastric cancer in association with *Helicobacter pylori* different virulence factors: A systematic review and meta-analysis. *Microb Pathog*, *118*, 214-219. <https://doi.org/10.1016/j.micpath.2018.03.004>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.
- Kumar, S., Nguyen, T. H., Nahar, S., Sarker, N., Baker, S., Bulach, D., . . . Rahman, M. (2021). A comparative whole genome analysis of *Helicobacter pylori* from a human dense South Asian setting. *Helicobacter*, *26*(1), e12766. <https://doi.org/10.1111/hel.12766>

- Qureshi, N. N., Gallaher, B., & Schiller, N. L. (2014). Evolution of amoxicillin resistance of *Helicobacter pylori* in vitro: characterization of resistance mechanisms. *Microb Drug Resist*, 20(6), 509-516. <https://doi.org/10.1089/mdr.2014.0019>
- Rhead, J. L., Letley, D. P., Mohammadi, M., Hussein, N., Mohagheghi, M. A., Eshagh Hosseini, M., & Atherton, J. C. (2007). A new *Helicobacter pylori* vacuolating cytotoxin determinant, the intermediate region, is associated with gastric cancer. *Gastroenterology*, 133(3), 926-936. <https://doi.org/10.1053/j.gastro.2007.06.056>
- Rimbara, E., Noguchi, N., Kawai, T., & Sasatsu, M. (2012). Fluoroquinolone resistance in *Helicobacter pylori*: role of mutations at position 87 and 91 of GyrA on the level of resistance and identification of a resistance conferring mutation in GyrB. *Helicobacter*, 17(1), 36-42. <https://doi.org/10.1111/j.1523-5378.2011.00912.x>
- Rizzato, C., Torres, J., Obazee, O., Camorlinga-Ponce, M., Trujillo, E., Stein, A., . . . Kato, I. (2020). Variations in *cag* pathogenicity island genes of *Helicobacter pylori* from Latin American groups may influence neoplastic progression to gastric cancer. *Sci Rep*, 10(1), 6570. <https://doi.org/10.1038/s41598-020-63463-0>
- Rizzato, C., Torres, J., Plummer, M., Muñoz, N., Franceschi, S., Camorlinga-Ponce, M., . . . Kato, I. (2012). Variations in *Helicobacter pylori* cytotoxin-associated genes and their influence in progression to gastric cancer: implications for prevention. *PLoS One*, 7(1), e29605. <https://doi.org/10.1371/journal.pone.0029605>
- Rodríguez - Gómez, E. R. (2012). *Prevalencia de la resistencia de Helicobacter pylori a tetraciclina en aislamientos colombianos* Pontificia Universidad Javeriana]. <http://hdl.handle.net/10554/11810>.
- Rodríguez Gómez, E. R., Otero Regino, W., Monterrey, P. A., & Trespalacios Rangel, A. A. (2020). *cagA* gene EPIYA motif genetic characterization from Colombian *Helicobacter pylori* isolates: Standardization of a molecular test for rapid clinical laboratory detection. *PLoS One*, 15(1), e0227275. <https://doi.org/10.1371/journal.pone.0227275>
- Román-Román, A., Martínez-Carrillo, D. N., Atrisco-Morales, J., Azúcar-Heziquio, J. C., Cuevas-Caballero, A. S., Castañón-Sánchez, C. A., . . . Fernández-Tilapa, G. (2017). genotype but not. *Gut Pathog*, 9, 18. <https://doi.org/10.1186/s13099-017-0167-z>
- Rubio, A., & Pérez-Pulido, A. J. (2021). Protein-Coding Genes of. *Genes (Basel)*, 12(3). <https://doi.org/10.3390/genes12030377>
- Ruiz-Linares, A. (2014). How genes have illuminated the history of early Americans and Latino Americans. *Cold Spring Harb Perspect Biol*, 7(6). <https://doi.org/10.1101/cshperspect.a008557>
- Safavi, M., Sabourian, R., & Foroumadi, A. (2016). Treatment of *Helicobacter pylori* infection: Current and future insights. *World J Clin Cases*, 4(1), 5-19. <https://doi.org/10.12998/wjcc.v4.i1.5>
- Salimzadeh, L., Bagheri, N., Zamanzad, B., Azadegan-Dehkordi, F., Rahimian, G., Hashemzadeh-Chaleshtori, M., . . . Shirzad, H. (2015). Frequency of virulence factors in *Helicobacter pylori*-infected patients with gastritis. *Microb Pathog*, 80, 67-72. <https://doi.org/10.1016/j.micpath.2015.01.008>
- Sanches, B. S., Martins, G. M., Lima, K., Cota, B., Moretzsohn, L. D., Ribeiro, L. T., . . . Coelho, L. G. (2016). Detection of *Helicobacter pylori* resistance to clarithromycin and fluoroquinolones in Brazil: A national survey. *World J Gastroenterol*, 22(33), 7587-7594. <https://doi.org/10.3748/wjg.v22.i33.7587>
- Saranathan, R., Levi, M. H., Wattam, A. R., Malek, A., Asare, E., Behin, D. S., . . . Szymczak, W. A. (2020). *Helicobacter pylori* Infections in the Bronx, New York: Surveying Antibiotic Susceptibility and Strain Lineage by Whole-Genome Sequencing. *J Clin Microbiol*, 58(3). <https://doi.org/10.1128/JCM.01591-19>

- Savoldi, A., Carrara, E., Graham, D. Y., Conti, M., & Tacconelli, E. (2018). Prevalence of Antibiotic Resistance in *Helicobacter pylori*: A Systematic Review and Meta-analysis in World Health Organization Regions. *Gastroenterology*, 155(5), 1372-1382.e1317. <https://doi.org/10.1053/j.gastro.2018.07.007>
- Schmid, E. N., von Recklinghausen, G., & Ansorg, R. (1990). Bacteriophages in *Helicobacter* (*Campylobacter*) *pylori*. *J Med Microbiol*, 32(2), 101-104. <https://doi.org/10.1099/00222615-32-2-101>
- Secka, O., Berg, D. E., Antonio, M., Corrah, T., Tapgun, M., Walton, R., . . . Thomas, J. E. (2013). Antimicrobial susceptibility and resistance patterns among *Helicobacter pylori* strains from The Gambia, West Africa. *Antimicrob Agents Chemother*, 57(3), 1231-1237. <https://doi.org/10.1128/AAC.00517-12>
- Secka, O., Vale, F. F., Buissonnière, A., Thomas, J. E., Mégraud, F., & Lehours, P. (2017). Phylogeographic agreement between prophage and bacterial housekeeping genes in *Helicobacter pylori* strains from The Gambia. *Helicobacter*. <https://doi.org/10.1111/hel.12394>
- Servetas, S. L., Kim, A., Su, H., Cha, J. H., & Merrell, D. S. (2018). Comparative analysis of the Hom family of outer membrane proteins in isolates from two geographically distinct regions: The United States and South Korea. *Helicobacter*, 23(2), e12461. <https://doi.org/10.1111/hel.12461>
- Shiota, S., Suzuki, R., Matsuo, Y., Miftahussurur, M., Tran, T. T., Binh, T. T., & Yamaoka, Y. (2014). *Helicobacter pylori* from gastric cancer and duodenal ulcer show same phylogeographic origin in the Andean region in Colombia. *PLoS One*, 9(8), e105392. <https://doi.org/10.1371/journal.pone.0105392>
- Shiota, S., Watada, M., Matsunari, O., Iwatani, S., Suzuki, R., & Yamaoka, Y. (2012). *Helicobacter pylori* *iceA*, clinical outcomes, and correlation with *cagA*: a meta-analysis. *PLoS One*, 7(1), e30354. <https://doi.org/10.1371/journal.pone.0030354>
- Smiley, R., Bailey, J., Sethuraman, M., Posecion, N., & Showkat Ali, M. (2013). Comparative proteomics analysis of sarcosine insoluble outer membrane proteins from clarithromycin resistant and sensitive strains of *Helicobacter pylori*. *J Microbiol*, 51(5), 612-618. <https://doi.org/10.1007/s12275-013-3029-5>
- Suerbaum, S., & Michetti, P. (2002). *Helicobacter pylori* infection. *N Engl J Med*, 347(15), 1175-1186. <https://doi.org/10.1056/NEJMra020542>
- Sulakvelidze, A. (2005). Phage therapy: an attractive option for dealing with antibiotic-resistant bacterial infections. *Drug Discov Today*, 10(12), 807-809. [https://doi.org/10.1016/S1359-6446\(05\)03441-0](https://doi.org/10.1016/S1359-6446(05)03441-0)
- Takeuchi, H., Kira, M., Konishi, S., Uchiyama, J., Matsuzaki, S., & Matsumura, Y. (2018). Polymorphisms in the *Helicobacter pylori* NY43 strain and its prophage-cured derivatives. *Microbiology*, 164(6), 877-882. <https://doi.org/10.1099/mic.0.000665>
- Talebi Bezmin Abadi, A. (2017). *Helicobacter pylori* treatment: New perspectives using current experience. *J Glob Antimicrob Resist*, 8, 123-130. <https://doi.org/10.1016/j.jgar.2016.11.008>
- Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., . . . Ostell, J. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res*, 44(14), 6614-6624. <https://doi.org/10.1093/nar/gkw569>
- Taylor, V. L., Fitzpatrick, A. D., Islam, Z., & Maxwell, K. L. (2019). The Diverse Impacts of Phage Morons on Bacterial Fitness and Virulence. *Adv Virus Res*, 103, 1-31. <https://doi.org/10.1016/bs.aivir.2018.08.001>

- Thorell, K., Hosseini, S., Palacios Gonzáles, R. V., Chaotham, C., Graham, D. Y., Paszat, L., . . . Sjöling, Å. (2016). Identification of a Latin American-specific BabA adhesin variant through whole genome sequencing of *Helicobacter pylori* patient isolates from Nicaragua. *BMC Evol Biol*, *16*, 53. <https://doi.org/10.1186/s12862-016-0619-y>
- Thorell, K., Lehours, P., & Vale, F. F. (2017). Genomics of *Helicobacter pylori*. *Helicobacter*, *22 Suppl 1*. <https://doi.org/10.1111/hel.12409>
- Thorell, K., Yahara, K., Berthenet, E., Lawson, D. J., Mikhail, J., Kato, I., . . . Falush, D. (2017). Rapid evolution of distinct *Helicobacter pylori* subpopulations in the Americas. *PLoS Genet*, *13*(2), e1006546. <https://doi.org/10.1371/journal.pgen.1006546>
- Thung, I., Aramin, H., Vavinskaya, V., Gupta, S., Park, J. Y., Crowe, S. E., & Valasek, M. A. (2016). Review article: the global emergence of *Helicobacter pylori* antibiotic resistance. *Aliment Pharmacol Ther*, *43*(4), 514-533. <https://doi.org/10.1111/apt.13497>
- Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., . . . Venter, J. C. (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature*, *388*(6642), 539-547. <https://doi.org/10.1038/41483>
- Torres, J., Correa, P., Ferreccio, C., Hernandez-Suarez, G., Herrero, R., Cavazza-Porro, M., . . . Morgan, D. (2013). Gastric cancer incidence and mortality is associated with altitude in the mountainous regions of Pacific Latin America. *Cancer Causes Control*, *24*(2), 249-256. <https://doi.org/10.1007/s10552-012-0114-8>
- Touchon, M., Bernheim, A., & Rocha, E. P. (2016). Genetic and life-history traits associated with the distribution of prophages in bacteria. *ISME J*, *10*(11), 2744-2754. <https://doi.org/10.1038/ismej.2016.47>
- Trespalcios, A. A., Otero Regino, W., & Mercado Reyes, M. (2010). Resistencia de *Helicobacter pylori* a metronidazol, claritromicina y amoxicilina en pacientes colombianos. *Revista Colombiana de Gastroenterología*, *25*, 31-38.
- Trespalcios, A. A., Otero, W., Caminos, J. E., Mercado, M. M., Avila, J., Rosero, L. E., . . . Graham, D. Y. (2013). Phenotypic and genotypic analysis of clarithromycin-resistant *Helicobacter pylori* from Bogotá D.C., Colombia. *J Microbiol*, *51*(4), 448-452. <https://doi.org/10.1007/s12275-013-2465-6>
- Trespalcios, A. A., Rimbara, E., Otero, W., Reddy, R., & Graham, D. Y. (2015). 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*, *81*(4), 251-255. <https://doi.org/10.1016/j.diagmicrobio.2014.12.003>
- Trespalcios-Rangél, A. A., Otero, W., Arévalo-Galvis, A., Poutou-Piñales, R. A., Rimbara, E., & Graham, D. Y. (2016). Surveillance of Levofloxacin Resistance in *Helicobacter pylori* Isolates in Bogotá-Colombia (2009-2014). *PLoS One*, *11*(7), e0160007. <https://doi.org/10.1371/journal.pone.0160007>
- Trujillo, E., Martínez, T., & Bravo, M. M. (2014). [Genotyping of *Helicobacter pylori* virulence factors *vacA* and *cagA* in individuals from two regions in Colombia with opposing risk for gastric cancer]. *Biomedica*, *34*(4), 567-573. <https://doi.org/10.1590/S0120-41572014000400009>
- Tshibangu-Kabamba, E., & Yamaoka, Y. (2021). *Helicobacter pylori* infection and antibiotic resistance - from biology to clinical implications. *Nat Rev Gastroenterol Hepatol*. <https://doi.org/10.1038/s41575-021-00449-x>
- Tuan, V. P., Narith, D., Tshibangu-Kabamba, E., Dung, H. D. Q., Viet, P. T., Sokomoth, S., . . . Yamaoka, Y. (2019). A Next-Generation Sequencing-Based Approach to Identify Genetic

Determinants of Antibiotic Resistance in Cambodian. *J Clin Med*, 8(6). <https://doi.org/10.3390/jcm8060858>

Turkina, M. V., Olofsson, A., Magnusson, K. E., Arnqvist, A., & Vikström, E. (2015). Helicobacter pylori vesicles carrying CagA localize in the vicinity of cell-cell contacts and induce histone H1 binding to ATP in epithelial cells. *FEMS Microbiol Lett*, 362(11). <https://doi.org/10.1093/femsle/fnv076>

Uchiyama, J., Takeuchi, H., Kato, S., Gamoh, K., Takemura-Uchiyama, I., Ujihara, T., . . . Matsuzaki, S. (2013). Characterization of Helicobacter pylori bacteriophage KHP30. *Appl Environ Microbiol*, 79(10), 3176-3184. <https://doi.org/10.1128/AEM.03530-12>

Uchiyama, J., Takeuchi, H., Kato, S., Takemura-Uchiyama, I., Ujihara, T., Daibata, M., & Matsuzaki, S. (2012). Complete genome sequences of two Helicobacter pylori bacteriophages isolated from Japanese patients. *J Virol*, 86(20), 11400-11401. <https://doi.org/10.1128/JVI.01767-12>

Umit, H., Tezel, A., Bukavaz, S., Unsal, G., Otkun, M., Soyly, A. R., . . . Bilgi, S. (2009). The relationship between virulence factors of Helicobacter pylori and severity of gastritis in infected patients. *Dig Dis Sci*, 54(1), 103-110. <https://doi.org/10.1007/s10620-008-0316-9>

Uribe Echeverry, P. T., Acosta Cerquera, M. A., Arturo Arias, B. L., Jaramillo Arredondo, M. D. S., Betancur Pérez, J. F., & Pérez Agudelo, J. M. (2018). Prevalencia genotípica de cagA y vacA en aislamientos de Helicobacter pylori de pacientes colombianos. *Revista Cubana de Medicina Tropical*, 70, 18-26.

Vale, F. F., & Lehours, P. (2018). Relating Phage Genomes to Helicobacter pylori Population Structure: General Steps Using Whole-Genome Sequencing Data. *Int J Mol Sci*, 19(7). <https://doi.org/10.3390/ijms19071831>

Vale, F. F., Nunes, A., Oleastro, M., Gomes, J. P., Sampaio, D. A., Rocha, R., . . . Lehours, P. (2017). Genomic structure and insertion sites of Helicobacter pylori prophages from various geographical origins. *Sci Rep*, 7, 42471. <https://doi.org/10.1038/srep42471>

Vale, F. F., Vadivelu, J., Oleastro, M., Breurec, S., Engstrand, L., Perets, T. T., . . . Lehours, P. (2015). Dormant phages of Helicobacter pylori reveal distinct populations in Europe. *Sci Rep*, 5, 14333. <https://doi.org/10.1038/srep14333>

van Doorn, L. J., Figueiredo, C., Sanna, R., Plaisier, A., Schneeberger, P., de Boer, W., & Quint, W. (1998). Clinical relevance of the cagA, vacA, and iceA status of Helicobacter pylori. *Gastroenterology*, 115(1), 58-66. [https://doi.org/10.1016/s0016-5085\(98\)70365-8](https://doi.org/10.1016/s0016-5085(98)70365-8)

van Tonder, A. J., Mistry, S., Bray, J. E., Hill, D. M., Cody, A. J., Farmer, C. L., . . . Brueggemann, A. B. (2014). Defining the estimated core genome of bacterial populations using a Bayesian decision model. *PLoS Comput Biol*, 10(8), e1003788. <https://doi.org/10.1371/journal.pcbi.1003788>

Vianna, J. S., Ramis, I. B., Halicki, P. C., Gastal, O. L., Silva, R. A., Junior, J. S., . . . da Silva, P. E. (2015). Detection of Helicobacter pylori CagA EPIYA in gastric biopsy specimens and its relation to gastric diseases. *Diagn Microbiol Infect Dis*, 83(2), 89-92. <https://doi.org/10.1016/j.diagmicrobio.2015.05.017>

Vilar E Silva, A., Junior, M. R., Vinagre, R. M., Santos, K. N., da Costa, R. A., Fecury, A. A., . . . Martins, L. C. (2014). Evaluation of the Pattern of EPIYA Motifs in the Helicobacter pylori cagA Gene of Patients with Gastritis and Gastric Adenocarcinoma from the Brazilian Amazon Region. *Int J Bacteriol*, 2014, 418063. <https://doi.org/10.1155/2014/418063>

Watada, M., Shiota, S., Matsunari, O., Suzuki, R., Murakami, K., Fujioka, T., & Yamaoka, Y. (2011). Association between Helicobacter pylori cagA-related genes and clinical outcomes in Colombia and Japan. *BMC Gastroenterol*, 11, 141. <https://doi.org/10.1186/1471-230X-11-141>

- Wattam, A. R., Abraham, D., Dalay, O., Disz, T. L., Driscoll, T., Gabbard, J. L., . . . Sobral, B. W. (2014). PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res*, 42(Database issue), D581-591. <https://doi.org/10.1093/nar/gkt1099>
- Wong, M. C. S., Huang, J., Chan, P. S. F., Choi, P., Lao, X. Q., Chan, S. M., . . . Liang, P. (2021). Global Incidence and Mortality of Gastric Cancer, 1980-2018. *JAMA Netw Open*, 4(7), e2118457. <https://doi.org/10.1001/jamanetworkopen.2021.18457>
- Xia, Y., Yamaoka, Y., Zhu, Q., Matha, I., & Gao, X. (2009). A comprehensive sequence and disease correlation analyses for the C-terminal region of CagA protein of *Helicobacter pylori*. *PLoS One*, 4(11), e7736. <https://doi.org/10.1371/journal.pone.0007736>
- Xu, C., Soyfoo, D. M., Wu, Y., & Xu, S. (2020). Virulence of *Helicobacter pylori* outer membrane proteins: an updated review. *Eur J Clin Microbiol Infect Dis*, 39(10), 1821-1830. <https://doi.org/10.1007/s10096-020-03948-y>
- Yahara, K., Lehours, P., & Vale, F. F. (2019). Analysis of genetic recombination and the pan-genome of a highly recombinogenic bacteriophage species. *Microb Genom*. <https://doi.org/10.1099/mgen.0.000282>.
- Yamaoka, Y. (2008). Roles of *Helicobacter pylori* BabA in gastroduodenal pathogenesis. *World J Gastroenterol*, 14(27), 4265-4272. <https://doi.org/10.3748/wjg.14.4265>
- Yamaoka, Y. (2009). *Helicobacter pylori* typing as a tool for tracking human migration. *Clin Microbiol Infect*, 15(9), 829-834. <https://doi.org/10.1111/j.1469-0691.2009.02967.x>
- Yamaoka, Y. (2010). Mechanisms of disease: *Helicobacter pylori* virulence factors. *Nat Rev Gastroenterol Hepatol*, 7(11), 629-641. <https://doi.org/10.1038/nrgastro.2010.154>
- Yamaoka, Y., Kodama, T., Gutierrez, O., Kim, J. G., Kashima, K., & Graham, D. Y. (1999). Relationship between *Helicobacter pylori* iceA, cagA, and vacA status and clinical outcome: studies in four different countries. *J Clin Microbiol*, 37(7), 2274-2279. <https://doi.org/10.1128/JCM.37.7.2274-2279.1999>
- Yamaoka, Y., Orito, E., Mizokami, M., Gutierrez, O., Saitou, N., Kodama, T., . . . Graham, D. Y. (2002). *Helicobacter pylori* in North and South America before Columbus. *FEBS Lett*, 517(1-3), 180-184. [https://doi.org/10.1016/s0014-5793\(02\)02617-0](https://doi.org/10.1016/s0014-5793(02)02617-0)
- Yepes, C. A., Rodríguez Varón, A., Ruiz Morales, Á., & Ariza, B. (2008). Resistencia antibiótica del *Helicobacter pylori* en el Hospital Universitario San Ignacio de Bogotá. *Acta Medica Colombiana*, 33, 11-14.
- Yonezawa, H., Osaki, T., Hanawa, T., Kurata, S., Ochiai, K., & Kamiya, S. (2013). Impact of *Helicobacter pylori* biofilm formation on clarithromycin susceptibility and generation of resistance mutations. *PLoS One*, 8(9), e73301. <https://doi.org/10.1371/journal.pone.0073301>
- You, Y., He, L., Zhang, M., & Zhang, J. (2015). Comparative genomics of a *Helicobacter pylori* isolate from a Chinese Yunnan Naxi ethnic aborigine suggests high genetic divergence and phage insertion. *PLoS One*, 10(3), e0120659. <https://doi.org/10.1371/journal.pone.0120659>
- Zhang, X. S., Tegtmeyer, N., Traube, L., Jindal, S., Perez-Perez, G., Sticht, H., . . . Blaser, M. J. (2015). A specific A/T polymorphism in Western tyrosine phosphorylation B-motifs regulates *Helicobacter pylori* CagA epithelial cell interactions. *PLoS Pathog*, 11(2), e1004621. <https://doi.org/10.1371/journal.ppat.1004621>
- Álvarez, A., Moncayo, J. I., Santacruz, J. J., Corredor, L. F., Reinoso, E., Martínez, J. W., & Beltrán, L. (2009). Resistencia a metronidazol y claritromicina en aislamientos de *Helicobacter pylori* de pacientes dispépticos en Colombia. *Revista médica de Chile*, 137, 1309-1314.

Łoś, M., & Węgrzyn, G. (2012). Pseudolysogeny. *Adv Virus Res*, 82, 339-349. <https://doi.org/10.1016/B978-0-12-394621-8.00019-4>

Šterbenc, A., Lunar, M. M., Homan, M., Luzar, B., Zidar, N., & Poljak, M. (2020). Prevalence of the. *Can J Infect Dis Med Microbiol*, 2020, 4080248. <https://doi.org/10.1155/2020/4080248>

ANEXOS

ANEXO 1. Bacteriophages of *Helicobacter pylori*

Muñoz, A. B., Stepanian, J., Trespalacios, A. A., & Vale, F. F. (2020). Bacteriophages of *Helicobacter pylori*. *Frontiers in microbiology*, *11*, 549084. <https://doi.org/10.3389/fmicb.2020.549084>



Bacteriophages of *Helicobacter pylori*

Angela B. Muñoz^{1,2*}, Johanna Stepanian¹, Alba Alicia Trespalacios¹ and Filipa F. Vale^{2*}

¹ Infectious Diseases Research Group, Microbiology Department, Sciences Faculty, Pontificia Universidad Javeriana, Bogotá, Colombia, ² Host–Pathogen Interactions Unit, Research Institute for Medicines (iMed-ULisboa), Faculdade de Farmácia, Universidade de Lisboa, Lisbon, Portugal

The bacterium *Helicobacter pylori* colonize the stomach in approximately half of the world's population. Infection with this bacterium is associated with gastritis, peptic ulcer, adenocarcinoma, and gastric mucosa-associated lymphoid tissue lymphoma. Besides being a pathogen with worldwide prevalence, *H. pylori* show increasingly high antibiotic resistance rates, making the development of new therapeutic strategies against this bacterium challenging. Furthermore, *H. pylori* is a genetically diverse bacterium, which may be influenced by the presence of mobile genomic elements, including prophages. In this review, we analyze these issues and summarize various reports and findings related to phages and *H. pylori*, discussing the relationship between the presence of these elements and the genomic diversity, virulence, and fitness of this bacterium. We also analyze the state of the knowledge on the potential utility of bacteriophages as a therapeutic strategy for *H. pylori*.

Keywords: *Helicobacter pylori*, bacteriophage, phage therapy, prophage genetic diversity, phage–host interaction

OPEN ACCESS

Edited by:

Eric Altermann, AgResearch Ltd, New Zealand

Reviewed by:

Jeremy J. Barr, Monash University, Australia
Sarah O'Flaherty, North Carolina State University,

United States Junko Akada, Oita University, Japan

*Correspondence: Angela B. Muñoz munozangela@javeriana.edu.co
Filipa F. Vale f.vale@ff.ulisboa.pt; vale.filipa@gmail.com

Specialty section: This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal *Frontiers in Microbiology*

Received: 04 April 2020

Accepted: 21 October 2020

Published: 12 November 2020

Citation: Muñoz AB, Stepanian J, Trespalacios AA and Vale FF (2020) Bacteriophages of *Helicobacter pylori*.

Front. Microbiol. 11:549084.

doi:10.3389/fmicb.2020.549084

INTRODUCTION

Helicobacter pylori is a spiral-shaped, microaerophilic, Gram-negative, pathogenic bacterium that colonizes and persists in the human stomach. *H. pylori* infection is considered the most frequent chronic bacterial infection worldwide (Camilo et al., 2017). All individuals infected with *H. pylori* present with chronic gastritis, while 20% develop peptic ulcer and less than 1% develop adenocarcinoma or mucosa-associated lymphoid tissue (MALT)-type gastric lymphoma. Further, *H. pylori* infection is recognized as the leading risk factor for the development of gastric cancer (Polk and Peek, 2010). Numerous factors have been described as responsible for the progression of the infection to more severe diseases: although these factors may be inherent to the host (Figueiredo et al., 2017; Miftahussurur et al., 2017), several *H. pylori* virulence factors have also been associated with malignancy. The most frequently associated are the Cag pathogenicity island (*cagPAI*) and the VacA vacuolating cytotoxin (Denic et al., 2020).

CagPAI is a genetic locus of 40 kb, comprising 31 genes that encode a type IV secretion system (T4SS). The T4SS enables the injection of bacterial components, such as the CagA oncoprotein, into host gastric epithelial cells (Soluri et al., 2020). VacA is a toxin secreted by *H. pylori* that inserts into host cell membranes to form chloride-sensitive channels and disrupt endolysosomal trafficking, causing an accumulation of dysfunctional lysosomes and autophagosomes (Denic et al., 2020). Most *H. pylori* strains isolated have the *vacA* gene. However, differences in the alleles of the signal (s) and middle (m) regions of this gene are responsible for differing levels of cytotoxicity (Oktem-Okullu et al., 2020).

In addition to the virulence of *H. pylori*, another critical challenge for clinicians has been to find an optimal treatment that takes into account the high levels of resistance of *H. pylori* to antibiotics (Camargo et al., 2014; Talebi Bezin Abadi, 2017).

H. pylori cure rates using antibiotic treatment have been reported to be as low as 57% (Dos Santos and Carvalho, 2015). The minimum acceptable rate for a first attempt based treatment is 90% (Nijevitch et al., 2014; Jones et al., 2017). Current treatments are based on the use of antibiotics combined with a proton pump inhibitor (PPI) (Malfertheiner et al., 2017). The PPI is used to increase intragastric pH because, at an acidic pH, *H. pylori* transforms into its antibiotic-resistant coccoid form (Ierardi et al., 2019). Furthermore, acidic pH decreases the antimicrobial activity and half-life of antibiotics (Otero et al., 2018).

The increasing rates of *H. pylori* antibiotic resistance have necessitated the development of new therapeutic strategies (Vitor and Vale, 2011). An emerging alternative treatment for antibiotic-resistant bacterial infections is phage-therapy, which uses bacteriophages, also known as phages, to eliminate a bacterial population (Lin D. M. et al., 2017). The resurgence of this therapy has been due mainly to studies showing that phages are highly specific and easy to isolate (Loc-Carrillo and Abedon, 2011); moreover, phage therapy has demonstrated effectiveness in treating various infections, even curing chronic infections (Abedon, 2019). Despite the benefits of this therapy, its use for the treatment of *H. pylori* infection is a distant goal, considering that the understanding of *H. pylori* phage biology is still in its infancy.

Phages, which are viral particles that infect bacteria, may be lytic, lysogenic, or pseudo-lysogenic. Lytic phages recognize bacterial surfaces and inject their nucleic acids into the host cell; then, they assemble, multiply, and finally disrupt the cell to release phage progeny that infect new bacterial cells (Sulakvelidze, 2005). In lysogenic phages, phage DNA integrates into the bacterial genome, forming prophages (also known as template phages). Prophages promote bacterial evolution via horizontal gene transfer and induce the transduction of several genes involved in the biological behaviors of the bacterium (Brüssow et al., 2004; Paterson et al., 2010; Touchon et al., 2016). Pseudo-lysogenic phages are present as episomes (i.e., the genetic material is not integrated into the bacterial genome) and postpone cell lysis in nutrient-depleted hosts (Uchiyama et al., 2013); both lytic and lysogenic phages, under certain conditions, may acquire episomal forms (Łoś and Węgrzyn, 2012).

Bacteriophages represent the absolute majority of all organisms in the biosphere (Hatfull and Hendrix, 2011). Among the various bacteriophages, the discovery of *H. pylori* prophages has been of particular importance in explaining this bacterium's tremendous genetic diversity. This review aims to collate and analyze relevant publications regarding phages related to *H. pylori* and to understand how these phages impact the diversity and virulence of this bacterium. Additionally, we summarize reports on lytic phages with activity against *H. pylori*.

H. pylori Prophages

The first observations of phage-like intracellular particles in preparations of *H. pylori* were made shortly after the

discovery of this bacterium (Marshall et al., 1987; Goodwin et al., 1989). In the early 1990s, a lysogenic strain of *H. pylori* that produced phage particles spontaneously was described (Schmid et al., 1990). Three years later, the lytic cycle of this phage was reproduced. Negative stain electron microscopy revealed that phage heads of around 50 to 60 nm and the DNA length was estimated to be 22,000 bp (Heintschel Von Heinegg et al., 1993).

With recent advances in sequencing technologies, studies aiming at the discovery of phages are increasing. Since 2011, investigations focusing on the analysis of *Helicobacter* spp. genomes have been published that corroborate the presence of prophages in the bacterial genome and analyze their functions in this context (Lehours et al., 2011; Luo et al., 2012; Uchiyama et al., 2012, 2013, 2016; Vale et al., 2015, 2017; Secka et al., 2017; Vale and Lehours, 2018) (Table 1). These studies report that prophages are present in around 20% of *H. pylori* isolates. The first of these studies reported a prophage sequence in *H. pylori* isolated from a patient with MALT lymphoma. This prophage, named PhiHp33, was isolated from *H. pylori* strain B45 and could be induced by UV light. Genomic sequence analysis demonstrated that *H. pylori* strains from different regions carry full or partial prophage sequences. Analysis of integrase and holin gene sequences allowed differentiation of strains according to their geographical origin, generating results consistent with classification based on multi-locus sequence typing (MLST) (Lehours et al., 2011; Vale et al., 2015; Secka et al., 2017).

In 2012, the isolation of a new phage (1961P) from Taiwanese *H. pylori* strains was reported. This phage showed characteristics compatible with belonging to the family *Podoviridae*. Further, sequence analysis identified other similar prophages integrated into the genomes of different *H. pylori* strains that had previously been sequenced (Luo et al., 2012). Additionally, in the same year, the complete genome sequences of two *H. pylori* bacteriophages (KHP30 and KHP40) isolated from Japanese patients were reported. These prophages were obtained after spontaneous release of viral particles by *H. pylori* strains (Uchiyama et al., 2012). The KHP30 phage was characterized as spherical, with a lipid envelope. It was present in infected bacteria as an episome, indicating that it likely represents a new phage family (Uchiyama et al., 2013). Subsequently, the possibility of lysogeny of active KHP30-like phages in 174 Japanese *H. pylori* strains was examined because their genomes contain a putative integrase gene; one strain, NY40, was found to include a KHP30-like prophage sequence. In this strain, the attachment sites, *attL* and *attR*, were almost the same place in the genome. Further, their sequences were detected previously by Lehours et al., suggesting an active parental KHP30-like phage integrated into the ancestral NY40 genome in a site-specific manner (Uchiyama et al., 2016).

Analysis of KHP30 has continued, with researchers investigating its effects on the behaviors of the host strain (NY43) and demonstrating differences between the characteristics of *H. pylori* strains infected with prophage and novel prophage-free derivative strains. A recent study indicated that the presence of bacteriophages induces genetic alterations in the host genome, leading to the continuous production of prophage-free derivatives that coexist in *H. pylori* microbial communities

TABLE 1 | Main bacteriophages of *H. pylori*.

Bacteriophage	Type	Family	Genome size (Kb)	Details	References
1961P 13.3 nm	Prophage	<i>Podoviridae</i>	26.8	Accession number: NC_019512.1 Head: 68–74 nm Tail: 23 ×	(Luo et al.,2012)
<i>Campylobacter piloridis</i> strain	Prophage	No data	No data	Intracellular particles of 40 nm in diameter	(Marshall et al.,1987)
<i>Campylobacter pylori</i> strain	Prophage	No data	No data	Intracellular particles of 85 nm in diameter	(Goodwin et al.,1989)
De-M53-M	Prophages	No data	28.1	Prophage population: hpNEurope Accession number: KX119205	(Vale et al.,2017)
Fr-ANT170-U	Prophage	No data	31.2	Prophage population: hpAfrica1 Accession number: KX119201	(Vale et al.,2017)
Fr-B41-M	Prophage	No data	29.4	Prophage population: hpSWEurope Accession number: KX119190	(Vale et al.,2017)
Fr-B58-M	Prophage	No data	22.6	Prophage population: hpEastAsia Accession number: KX119193	(Vale et al.,2017)
Fr-G12-G	Prophage	No data	28.6	Prophage population: hpEastAsia Accession number: KX119194	(Vale et al.,2017)
Fr-GC43-A	Prophage	No data	33.0	Prophage population: hpEastAsia Accession number: KX119195	(Vale et al.,2017)
Fr-MEG235-U	Prophage	No data	31.2	Prophage population: hpAfrica1 Accession number: KX119200	(Vale et al.,2017)
HP1	Prophage	<i>Siphoviridae</i>	22	Head:50–60 nm Tail:170 × 9.5 nm	(Heintschel Von Heinegg et al.,1993)
KHP30	Prophage	Cannot be classified / new family	26.2	Episome Accession number: NC_019928.1 Head: 67–71 nm Without tail	(Uchiyama et al.,2012, 2013;Takeuchi et al., 2018)
KHP40	Prophage	No data	26.4	KHP30-like Accession number: NC_019931.1	(Uchiyama et al.,2012)
PhiHp33	Prophage	<i>Siphoviridae</i>	24.6	Inducible by UV light Accession number: NC_016568 Head: 55–70 nm Tail: 92 × 6 nm	(Lehours et al.,2011)
Pt-1293-U	Prophage	No data	30.1	Prophage population: hpAfrica1 Accession number: KX119202	(Vale et al.,2017)
Pt-1846-U	Prophage	No data	28.0	Prophage population: hpAfrica1 Accession number: KX119176	(Vale et al.,2017)
Pt-1918-U	Prophage	No data	28.7	Prophage population: hpSWEurope Accession number: KX119192	(Vale et al.,2017)
Pt-212-99R-U	Prophage	No data	23.0	Prophage population: hpAfrica1 Accession number: KX119193	(Vale et al.,2017)
Pt-228_99-G	Prophage	No data	30.1	Prophage population: hpAfrica1 Accession number: KX119175	(Vale et al.,2017)
Pt-4472-G	Prophage	No data	27.6	Prophage population: hpSWEurope Accession number: KX119190	(Vale et al.,2017)
Pt-4481-G	Prophage	No data	25.4	Prophage population: hpAfrica1 Accession number: KX119196	(Vale et al.,2017) Pt-4497-
U	Prophage	No data	29.4	Prophage population: hpSWEurope Accession number: KX119191	(Vale et al.,2017) Pt-5322-G
Prophage	No data	28.3	Prophage population: hpAfrica1 Accession number: KX119198	(Vale et al.,2017) Pt-5771-G	Prophage
No data	29.8	Prophage population: hpAfrica1 Accession number: KX119199	(Vale et al.,2017) Pt-B89-G	Prophage	No data
27.4	Prophage population: hpAfrica1 Accession number: KX119203	(Vale et al.,2017) Pt-B92-G	Prophage	No data	30.5
Prophage population: hpAfrica1 Accession number: KX119197	(Vale et al.,2017) Sw-A626-G	Prophage	No data	31	Prophage
population: hpNEurope Accession number: KX119177	(Vale et al.,2017) Sw-577-G	Prophage	No data	26.9	Prophage population:
hpNEurope Accession number: KX119204	(Vale et al.,2017) UK-EN31-U	Prophage	No data	30.5	Prophage population: hpNEurope
Accession number: KX119174	(Vale et al.,2017) UK-EN32-U	Prophages	No data	29.9	Prophage population: hpNEurope
Accession number: KX119206	(Vale et al.,2017) ϕ HPE1 Lytic <i>Podoviridae</i>	No data	Titer: 10 ⁹ PFU/ml Head: 62 nm Short noncontractile tail: 12 × 6 nm		(Abdel-Halim and Askora,2013)
ϕ HPE2	Lytic	<i>Siphoviridae</i>	No data	Titer: 10 ¹⁰ PFU/ml Head: 95 nm Tail: 180 × 15 nm	(Abdel-Halim and Askora,2013)
Hp ϕ	Lytic	No data	No data	Used in synergy with lactoferricin and hydroxyapatite. Adsorption rate: 1.89 × 10 ⁹ mL/min Latent period: 45 min Burst size: 80 PFU	(Cuomo et al.,2020)
<i>H. pylori</i> strain BAPOUI	Lytic	No data	No data	Obtained from human feces Head: 100 nm Without tail	(Vale et al.,2008)

and confer diversity in morphology, motility, viability, and pathogenicity. Further, this study reported that KHP30 prophage could specifically suppress CagA expression (Takeuchi et al., 2018). Another study conducted using an *H. pylori* strain isolated from an ethnic minority patient in China revealed the presence of a 32.5 kb prophage integrated into the genome and demonstrated that most of its genes (30/33) were the same as

those of KHP30; however, in this case, the prophage was inserted between two putative virulence genes, *oipA* and *hombB*, which encode adherence factors that mediate interactions between *H. pylori* and the host microenvironment. The authors suggest that the presence of a prophage inserted in this region indicates that HomB and OipA may act as receptors for the phage (You et al.,2015).

In addition to the characterization of *H. pylori* prophages, the relationship between the presence of prophages and bacterial genetic diversity has been widely discussed. *H. pylori* is a genetically diverse species. This diversity has been associated with the geographic origin of populations (Matta et al., 2017) and likely influenced by the presence of mobile genomic elements, such as prophages, among other factors (Vale and Lehours, 2018). Since 2015, Vale et al. have reported that *H. pylori* prophages may contribute to bacterial genetic diversity. Most observations indicated a phylogeographic agreement between phage and bacterial genes, suggesting a co-evolution model between the virus and its bacterial host. Nonetheless, some prophages were assigned to populations distinct from their host, while others exhibited recombination signs among populations (Yahara et al., 2019). The phylogeographic distribution of

H. pylori prophages can be characterized by Bayesian clustering analysis, based on a typing scheme that includes the sequences of two prophage genes: integrase (which is responsible for the integration of the phage genome into the bacterial chromosome) and holin (which is involved in cell lysis during a lytic cycle). Four prophage populations have been described: hpAfrica1, hpEastAsia, hpNEurope, and hpSWEurope, notably subdividing the European population (Vale et al., 2015). Recently, based on the whole-genome analysis of *H. pylori*, two subpopulations of hpEurope were also observed (Thorell et al., 2017).

Available information on *H. pylori* prophage diversity indicates that prophages and bacteria share a complex evolutionary history (Vale et al., 2017) and that the bacterial genome has been widely modified, in diverse regions, via horizontal gene transfer (Takeuchi et al., 2018). A few studies have investigated the presence of phages in other *H. pylori* strains; for example, a study from 2016 reported the detection of a 31.7 kb prophage in an *H. pylori* strain isolated from a Mexican patient with gastric cancer (Mucito-Varela et al., 2016). Further, Kumar et al. (2017) analyzed *H. pylori* strains from Kuwait and reported a prophage in a strain isolated from a patient with active chronic gastritis and focal intestinal metaplasia. In silico analyses have revealed prophage orthologous genes in the genomes of bacterial strains with virulence genes such as *cagA* or *vacA* (Kyrillos et al., 2016). Reports suggest that prophages inserted in the *H. pylori* genome represent essential elements for the adaptation of this bacterium to hostile environments, because metabolic advantages or tolerance mechanisms that can be influenced by prophages may improve *H. pylori* competitiveness. Nevertheless, no specific prophage functions have, as yet, been recognized in *H. pylori*, and the roles of prophage in disease have rarely been recorded (Vale et al., 2017). Long-term colonization of the human stomach by *H. pylori* has allowed it to evolve together with its human host for more than 100,000 years (Moodley et al., 2012). Furthermore, because of frequent mutation and recombination events and horizontal gene transfer events, *H. pylori* is an extremely diverse bacterium. Specifically, *H. pylori* is naturally competent for transformation and highly competent in recombination, making the exchange of chromosomal DNA fragments between strains frequent and highly efficient (Fischer et al., 2020). Phages have generally been considered as actively involved in the processes of

recombination and horizontal genetic exchange. In this context, a recent study presented a quantitative analysis of the described recombination events of *H. pylori* phage genes, revealing frequent recombination among phage core genes previously reported as less prone to recombination. This observation appears to reflect the co-evolutionary relationships of these phages with their hosts, revealing the evolutionary arms race that exists, either to contribute to phage escape from bacterial immunity or to protect hosts that produce defective phages. The high recombination rate in *H. pylori* phages suggests that they are among the most recombinogenic phages on earth (Yahara et al., 2019). To further advance knowledge of this topic, it is important to investigate the biological and genomic characteristics of *H. pylori* phages in detail. Likewise, valuable information may emerge from the continued search for phage in strains associated with different pathologies and from various geographical locations. Furthermore, important insights may be obtained through investigation of phage insertion sites and their roles in the bacterial genome, considering that prophages can act as gene reservoirs, which may benefit pathogens in ways that are just beginning to be determined.

Lytic Bacteriophages of *H. pylori*

The rise of antibiotic resistance increased the interest in studying bacteriophages, particularly the lytic ones. Indeed, phage therapy has various potential advantages than antibiotic use because phages and phage lysins, are highly specific, affecting the target strain but not the microbiome (Brüssow, 2017). Further, phages only replicate at the infection site, and no secondary effects have been described (Matsuzaki et al., 2005). Although these findings indicate the potential to use phages to eradicate *H. pylori*, reports in this field are scarce.

At the time of this review, there were only three studies on lytic bacteriophages (Table 1): the first was published by Vale et al. (2008), who isolated one lytic phage from human feces. Not many characteristics of this phage are known. It was only reported that it is a phage without a tail and an approximate size of 100 nm. The second study was published in 2013 by Abdel-Haliem and Askora, who isolated and characterized two lytic phages (OHPE1 and OHPE2) from wastewater (Abdel-Haliem and Askora, 2013). These phages were classified, by electronic microscopy, as belonging to the *Podoviridae* and *Siphoviridae* families, respectively. The titers for OHPE1 and OHPE2 were 10^9 PFU/ml and 10^{10} PFU/ml, respectively. These studies did not offer data on the latent period, burst size, or antimicrobial potential of these phages.

A third recent study by Cuomo et al. (2020) describes a new therapeutic option based on using a lytic phage of *H. pylori* (Hp ϕ), isolated from gastric biopsies. The phage was tested alone and combined with lactoferrin (LA) and adsorbed on hydroxyapatite (HA) nanoparticles. The characterization of Hp ϕ revealed that the adsorption rate was 1.89×10^9 mL/min, the latent period was 45 min, and the burst size was only 80 PFU. These analyses showed that the use of the complex (Hp ϕ + LF-HA) enhances the activity of Hp ϕ up to 4 times. The use of HA as a vehicle can improve the natural properties of both Hp ϕ and LF because it protects the gastric acid environment.

An important finding, considering that phages are often sensitive to extreme pH values such as those associated with stomach acidity (Dąbrowska and Abedon, 2019).

The studies available to date have not addressed certain aspects important for defining phage-therapy. The genome sequences of the phages remain unavailable. Genomic analysis could determine whether the phage genome comprises genes involved in lysogenic events, encode toxins, or represent determinants of antimicrobial resistance, or whether they code for lytic cycle repressors, or site-specific integrases or recombinases (Fernández et al., 2019). Likewise, considering the specific defense mechanisms mounted by bacteria against bacteriophage infections, it is essential to address the possible resistance that

H. pylori may develop against these phages (Hyman and Abedon, 2010). The study by Cuomo et al. (2020) does not define the minimum effective combined doses of H_p φ and LF-HA.

Additional to the use of lytic phages, the use of phage lysins and related bacteriolytic enzymes are also being considered in phage therapy (Fischetti, 2018). Many phages have polysaccharide depolymerases and lysins that recognize, bind, and degrade the polysaccharide compounds (Latka et al., 2017). The enzyme holin could also be used as a therapeutic alternative. This enzyme is involved in the holin-endolysin pathway responsible for the first steps of bacterial lysis, which begins when the holin forms micrometer-scale holes in the bacterial inner membrane, releasing active endolysin into the periplasm to degrade peptidoglycan (Cahill and Young, 2019). Studies have evaluated the effect of enzymes for the treatment of bacterial infections mainly caused by gram-negative bacteria, and have shown efficacy in reducing bacterial virulence (D'andrea et al., 2017; Lin H. et al., 2017; Majkowska-Skrobek et al., 2018; Solovieva et al., 2018). Considering the presence of prophages in *H. pylori* and the background in other bacterial models, the investigation of these kinds of enzymes in *H. pylori* is feasible.

CONCLUSION

The findings reviewed here show that *H. pylori* genetic diversity is influenced by prophages, which are frequently found in strains of this bacterium, represent a structured population, and are highly recombinogenic. The additional genetic diversity that phages

may provide to *H. pylori* confers advantages to this bacterium in terms of persistence and colonization of different human populations. The study of *H. pylori* phages has gained importance considering the substantial challenges posed by infection with this bacterium, including low eradication rates and antibiotic resistance. Phages appear to be a promising approach for the design of future therapies. Advances in research to improve understanding of phage and *H. pylori* interactions are required, and studies investigating whether prophage genes belonging to the lysis cassette are useful for phage-therapy are warranted. Further, identification of more lytic phages of *H. pylori* would allow considering the phage-therapy as an alternative approach for eradicating *H. pylori*, which would undoubtedly mark a milestone in the history of applications of *H. pylori* phages.

AUTHOR CONTRIBUTIONS

AM, JS, AT, and FV conceived the review. AM and JS collected the information. AM and FV wrote the manuscript. All authors contributed to the manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by the following sources: AM is a recipient of a scholarship from the Centro de Estudios Interdisciplinarios Básicos y Aplicados (CEIBA Foundation), Colombia. AM, JS, and AT are recipients of a project grant (120380763025/2018) from the Departamento Administrativo de Ciencia, Tecnología e Innovación de Colombia (Colciencias) and a project grant (PPTA_7676) from Research Vice-Rector, Pontificia Universidad Javeriana. FV is the recipient of a project grant (PTDC/BTM-SAL/28978/2017) from the Fundação para a Ciência e a Tecnologia (FCT).

ACKNOWLEDGMENTS

We thank the entities that financially supported the development of this work.

REFERENCES

- Abdel-Haliem, M. E. F., and Askora, A. (2013). Isolation and characterization of bacteriophages of *Helicobacter pylori* isolated from Egypt. *Future Virol.* 8, 821–826. doi: 10.2217/fvl.13.58
- Abedon, S. T. (2019). Use of phage therapy to treat long-standing, persistent, or chronic bacterial infections. *Adv. Drug Deliv. Rev.* 145, 18–39. doi: 10.1016/j.addr.2018.06.018
- Brüssow, H. (2017). Infection therapy: the problem of drug resistance - and possible solutions. *Microb. Biotechnol.* 10, 1041–1046. doi: 10.1111/1751-7915.12777
- Brüssow, H., Canchaya, C., and Hardt, W. D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol. Mol. Biol. Rev.* 68, 560–602. doi: 10.1128/MMBR.68.3.560-602.2004
- Cahill, J., and Young, R. (2019). Phage lysis: multiple genes for multiple barriers. *Adv. Virus Res.* 103, 33–70. doi: 10.1016/bs.aivir.2018.09.003
- Camargo, M. C., García, A., Riquelme, A., Otero, W., Camargo, C. A., Hernandez-García, T., et al. (2014). The problem of *Helicobacter pylori* resistance to antibiotics: a systematic review in Latin America. *Am. J. Gastroenterol.* 109, 485–495. doi: 10.1038/ajg.2014.24
- Camilo, V., Sugiyama, T., and Touati, E. (2017). Pathogenesis of *Helicobacter pylori* infection. *Helicobacter* 22(Suppl. 1), e12405. doi: 10.1111/hel.12405
- Cuomo, P., Papaiani, M., Fulgione, A., Guerra, F., Capparelli, R., and Medaglia, C. (2020). An innovative approach to control. *Microorganisms* 8:1214. doi: 10.3390/microorganisms8081214
- Dąbrowska, K., and Abedon, S. T. (2019). Pharmacologically aware phage therapy: pharmacodynamic and pharmacokinetic obstacles to phage antibacterial action in animal and human bodies. *Microbiol. Mol. Biol. Rev.* 83:e00012-19. doi: 10.1128/MMBR.00012-19
- D'andrea, M. M., Marmo, P., Henrici De Angelis, L., Palmieri, M., Ciacci, N., Di Lallo, et al. (2017). φBO1E, a newly discovered lytic bacteriophage targeting carbapenemase-producing *Klebsiella pneumoniae* of the pandemic

- Clonal Group 258 clade II lineage. *Sci. Rep.* 7:2614. doi: 10.1038/s41598-017-02788-9
- Denic, M., Touati, E., and De Reuse, H. (2020). Review: pathogenesis of *Helicobacter pylori* infection. *Helicobacter* 25(Suppl. 1):e12736. doi: 10.1111/hel. 12736
- Dos Santos, A. A., and Carvalho, A. A. (2015). Pharmacological therapy used in the elimination of *Helicobacter pylori* infection: a review. *World J. Gastroenterol.* 21, 139–154. doi: 10.3748/wjg.v21.i1.139
- Fernández, L., Gutiérrez, D., García, P., and Rodríguez, A. (2019). The perfect bacteriophage for therapeutic applications—a quick guide. *Antibiotics* 8:126. doi: 10.3390/antibiotics8030126
- Figueiredo, C., Camargo, M. C., Leite, M., Fuentes-Pananá, E. M., Rabkin, C. S., and Machado, J. C. (2017). Pathogenesis of gastric cancer: genetics and molecular classification. *Curr. Top. Microbiol. Immunol.* 400, 277–304. doi: 10.1007/978-3-319-50520-6_12
- Fischer, W., Tegtmeyer, N., Stingl, K., and Backert, S. (2020). Four chromosomal type IV secretion systems in *Helicobacter pylori*: composition, structure and function. *Front. Microbiol.* 11:1592. doi: 10.3389/fmicb.2020.01592
- Fischetti, V. A. (2018). Development of phage lysins as novel therapeutics: a historical perspective. *Viruses* 10:310. doi: 10.3390/v10060310
- Goodwin, C. S., Armstrong, J. A., and Peters, M. (1989). “Microbiology of *C. pylori*,” in *Campylobacter pylori in Gastritis and Peptic Ulcer Disease*, ed. D. Y. Graham (New York, NY: Igaku-Shoin Medical Publishers), 25–49.
- Hatfull, G. F., and Hendrix, R. W. (2011). Bacteriophages and their genomes. *Curr. Opin. Virol.* 1, 298–303. doi: 10.1016/j.coviro.2011.06.009
- Heintschel Von Heinegg, E., Nalik, H. P., and Schmid, E. N. (1993). Characterisation of a *Helicobacter pylori* phage (HP1). *J. Med. Microbiol.* 38, 245–249. doi: 10.1099/00222615-38-4-245
- Hyman, P., and Abedon, S. T. (2010). Bacteriophage host range and bacterial resistance. *Adv. Appl. Microbiol.* 70, 217–248. doi: 10.1016/S0065-2164(10)70007-1
- Ierardi, E., Losurdo, G., Fortezza, R. F., Principi, M., Barone, M., and Leo, A. D. (2019). Optimizing proton pump inhibitors in *Helicobacter pylori* treatment: old and new tricks to improve effectiveness. *World J. Gastroenterol.* 25, 5097–5104. doi: 10.3748/wjg.v25.i34.5097
- Jones, N. L., Koletzko, S., Goodman, K., Bontems, P., Cadranet, S., Casswall, T., et al. (2017). Joint ESPGHAN/NASPGHAN guidelines for the management of *Helicobacter pylori* in children and adolescents (Update 2016). *J. Pediatr. Gastroenterol. Nutr.* 64, 991–1003. doi: 10.1097/MPG.0000000000001594
- Kumar, N., Albert, M. J., Al Abkal, H., Siddique, I., and Ahmed, N. (2017). What constitutes an Arabian *Helicobacter pylori*? Lessons from comparative genomics. *Helicobacter* 22:e12323. doi: 10.1111/hel.12323
- Kyrillos, A., Arora, G., Murray, B., and Rosenwald, A. G. (2016). The presence of phage orthologous genes in *Helicobacter pylori* correlates with the presence of the virulence factors CagA and VacA. *Helicobacter* 21, 226–233. doi: 10.1111/hel.12282
- Latka, A., Maciejewska, B., Majkowska-Skrobek, G., Briers, Y., and Drulis-Kawa, Z. (2017). Bacteriophage-encoded virion-associated enzymes to overcome the carbohydrate barriers during the infection process. *Appl. Microbiol. Biotechnol.* 101, 3103–3119. doi: 10.1007/s00253-017-8224-6
- Lehours, P., Vale, F. F., Bjursell, M. K., Melefors, O., Advani, R., Glavas, S., et al. (2011). Genome sequencing reveals a phage in *Helicobacter pylori*. *mBio* 2:e00239-11. doi: 10.1128/mBio.00239-11
- Lin, D. M., Koskella, B., and Lin, H. C. (2017). Phage therapy: an alternative to antibiotics in the age of multi-drug resistance. *World J. Gastrointest. Pharmacol. Ther.* 8, 162–173. doi: 10.4292/wjgpt.v8.i3.162
- Lin, H., Paff, M. L., Molineux, I. J., and Bull, J. J. (2017). Therapeutic application of phage capsule depolymerases against K1, K5, and K30 capsulated. *Front. Microbiol.* 8:2257. doi: 10.3389/fmicb.2017.02257
- Loc-Carrillo, C., and Abedon, S. T. (2011). Pros and cons of phage therapy. *Bacteriophage* 1, 111–114. doi: 10.4161/bact.1.2.14590
- Łoś, M., and Węgrzyn, G. (2012). Pseudolysogeny. *Adv. Virus Res.* 82, 339–349. doi: 10.1016/B978-0-12-394621-8.00019-4
- Luo, C. H., Chiou, P. Y., Yang, C. Y., and Lin, N. T. (2012). Genome, integration, and transduction of a novel temperate phage of *Helicobacter pylori*. *J. Virol.* 86, 8781–8792. doi: 10.1128/JVI.00446-12
- Majkowska-Skrobek, G., Latka, A., Berisio, R., Squeglia, F., Maciejewska, B., Briers, Y., et al. (2018). Phage-borne depolymerases decrease. *Front. Microbiol.* 9:2517. doi: 10.3389/fmicb.2018.02517
- Malfetheriner, P., Megraud, F., O’morain, C. A., Gisbert, J. P., Kuipers, E. J., Axon, A. T., et al. (2017). Management of *Helicobacter pylori* infection—the Maastricht V/florence consensus report. *Gut* 66, 6–30. doi: 10.1136/gutjnl-2016-312288
- Marshall, B. J., Armstrong, J. A., Francis, G. J., Nokes, N. T., and Wee, S. H. (1987). Antibacterial action of bismuth in relation to *Campylobacter pyloridis* colonization and gastritis. *Digestion* 37(Suppl. 2), 16–30. doi: 10.1159/000199555
- Matsuzaki, S., Rashel, M., Uchiyama, J., Sakurai, S., Ujihara, T., Kuroda, M., et al. (2005). Bacteriophage therapy: a revitalized therapy against bacterial infectious diseases. *J. Infect. Chemother.* 11, 211–219. doi: 10.1007/s10156-005-0408-9
- Matta, A. J., Pazos, A. J., Bustamante-Rengifo, J. A., and Bravo, L. E. (2017). Genomic variability of *Helicobacter pylori* isolates of gastric regions from two Colombian populations. *World J. Gastroenterol.* 23, 800–809. doi: 10.3748/wjg.v23.i5.800
- Miftahussurur, M., Yamaoka, Y., and Graham, D. Y. (2017). *Helicobacter pylori* as an oncogenic pathogen, revisited. *Expert Rev. Mol. Med.* 19:e4. doi: 10.1017/erm.2017.4
- Moodley, Y., Linz, B., Bond, R. P., Nieuwoudt, M., Soodyall, H., Schlebush, C. M., et al. (2012). Age of the association between *Helicobacter pylori* and man. *PLoS Pathog.* 8:e1002693. doi: 10.1371/journal.ppat.1002693
- Mucito-Varela, E., Castillo-Rojas, G., Cevallos, M. A., Lozano, L., Merino, E., López-Leal, G., et al. (2016). Complete genome sequence of *Helicobacter pylori* strain 7C isolated from a mexican patient with chronic gastritis. *Genome Announc.* 4:e01503-15. doi: 10.1128/genomeA.01503-15
- Nijevitch, A. A., Idrisov, B., Akhmadeeva, E. N., and Graham, D. Y. (2014). Choosing optimal first-line *Helicobacter pylori* therapy: a view from a region with high rates of antibiotic resistance. *Curr. Pharm. Des.* 20, 4510–4516. doi: 10.2174/13816128113196660728
- Oktem-Okullu, S., Cekic-Kipritci, Z., Kilic, E., Seymen, N., Mansur-Ozen, N., Sezerman, U., et al. (2020). Analysis of correlation between the seven important *Helicobacter pylori* (H. pylori) virulence factors and drug resistance in patients with gastritis. *Gastroenterol. Res. Pract.* 2020:3956838. doi: 10.1155/2020/3956838
- Otero, R. W., Gómez, Z. M., Otero, P. L., and Trespalacios, R. A. (2018). *Helicobacter pylori*: how do we treat it in 2018? *Rev. Gastroenterol. Peru.* 38, 54–63.
- Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A. J., Thomson, N. R., et al. (2010). Antagonistic co-evolution accelerates molecular evolution. *Nature* 464, 275–278. doi: 10.1038/nature08798
- Polk, D. B., and Peek, R. M. (2010). *Helicobacter pylori*: gastric cancer and beyond. *Nat. Rev. Cancer* 10, 403–414. doi: 10.1038/nrc2857
- Schmid, E. N., Von Recklinghausen, G., and Ansorg, R. (1990). Bacteriophages in *Helicobacter (Campylobacter) pylori*. *J. Med. Microbiol.* 32, 101–104. doi: 10.1099/00222615-32-2-101
- Secka, O., Vale, F. F., Buissonnière, A., Thomas, J. E., Mégraud, F., and Lehours, P. (2017). Phylogeographic agreement between prophage and bacterial housekeeping genes in *Helicobacter pylori* strains from The Gambia. *Helicobacter* 22:e12394. doi: 10.1111/hel.12394
- Solovieva, E. V., Myakinina, V. P., Kislichkina, A. A., Krasilnikova, V. M., Verevkin, V. V., Mochalov, V. V., et al. (2018). Comparative genome analysis of novel Podoviruses lytic for hypermucoviscous *Klebsiella pneumoniae* of K1, K2, and K57 capsular types. *Virus Res.* 243, 10–18. doi: 10.1016/j.virusres.2017.09.026
- Soluri, M. F., Puccio, S., Caredda, G., Edomi, P., D’elios, M. M., Cianchi, F., et al. (2020). Defining the *Helicobacter pylori* disease-specific antigenic repertoire. *Front. Microbiol.* 11:1551. doi: 10.3389/fmicb.2020.01551
- Sulakvelidze, A. (2005). Phage therapy: an attractive option for dealing with antibiotic-resistant bacterial infections. *Drug Discov. Today* 10, 807–809. doi: 10.1016/S1359-6446(05)03441-0
- Takeuchi, H., Kira, M., Konishi, S., Uchiyama, J., Matsuzaki, S., and Matsumura, Y. (2018). Polymorphisms in the *Helicobacter pylori* NY43 strain and its prophage-cured derivatives. *Microbiology* 164, 877–882. doi: 10.1099/mic.0.000665

- Vale, F. F., Nunes, A., Oleastro, M., Gomes, J. P., Sampaio, D. A., Rocha, R., et al. (2017). Genomic structure and insertion sites of *Helicobacter pylori* prophages from various geographical origins. *Sci. Rep.* 7:42471. doi: 10.1038/srep42471
- Vale, F. F., Vadivelu, J., Oleastro, M., Breurec, S., Engstrand, L., Perets, T. T., et al. (2015). Dormant phages of *Helicobacter pylori* reveal distinct populations in Europe. *Sci. Rep.* 5:14333. doi: 10.1038/srep14333
- Vítor, J. M., and Vale, F. F. (2011). Alternative therapies for *Helicobacter pylori*: probiotics and phytomedicine. *FEMS Immunol. Med. Microbiol.* 63, 153–164. doi: 10.1111/j.1574-695X.2011.00865.x
- Yahara, K., Lehours, P., and Vale, F. F. (2019). Analysis of genetic recombination and the pan-genome of a highly recombinogenic bacteriophage species. *Microb. Genom.* 5:e000282. doi: 10.1099/mgen.0.000282
- You, Y., He, L., Zhang, M., and Zhang, J. (2015). Comparative genomics of a *Helicobacter pylori* isolate from a Chinese Yunnan Naxi ethnic aborigine suggests high genetic divergence and phage insertion. *PLoS One* 10:e0120659. doi: 10.1371/journal.pone.0120659

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Muñoz, Stepanian, Trespalacios and Vale. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

ANEXO 2. Draft genome sequences of 29 *Helicobacter pylori* strains isolated from Colombia.

Muñoz AB, Stepanian J, Acosta C, Solano-Gutierrez JS, Vale FF, Trespalacios- Rangel AA. 2021. Draft genome sequences of 29 *Helicobacter pylori* strains isolated from Colombia. *Microbiol Resour Announc* 10:e00218-21. <https://doi.org/10.1128/MRA.00218-21>



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

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

^aInfectious Diseases Research Group, Microbiology Department, Sciences Faculty, Pontificia Universidad Javeriana, Bogotá, Colombia

^bAXOMICS-Sequencing Center, Biological Science Department, Science School, Universidad EAFIT, Medellín, Colombia

^cHost-Pathogen Interactions Unit, Research Institute for Medicines (iMed-ULisboa), Faculdade de Farmácia, Universidade de Lisboa, Lisbon, Portugal

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% IsoVitalax (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 Trimmomatic 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.

Volume 10 Issue 19 e00218-21

Citation Muñoz AB, Stepanian J, Acosta C, Solano-Gutierrez JS, Vale FF, Trespalacios-Rangel AA. 2021. Draft genome sequences of 29 *Helicobacter pylori* strains isolated from Colombia. Microbiol Resour Announc 10:e00218-21. <https://doi.org/10.1128/MRA.00218-21>.

Editor Catherine Putonti, Loyola University Chicago

Copyright © 2021 Muñoz et al. This is an

open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Angela B. Muñoz, munoangela@javeriana.edu.co, or Alba A. Trespalacios-Rangel, alba.trespacios@javeriana.edu.co.

Received 28 February 2021 Accepted 16 April 2021 Published 13 May 2021

TABLE 1 Genome statistics of sequences reported

Strain name	GenBank accession no.	SRA accession no.	No. of CDS ^a	Genome size (bp)	GC content (%)	MLST	No. of contigs	N_{50} value (bp)	Genome coverage (x)	No. of raw reads
COL 1-PUJ	JAFCHS000000000	SRR13796410	1,582	1,679,429	38.8	HpEurope	72	58,624	135	1,163,956
COL 2-PUJ	JACSDV000000000	SRR13796434	1,632	1,607,581	39	HpEurope	46	82,542	392	2,633,600
COL 5-PUJ	JACSDU000000000	SRR13796433	1,667	1,624,361	39.1	HpEurope	133	21,560	105	850,310
COL 6-PUJ	JACSDT000000000	SRR13796422	1,673	1,625,175	39.1	HpEurope	133	20,478	95	745,268
COL 8-PUJ	JAFCHT000000000	SRR13796409	1,664	1,661,424	38.9	HpEurope	43	93,756	286	2,823,146
COL 9-PUJ	JAFCHU000000000	SRR13796408	1,653	1,613,788	39	HpEurope	47	84,065	311	2,073,864
COL 10-PUJ	JAFCHV000000000	SRR13796407	1,680	1,642,843	39.4	HpEurope	58	107,835	375	2,741,078
COL 11-PUJ	JAFCHW000000000	SRR13796406	1,663	1,626,191	39.4	HpEurope	58	86,547	221	1,639,572
COL 12-PUJ	JAFCHX000000000	SRR13796432	1,633	1,637,127	38.9	HpEurope	43	96,559	368	2,773,188
COL 13-PUJ	JAFCHY000000000	SRR13796431	1,714	1,658,899	39.3	HpEurope	58	59,276	554	4,615,312
COL 14-PUJ	JAFCHZ000000000	SRR13796430	1,701	1,673,807	38.9	HpEurope	54	80,670	256	1,758,244
COL 15-PUJ	JAFCIA000000000	SRR13796429	1,563	1,546,556	39.2	HpEurope	46	60,019	133	987,226
COL 16-PUJ	JACSDS000000000	SRR13796412	1,665	1,634,541	39.3	HpEurope	64	54,632	482	2,363,616
COL 18-PUJ	JAF CIB000000000	SRR13796428	1,662	1,658,604	38.9	HpEurope	37	93,756	580	436,440
COL 19-PUJ	JAF C IC000000000	SRR13796427	1,654	1,656,342	38.9	HpEurope	42	81,649	371	2,363,966
COL 20-PUJ	JAF C ID000000000	SRR13796426	1,674	1,619,537	39	HpEurope	56	59,593	340	2,201,658
COL 21-PUJ	JAF C IE000000000	SRR13796425	1,670	1,618,229	39	HpEurope	55	58,359	435	3,168,290
COL 23-PUJ	JACSDR000000000	SRR13796411	1,661	1,614,131	39.1	HpEurope	91	32,948	99	762,738
COL 24-PUJ	JAF C IF000000000	SRR13796424	1,674	1,622,244	39	HpEurope	119	25,019	119	922,348
COL 25-PUJ	JAF C IG000000000	SRR13796423	1,681	1,662,282	38.9	HpEurope	88	37,899	113	909,004
COL 26-PUJ	JAF C IH000000000	SRR13796421	1,631	1,619,895	39	HpEurope	52	97,025	208	1,447,706
COL 27-PUJ	JAF C II000000000	SRR13796420	1,656	1,653,551	38.9	HpEurope	36	80,500	359	2,247,256
COL 28-PUJ	JAF C IJ000000000	SRR13796419	1,671	1,653,809	38.9	HpEurope	23	132,947	479	3,095,502
COL 29-PUJ	JAF C IK000000000	SRR13796418	1,585	1,586,826	39.2	HpEurope	31	144,624	340	2,443,034
COL 30-PUJ	JAF C IL000000000	SRR13796417	1,596	1,587,263	39.1	HpEurope	35	92,323	339	2,478,576
COL 31-PUJ	JAF C IM000000000	SRR13796416	1,795	1,711,739	39	HpEurope	14	41,620	363	2,430,084
COL 49-PUJ	JAF C IN000000000	SRR13796415	1,786	1,735,252	39	HpEurope	96	59,657	255	2,107,660
COL 50-PUJ	JAF C IO000000000	SRR13796414	1,726	1,711,312	38.9	HpEurope	57	118,609	522	4,410,236
COL 51-PUJ	JAF C IP000000000	SRR13796413	1,656	1,666,802	38.9	HpEurope	39	82,225	404	2,801,844

^aCDS, 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](https://doi.org/10.6026/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-Rector, Pontificia Universidad Javeriana (PPTA_7676) and F.P.I.T. BanRepCultural (project 3956). F.F.V. 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; 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

- 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>.
- Peek RM, Blaser MJ. 2002. Helicobacter pylori and gastrointestinal tract adenocarcinomas. *Nat Rev Cancer* 2:28–37. <https://doi.org/10.1038/nrc703>.
- Camilo V, Sugiyama T, Touati E. 2017. Pathogenesis of Helicobacter pylori infection. *Helicobacter* 22:e12405. <https://doi.org/10.1111/hel.12405>.
- Bravo LE, Cortés A, Carrascal 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.
- 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>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski 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>.
- Tatusova T, DiCuccio M, Badretdin A, Chetverin 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>.
- 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.14826.1>.
- 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, Katzowitsch 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>.
- 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>.
- 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>.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- 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>.
- 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.
- 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>.
- Vale FF, Vadivelu J, Oleastro M, Breurec S, Engstrand L, Perets TT, Mégraud F, Lehours P. 2015. Dormant phages of Helicobacter pylori reveal distinct populations in Europe. *Sci Rep* 5:14333. <https://doi.org/10.1038/srep14333>.
- 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>.

ANEXO 3. An American lineage of *Helicobacter pylori* prophages found in Colombia.

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>

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

Angela B Muñoz^{1,2}  | Alba A Trespalacios-Rangel¹  | Filipa F Vale² 

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

²Host-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/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 hp-SWEurope, 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 populations¹ 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 diversity¹⁸ 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 2015⁷ 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- AAGYTTTTAGMGTTTTGYG, 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* prophages, 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.

Accession number Genbank	Strain	Blast Database	Integrase			Holin			Country	PST	MLST
			Query cov.	E-value	% Id.	Query cov.	E-value	% Id.			
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
MVUO01000000	HP15012	WGS	84%	2.00E-88	82.90%	100%	3.00E-67	86,64%	Australia	hpNEurope	hpEurope
MVYB01000000	HP04042	WGS	100%	0.00E+00	95.93%	100%	3.00E-67	86,64%	Australia	hpSWEurope	hpEurope
MVVZ01000000	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
NYJJ01000000	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 2.3.4²⁷⁻²⁹ program input file using *xmfa2struct* by X. Didelot and D. Falush (<http://www.xavie.rdidelot.xtremehost.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 Harvester³⁰ 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-parameter³¹ 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 *ureI*) 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 ISCo1, 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

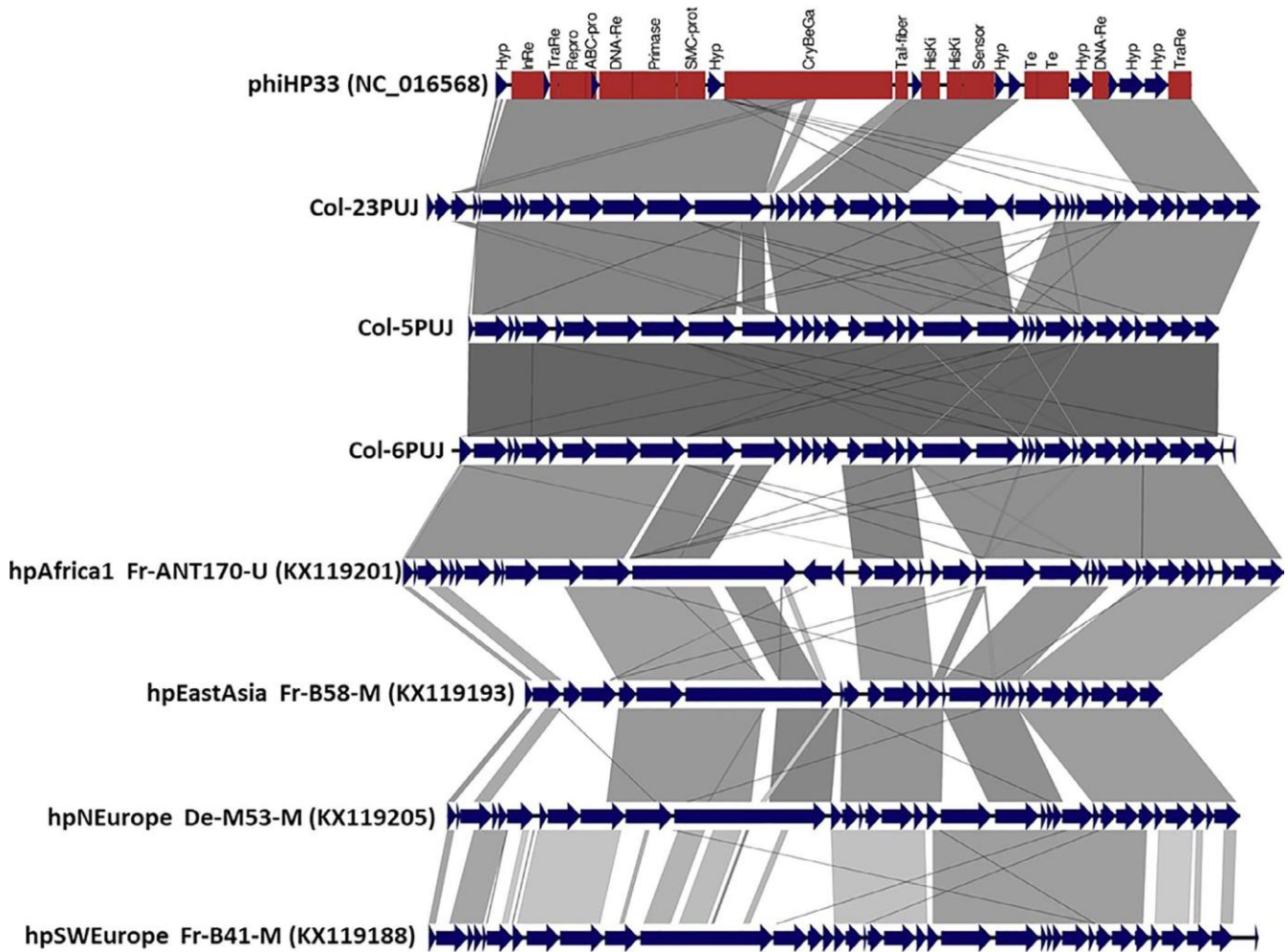


FIGURE 1 Synteny in genome organization of prophages from different populations. Using the BLASTn and Easyfig software, a phage comparison chart was constructed. The three complete Colombian prophages were used, and as a reference, phiHP33 and some phages (one for each population group) previously described 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.

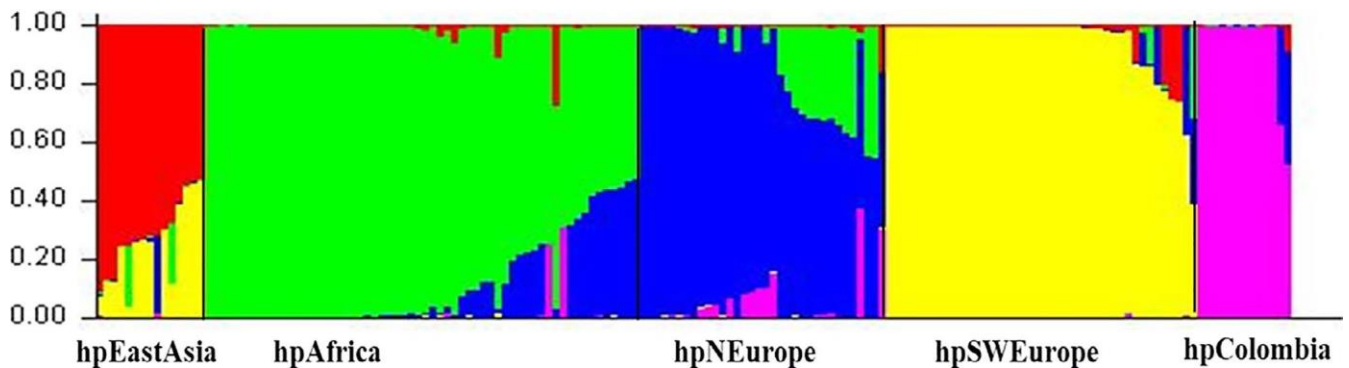


FIGURE 2 *H. pylori* prophage populations. Bayesian population assignments using Structure 2.3.4 software and the admixture model to analyze integrase and holin genes sequences. The plot shows five colored segments (K) representing the membership coefficients in each cluster. The fifth population was 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 method³⁰ 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

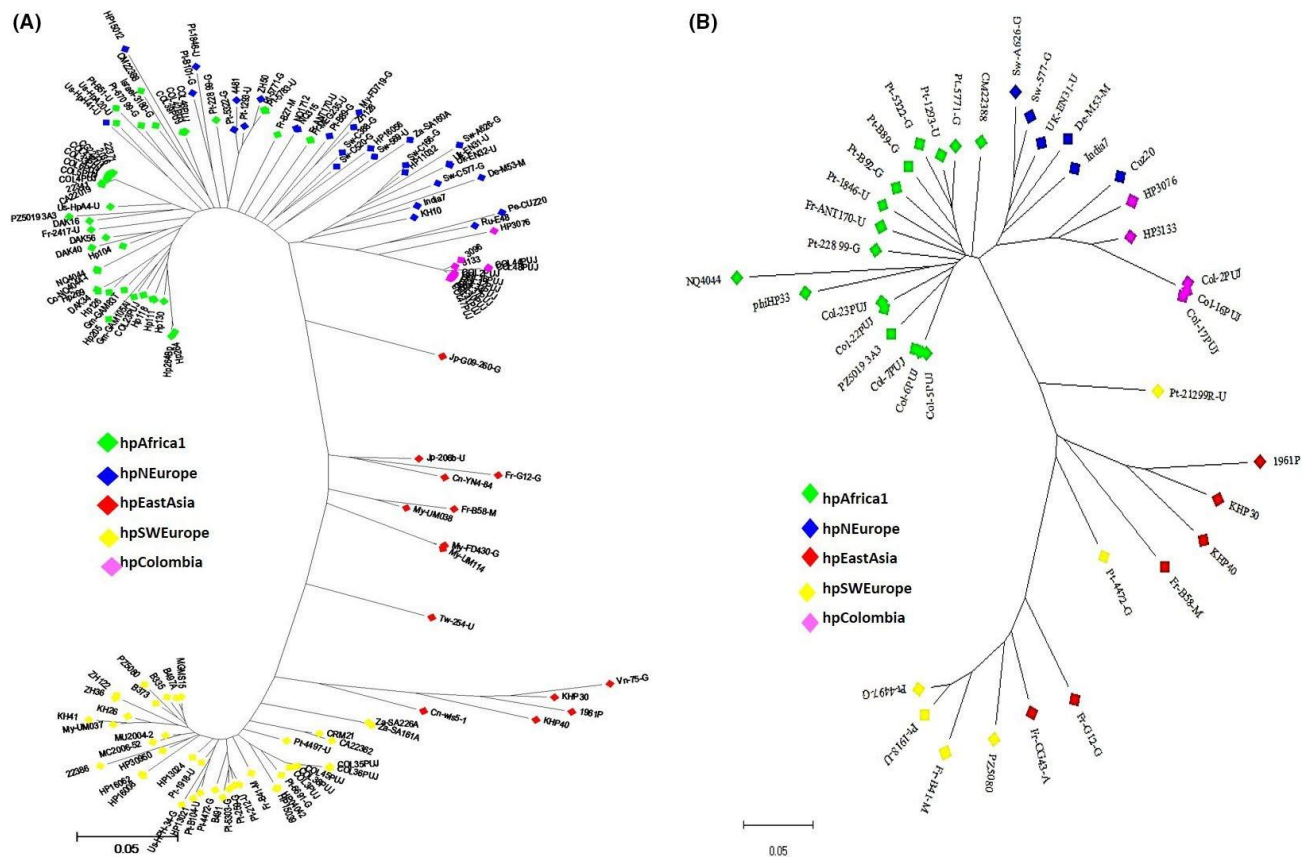


FIGURE 3 Phylogenetic distribution of the *H. pylori* prophages. A. Neighbor-joining phylogenetic tree topology (using the Kimura two-parameter model, running 1000 bootstrapping replications) of nucleotide alignments of the concatenated integrase and holin genes from 167 prophage sequences (27 Colombian 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.

5 | 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 database³⁸ 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 genes³⁷ 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 are 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) (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 (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 bacteria⁴²; finally, the prophage present in the strain PZ5019 (available from NCBI) (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 pathways¹⁹ 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.

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

- 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).
- 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.
- Paterson S, Vogwill T, Buckling A, et al. Antagonistic coevolution accelerates molecular evolution. *Nature*. 2010;464(7286):275-278.
- 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.
- Lehours P, Vale FF, Bjursell MK, et al. Genome sequencing reveals a phage in *Helicobacter pylori*. *MBio*. 2011;2(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).
- Vale FF, Vadivelu J, Oleastro M, et al. Dormant phages of *Helicobacter pylori* reveal distinct populations in Europe. *Sci Rep*. 2015;5:14333.
- 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.
- 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).
- 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.
- Uchiyama J, Takeuchi H, Kato S-I, et al. Characterization of *Helicobacter pylori* bacteriophage KHP30. *Appl Environ Microbiol*. 2013;79(10):3176-3184.
- 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).
- Muñoz AB, Stepanian J, Trespalacios AA, Vale FF. Bacteriophages of *Helicobacter pylori*. *Front. Microbiol*. 2020;11:549084.
- Mégraud F, Lehours P, Vale FF. The history of *Helicobacter pylori*: from phylogeography to paleomicrobiology. *Clin Microbiol Infect*. 2016;22(11):922-927.
- 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.
- Bravo LE, Cortés A, Carrascal E, et al. *Helicobacter pylori*: patología y prevalencia en biopsias gástricas en Colombia. *Colombia Médica*. 2003;34(3).
- Camilo V, Sugiyama T, Touati E. Pathogenesis of *Helicobacter pylori* infection. *Helicobacter*. 2017;22(Suppl):1.
- Thorell K, Lehours P, Vale FF. Genomics of *Helicobacter pylori*. *Helicobacter*. 2017;22(Suppl):1.
- 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).
- Muñoz-Ramírez 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.
- Thorell K, Yahara K, Berthenet E, et al. Rapid evolution of distinct *Helicobacter pylori* subpopulations in the Americas. *PLoS Genet*. 2017;13(2):e1006546.
- 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.
- 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.
- 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.
- Gutiérrez-Escobar AJ, Trujillo E, Acevedo O, Bravo MM. Phylogenomics of Colombian. *Gut Pathog*. 2017;9:52.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772-780.
- 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.
- 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.
- Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155(2):945-959.
- 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.
49. de Sablet T, Piazzuelo MB, Shaffer CL, et al. Phylogeographic origin of *Helicobacter pylori* is a determinant of gastric cancer risk. *Gut.* 2011;60(9):1189-1195.
50. 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>

ANEXO 4. Información de genomas colombianos de *H. pylori* recolectados de bases de datos públicas.

En esta base de datos se recolecta la información demográfica, clínica y las estadísticas de ensamblaje de los genomas de *H. pylori* colombianos disponibles en bases de datos públicas. (Consultado julio de 2020). Esta base de datos se deriva de la consulta en la base de datos Patric.

Teniendo en cuenta el tamaño de la tabla, se decidió depositar la base de datos completa en un repositorio público. Para acceder a la misma se puede consultar el siguiente enlace: <https://zenodo.org/record/4898561#.YLmb5KgzbIU>

Para citación se puede usar la siguiente referencia:

Muñoz, Angela B. (2021). Colombian *H. pylori* genomes in public databases. (Version 1.0) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.4898561>