

Filogenómica de aislamientos colombianos de *Helicobacter pylori*
procedentes de una zona de alto riesgo de cáncer gástrico

ANDRES JULIAN GUTIERREZ ESCOBAR



PONTIFICIA UNIVERSIDAD JAVERIANA

FACULTAD DE CIENCIAS

DOCTORADO EN CIENCIAS BIOLÓGICAS

Bogotá, D.C; 2018

Filogenómica de aislamientos colombianos de *Helicobacter pylori*
procedentes de una zona de alto riesgo de cáncer gástrico

Andrés Julián Gutiérrez Escobar

Tesis presentada como requisito parcial para optar al título de:
Doctor en Ciencias Biológicas

Tutor
Orlando Acevedo, Ph. D

Cotutoras:
María Mercedes Bravo MSc
Gina Méndez Callejas Ph. D

Pontificia Universidad Javeriana Bogotá
Facultad de Ciencias
Doctorado en Ciencias Biológicas
Bogotá D.C
Octubre de 2018

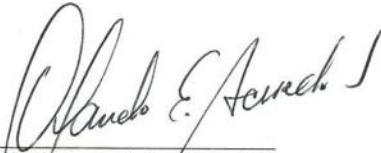
NOTA DE ADVERTENCIA

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

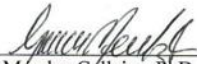
Artículo 23 de la Resolución N° 13 de julio de 1946.

Nota de aceptación:


Aprobado por el comité de Grado en cumplimiento con los requisitos exigidos por la Pontificia Universidad Javeriana para optar al título de Doctor en Ciencias Biológicas.



Orlando Acevedo Sarmiento. PhD
Tutor


María Mercedes Bravo. MSc
Cotutora

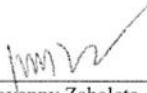

Gina Méndez Callejas. PhD.
Cotutora


Dr. William Otero Regino. MD.
Jurado


Juan Manuel Anzola. PhD.
Jurado


Álvaro Pazos. PhD.
Jurado


Andrés Felipe Aristizabal. PhD.
Jurado


Jovanny Zabaleta. PhD.
Jurado

Filogenómica de aislamientos colombianos de *Helicobacter pylori*
procedentes de una zona de alto riesgo de cáncer gástrico

ANDRES JULIAN GUTIERREZ ESCOBAR

APROBADO


Concepción Puerta Bula Ph.D.
Decana
Facultad de Ciencias




Alba Alicia Trespalacios Rangel Ph. D
Directora de posgrado
Facultad de Ciencias



AGRADECIMIENTOS

En primer lugar, quiero agradecer a Dios y al Señor Jesucristo por haber permitido los espacios, las coincidencias, todos los azares y contingencias que hicieron posible el cumplimiento de este extraño periodo de mi vida. Que esta infima contribución sea solo para tu gloria Señor.

A la Pontificia Universidad Javeriana, a la Universidad de Ciencias Aplicadas y Ambientales, al Instituto Nacional de Cancerología y a Colciencias. Por el apoyo y patrocinio de la presente tesis doctoral.

A mi tutor y cotutoras por su confianza y por dejarme trabajar con libertad creativa y sin exigencias administrativas arcaicas.

A mi padre Alvaro Gutiérrez y a mi madre Martha Lucia Escobar, a mi hermano y hermana y mi abuela Aurora Arias por haber inculcado en mi el espíritu científico. Por su amor sin condiciones.

A mi esposa Adela Lozano. Por ser la mujer mas espectacular, mas hermosa, mas brillante, mas dulce y mas linda. Gracias por tu apoyo, por lidiar con mis cambios de humor, por lidiar con mi depresión y mi ansiedad. Tu existencia es lo mas maravilloso y dulce que tengo. Amada esposa, gracias por hacer que mi tiempo aquí en esta tierra transcurra hermosamente.

A mi hijo, porque eres mi inspiración y porque me encanta ser tu Papá. Tu existencia es el marco perfecto para darle sentido completo a mi vida.

Contenido

1.	Resumen	10
2.	Planteamiento del problema	13
3.	Justificación	14
4.	Objetivos.....	15
4.1.	Objetivo general.....	15
4.2.	Objetivos específicos	15
5.	Marco teórico.....	16
5.1.	<i>Helicobacter pylori</i> : características generales	16
5.2.	Citotoxina vacuolizante VacA	18
5.3.	Adhesinas como factores de virulencia.....	19
5.3.1.	AlpA/AlpB	20
5.3.2.	HorB	21
5.4.	Tipα.....	22
5.5.	Respuesta inmune frente a la infección por <i>Helicobacter pylori</i>	23
5.6.	Patologías gástricas asociadas con la infección por <i>Helicobacter pylori</i>	26
5.6.1.	<i>Helicobacter pylori</i> : de gastritis a cáncer gástrico	28
5.6.2.	<i>Helicobacter pylori</i> y adenocarcinoma gástrico.....	29
5.6.3.	<i>Helicobacter pylori</i> y úlcera péptica	31
5.6.4.	<i>Helicobacter pylori</i> y linfoma tipo MALT	32
5.7.	Aspectos evolutivos moleculares y genéticos poblacionales de bacterias.....	33
5.8.	Aspectos topológicos de la evolución a escala de secuencia y estructura proteica.....	37
5.9.	Aspectos básicos de genética de poblaciones bacterianas	41
5.10.	Genómica evolutiva y estructura poblacional de <i>Helicobacter pylori</i>	44
6.	Metodología.....	47
6.1.	Ensamble y anotación	47
6.2.	Determinación del contenido de genes	48
6.3.	Análisis de locus múltiples (MLST).....	49
6.4.	Análisis filogenómicos: métodos basados en alineamiento de secuencias y métodos libres de alineamiento.....	49
6.4.1.	Análisis basados en alineamientos de fragmentos de secuencia.	50
6.4.2.	Análisis basados en kmeros sin alineamiento de secuencias.	50

6.5.	Análisis filogenético y poblacional para factores de virulencia	51
6.6.	Análisis de divergencia funcional y selección positiva.	52
6.7.	Predicción de la estructura proteica y generación de redes de interacción de residuos	53
7.	Resultados.....	55
7.1.	Contenido de genes	55
7.2.	Análisis de locus múltiples (MLST).....	57
7.3.	Análisis filogenómicos de aislamientos colombianos de <i>H. pylori</i>	59
7.4.	Diversidad genética de factores de virulencia	62
7.5.	Determinación de eventos de divergencia funcional y selección positiva para los factores de virulencia	74
7.6.	Carga mutacional diferencial para los factores de virulencia por grupos de patología gástrica y análisis topológico	78
8.	Discusión	88
9.	Conclusiones.....	104
10.	Referencias	106
11.	Anexos.....	137
11.1.	Anexo 1: Estadística de los genomas de <i>Helicobacter pylori</i>	137
11.2.	Anexo 2: Análisis Inmuno-Informáticos	137
11.3.	Anexo 3: Artículos científicos	137
11.4.	Productos obtenidos.....	137

Índice de Figuras

Figura 1. Sistema de amortiguamiento de <i>Helicobacter pylori</i>	18
Figura 2. Características histopatológicas de la mucosa gástrica generadas por la infección crónica de <i>Helicobacter pylori</i>	31
Figura 3. Análisis de contenido de genes..	56
Figura 4. Árbol del Pan genoma de <i>Helicobacter pylori</i>	57
Figura 5. Análisis de locus múltiples.....	58
Figura 6. Filogenómica de aislados colombianos de <i>Helicobacter pylori</i>	60
Figura 7. Árbol filogenético para los PNS conservados analizados con kSNP v3.0.....	61
Figura 8. Análisis de Ka/Ks contra posición nucleotídica.....	65
Figura 9. Reconstrucción filogenética para la adhesina AlpA de <i>Helicobacter pylori</i>	67
Figura 10. Reconstrucción filogenética para la adhesina HorB de <i>Helicobacter pylori</i>	69
Figura 11. Reconstrucción filogenética para la adhesina VacA de <i>Helicobacter pylori</i>	71
Figura 12. Reconstrucción filogenética para la adhesina Tipa de <i>Helicobacter pylori</i>	73
Figura 13. Modelo del dominio VacAp555Col usando el servidor I-TASSER.	81

Índice de Tablas

Tabla 1. Análisis de diferenciación y flujo genético de tres alelos de aislamientos colombianos de <i>H. pylori</i>	55
Tabla 2. Análisis de divergencia funcional en factores de patogenicidad de aislamientos colombianos <i>H. pylori</i>	68
Tabla 3. Análisis de selección positiva en factores de patogenicidad de aislamientos colombianos de <i>H. pylori</i>	69
Tabla 4. Determinación de residuos polimórficos significativamente diferentes entre grupos de patología gástrica mediante el uso del servidor Meta-CATS.....	70
Tabla 5. Residuos topológicos esenciales para cada parámetro topológico estudiado.....	74
Tabla 6. Comparación de los parámetros topológicos básicos entre las redes de referencia y las redes de los modelos.....	76
Tabla 7. Relación entre los sitios topológicos esenciales con tipo de selección natural y predicción inmune.....	77
Tabla 8. Utilidad de los sitios topológicos esenciales para la diferenciación de grupos de patología gástrica.....	78

1. Resumen

Helicobacter pylori es una bacteria que coloniza la mucosa gástrica del 50% de la población mundial y está asociada a la aparición de patologías como gastritis, úlceras pépticas y cáncer gástrico. En Colombia, el 80% de la población está infectada con *H. pylori* y el cáncer gástrico es la primera causa de muerte por cáncer en hombres y la tercera para mujeres. *H. pylori* tiene tasa mutacional y recombinogénica una muy alta, que le ha permitido coevolucionar a la par con los humanos desde su migración fuera de África. El estudio de su estructura poblacional ha permitido evidenciar eventos migratorios humanos tanto ancestrales como modernos. El objetivo de este trabajo fue describir las relaciones filogenómicas de 103 aislamientos colombianos de *H. pylori* obtenidos en las ciudades de Bogotá, Tunja y poblaciones aledañas, pertenecientes a la zona andina, región de alto riesgo para cáncer gástrico.

Los genomas de los 103 aislamientos se analizaron por métodos filogenómicos basados en contenido de genes, alineamientos de secuencias, métodos libres de alineamiento y por tipificación de locus múltiples, que mostraron un posible evento evolutivo en la zona de estudio, en la cual emergieron linajes independientes de la bacteria posiblemente como respuesta al haplotipo mestizo emergente en la zona. Posteriormente se hicieron análisis filogenéticos, de genética poblacional y topológicos de cuatro factores de patogenicidad: AlpA, VacA, HorB y Tip α con el fin de evaluar si estos factores, importantes para la supervivencia de la bacteria, mostraban un comportamiento evolutivo similar al observado a escala genómica y así, confirmar la presencia de linajes independientes. Los cuatro árboles filogenéticos obtenidos permitieron evidenciar la presencia de estos linajes independientes,

y los estadísticos de genética poblacional indicaron que su evolución se debió a una rápida expansión poblacional producto de un cuello de botella, en el que probablemente la selección inmune pudo tener un papel importante en la aparición de las variantes alélicas observadas. Los patrones de variación genética y microevolutivos de estos factores de patogenicidad se debieron probablemente a eventos de divergencia funcional, selección positiva, purificadora e inmune, lo que, además, dio un sustento adicional a la presencia de linajes independientes de la bacteria en la zona de estudio. Adicionalmente, se identificaron residuos polimórficos en VacA y AlpA que no habían sido reportados previamente en la literatura y que podrían tener potencial para diferenciar las patologías gástricas asociadas a la infección, es decir, probablemente exista una carga mutacional diferencial entre aislamientos obtenidos de diferentes patologías gástricas para estos factores de virulencia. Finalmente, para determinar si la diferenciación evolutiva observada en VacA y AlpA tuvo impacto a nivel de su estructura proteica, se realizó un estudio a escala topológica, a partir de las redes de interacción de residuos para ambos factores. Se identificó que la conectividad de las proteínas es un factor selectivo sinérgico que contribuye con la diferenciación poblacional observada a nivel de la secuencia génica, además, se lograron identificar residuos topológicamente esenciales que también tienen el potencial de diferenciar los grupos de patologías gástricas.

En este estudio se evidenció un proceso de diferenciación genética a escala genómica, filogenética y poblacional de aislamientos de *Helicobacter pylori* que probablemente se debió a dos procesos: i) Al desarrollo de un nuevo haplotipo humano que indujo un proceso microevolutivo y ii) A un proceso de selección natural probablemente de tipo inmune, que indujo la diversificación alélica tanto a nivel de secuencia, como a nivel de estructura proteica

de factores de patogenicidad, que dieron origen a linajes independientes de *Helicobacter pylori* en una zona de alto riesgo de cáncer gástrico en Colombia.

2. Planteamiento del problema

El cáncer gástrico ocupa el quinto lugar en incidencia y el tercero en mortalidad a nivel mundial (IARC). En Colombia este cáncer es el primero en mortalidad en hombres y el cuarto en mujeres (Sierra et al. 2016) y se considera que la infección por *Helicobacter pylori* es el principal factor de riesgo para su desarrollo (Helicobacter and Cancer Collaborative Group). Se ha determinado que, aunque la prevalencia de la infección por *Helicobacter pylori* en Colombia es alta y generalizada (Bravo et al. 2003), la incidencia de cáncer gástrico aumenta con el gradiente altitudinal, es decir, que en zonas montañosas como la región andina es alta y en la región pacífica es baja (Torres et al. 2013; Pardo et al. 2006).

La evidencia sobre el tipo de ancestría de la bacteria en la región andina colombiana es contradictoria. En un estudio con aislamientos de la bacteria procedentes de esta región, encontraron una ancestría exclusiva de tipo europeo debido al proceso de la conquista de América, lo cual ha sido considerado como un predictor fuerte del riesgo de cáncer gástrico. Sin embargo, otros estudios han encontrado una mezcla entre el tipo europeo, africano y trazas de asiático (de Sable et al., 2011; Kodaman et al., 2014). Contradictoriamente, otro estudio sugirió que la ancestría no solo era de tipo europeo, sino que las poblaciones colombianas presentaban un grupo poblacional propio (Shiota et al., 2014). Es bien conocido que Bogotá fue un centro importante para la entrada y establecimiento de europeos durante la conquista, pero es poco probable que la bacteria de tipo exclusivamente europeo no haya sufrido procesos microevolutivos locales para adaptarse a su nuevo hospedero; este tipo de eventos ya han sido reportados previamente para la bacteria (Kumar et al., 2017; Kumar et al., 2015). Sin embargo, a la fecha no se ha realizado ningún estudio que haya explorado

detalladamente la evolución del genoma de la bacteria en la zona del presente estudio y cómo la selección natural ha dado forma a sus patrones microevolutivos a escala local. Tampoco se conoce si han evolucionado variantes alélicas para factores de virulencia que estén relacionadas con las patologías gástricas características de esta zona del país.

3. Justificación

Una necesidad de la comunidad médica mundial y para la cual Colombia no es una excepción es contar con un diagnóstico molecular rápido y eficaz contra la infección por *Helicobacter pylori* (Kumar S et al., 2014). Se sabe que el proceso coevolutivo que ocurre entre *Helicobacter pylori* y hospedero es un modulador importante para el desarrollo de lesiones gástricas (Guevara et al., 2016) entonces, una clave para orientar la búsqueda de un diagnóstico molecular eficaz proviene del estudio de los procesos evolutivos a escala genómica de la bacteria.

Entender cómo ha evolucionado el genoma de la bacteria en la región andina colombiana, no solo permitirá comprender un poco más los patrones de ancestría en la zona y dar más evidencia a la comunidad internacional, sobre cómo ha sido la evolución de *Helicobacter pylori* después de la conquista de América, sino también, identificar los patrones microevolutivos que han dado origen a variantes alélicas para factores de virulencia, que puedan brindar información básica sobre la existencia de cargas mutacionales diferenciales entre aislamientos obtenidos a partir de diferentes grupos de patologías gástricas y así, establecer los pasos iniciales para el desarrollo de herramientas diagnósticas moleculares más precisas, contra la infección por *Helicobacter pylori* a escala local.

4. Objetivos

4.1. Objetivo general

Describir el proceso evolutivo a escala filogenómica de aislamientos colombianos de *Helicobacter pylori* procedentes de una zona de alto riesgo de cáncer gástrico.

4.2. Objetivos específicos

- Determinar las relaciones filogenómicas de aislamientos colombianos de *Helicobacter pylori* con respecto a sus referentes globales.
- Determinar sí las variantes alélicas codificantes para factores de virulencia de *Helicobacter pylori* se han diferenciado genéticamente en los aislamientos colombianos con respecto a sus referentes globales.
- Determinar sí los polimorfismos proteicos en factores de virulencia de *H. pylori* han generado cambios topológicos, que diferencien los aislamientos colombianos con respecto a sus referentes globales.

5. Marco teórico

5.1. *Helicobacter pylori*: características generales

Helicobacter pylori es una bacteria con forma espiral, Gram-negativa, microaerofílica, fastidiosa y que mide entre 2 a 4 μm de largo y de 0,5 a 1 μm de ancho. La bacteria posee de 2 a 6 flagelos polares que le permiten gran capacidad de movimiento. El crecimiento en cultivo es relativamente lento con un rango de 3 a 5 días a 37°C y bajo una atmósfera del 5% de oxígeno y 10% dióxido de carbono. Sus colonias son pequeñas, brillantes y traslúcidas (Kenneth y McColl., 2010). La bacteria es ureasa, catalasa y oxidasa positiva y puede catabolizar glucosa, pero no otro tipo de azúcares (Berg et al., 1997; Doig et al., 1999, Marais et al., 1999; Nedenskov., 1994). La capacidad patógena de *Helicobacter pylori* quedó demostrada en experimentos de auto-ingestión realizados por Marshall (Marshall et al., 1985; Morris y Nicholson., 1987) y con voluntarios (Morris et al., 1991) donde se comprobó que podía colonizar el estómago e inducir la inflamación de la mucosa gástrica.

El estómago humano es un ambiente extremadamente ácido y en constante movimiento y recambio de su contenido, con el fin de sobrevivir en él, *Helicobacter pylori* tiene sistemas enzimáticos que funcionan de manera sinérgica y que le ayudan a soportar la hostilidad de la mucosa gástrica. El primero de estos sistemas está constituido por la ureasa citoplasmática que corresponde al 15% del total de proteínas expresadas por la bacteria. Cuando la bacteria se encuentra en un ambiente que está debajo de pH 6.5, se activa un canal ubicado en la membrana citoplasmática denominado UreI que permite la entrada de urea del hospedero. Esta urea es entonces hidrolizada por la ureasa y produce amonio y dióxido de carbono

estableciendo un sistema amortiguador en el citoplasma bacteriano (Weeks et al. 2000). El segundo sistema enzimático importante para la supervivencia de la bacteria está regulado por la anhidrasa carbónica α . Esta enzima se localiza en el periplasma y cataliza la hidratación reversible del dióxido de carbono previamente producido por la ureasa a bicarbonato permitiendo la amortiguación del pH también en el periplasma. Tanto la ureasa como la anhidrasa carbónica están reguladas por el operón ArsRS (Pflock et al. 2005; Wen et al. 2007) que también regula otras enzimas no esenciales pero que son importantes para tolerar la acidez estomacal como, por ejemplo: la amidasa AmiE; la formamidasa AmiF; y la arginasa RocF (Pflock et al. 2006).

Para evitar ser eliminada la bacteria presenta un sistema de flagelos unipolares que la hacen ser altamente móvil y puede censar el pH y los iones de bicarbonato. De esta manera la bacteria puede migrar rápidamente desde el lumen que es muy ácido hacia la capa de moco que se ubica cerca de la superficie del epitelio gástrico que es más neutra. Este gradiente de pH a través de la mucosa es establecido mediante la secreción de iones bicarbonato las células del epitelio gástrico. Este gradiente de pH de la mucosa también tiene un efecto sobre su viscosidad, ya que a medida que se acerca a la superficie del epitelio su densidad también disminuye (Celli et al. 2009). La bacteria una vez ha penetrado la capa mucosa induce una reducción de su viscosidad al incrementar localmente el pH mediante el sistema ureasa que cataliza la transformación de urea a amonio (Celli et al. 2009). La bacteria censa la urea que produce el hospedero mediante el quimiorreceptor TlpB y así puede ubicarse en el epitelio. *Helicobacter pylori* detecta un rango Nanomolar de urea (Huang et al. 2015; Suchs et al., 2011). Esta ubicación le permite entonces descargar sus productos sobre la célula hospedera (Keilberg and Ottemann 2016) ver Figura 1.

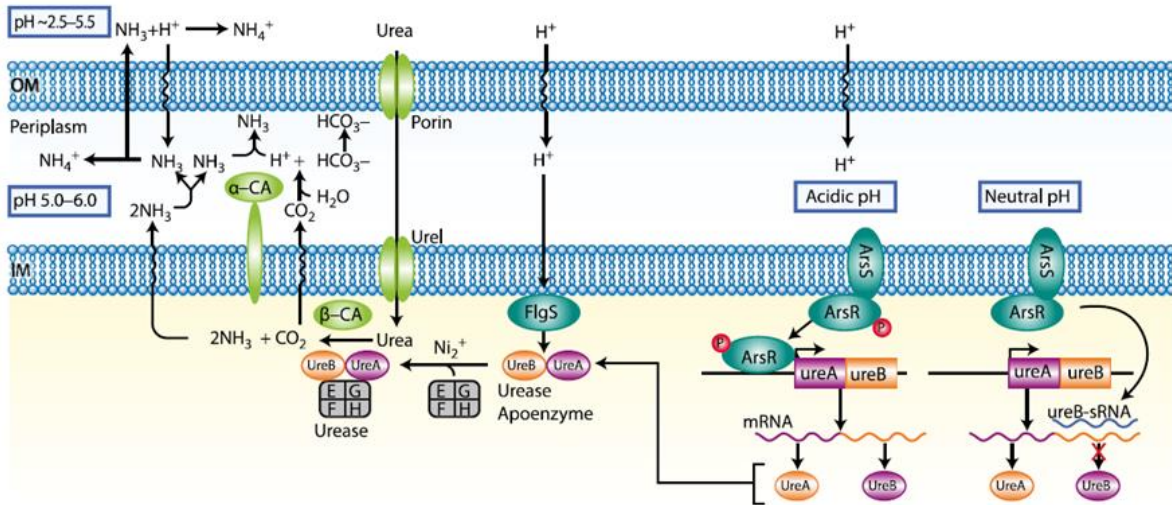


Figura 1. Sistema de amortiguamiento de *Helicobacter pylori*. Tomada y modificada de: Sachs G, David RS, Wen Y. Gastric Infection by *Helicobacter pylori*. *Curr Gastroenterol Rep* (2011) 13:540–546.

5.2. Citotoxina vacuolizante VacA

VacA es una toxina que inicialmente se describió como una proteína con capacidad de abrir poros en la célula hospedera y generar vacuolas (Foegeding et al. 2016). Posteriormente se han descrito múltiples funciones para esta toxina como, inducción de apoptosis en las células epiteliales, autofagia, inhibición de la activación de linfocitos T, todas ellas mediadas por la formación de canales selectivos dependientes de aniones (Kim and Blanke 2012; Terebiznik et al. 2009; Gebert et al. 2003).

A diferencia del gen *cagA*, el gen codificante para la toxina VacA está presente en todas las cepas de la bacteria y se caracteriza por ser altamente polimórfico. Se han descrito dos variantes alélicas para la zona de la señal peptídica (s1/s2), para la región intermedia (i1/i2), y para la región media (m1/m2). La región media es importante para la unión a la célula epitelial, la variante m1 tiene un rango de unión a diferentes tipos celulares más amplio que la variante m2. Los alelos s2 e i2 tienen una menor capacidad vacuolizante que sus

contrapartes s1 e i1 (Atherton et al. 1995; Rhead et al. 2007). Al parecer, todas las variantes de la toxina son importantes en la tolerización de las células dendríticas, en la diferenciación de linfocitos T reguladores y en la protección contra el asma en modelo de ratón (Oertli et al. 2013). Recientemente se describió que la expresión del gen *vacA* aumenta debido a un loop en la región 5' del transcrito (Amilon et al. 2015). La presencia de este loop se asocia con mayor inflamación y riesgo de patologías pre malignas más severas en la mucosa gástrica humana (Sinnott et al. 2016).

Los análisis de genotipificación de aislamientos clínicos de la bacteria han mostrado que la mayoría de las cepas que presentan la isla de patogenicidad *Cag* presentan la variante de la toxina s1/i1 mientras que las que no tienen la isla presentan usualmente la variante s2/i2. Esta asociación no se debe a la clonalidad, o a ligamiento o dependencia funcional; de hecho, las cepas en las que se ha hecho una delección del gen *cagA* continúan presentando actividad de la toxina VacA (Tummuru et al. 1994) y a su vez, en mutantes nulos para VacA, la bacteria aún puede inyectar CagA en el citoplasma de la célula hospedera (Argent et al. 2008). Lo que sí es claro es que estas proteínas se regulan entre ellas durante la infección para evitar un daño excesivo al epitelio gástrico del hospedero (Argent et al., 2008; Asahi et al. 2003; Fujikawa et al. 2003).

5.3. Adhesinas como factores de virulencia

Helicobacter pylori necesita desplazarse en el gradiente de pH hasta llegar a la mucosa gástrica y debe adherirse a ella, para esto emplea un amplio grupo de proteínas ubicadas en su membrana externa denominadas OMP (*outer membrane protein*). La adhesión es muy importante porque por un lado es un factor de virulencia de la bacteria (Magalhaes y Reis.,

2010) y por otro, porque evita que sea eliminada por el vaciamiento gástrico o el desprendimiento del moco por regeneración (Magalhaes y Reis., 2010; Odenbreit et al., 2002; Shao et al., 2005). Las adhesinas se han relacionado con la patología gástrica del hospedero (Kusters et al., 2006; Ramírez et al., 1999) y con la virulencia y adaptación de las cepas. (Odenbreit et al., 2009). Las adhesinas se han agrupado en 5 familias: i) Hop y Hor, ii) Hof, iii) Hom, iv) FecA y FrpB y v) Hef (Alm et al., 2000). Las adhesinas de *H. pylori* más estudiadas son: BabA (Ilver et al. 1998), SabA (Mahdavi et al. 2002), OipA (Yamaoka et al. 2000), HopZ (Peck et al. 1999), AlpA/AlpB (Odenbreit et al. 1999; 2002) y HorB (Snelling et al. 2007).

5.3.1. AlpA/AlpB

Las proteínas de adhesión AlpA/B son codificadas en el locus AlpAB y tienen un tamaño promedio de 518aa (Odenbreit et al. 1999, Senkovich et al., 2011; Kim et al., 2002; Jonge et al., 2004). La adhesina AlpA tiene un peso molecular de 56.1 kDa y es especie específica para *H. pylori* pues no se han encontrado homólogos ni en el género *Helicobacter* ni en otras especies bacterianas; además estas dos adhesinas son expresadas constitutivamente por todos los aislamientos clínicos estudiados hasta el momento (Odenbreit et al., 2009). Los estudios de delección del locus AlpAB han mostrado que las cepas mutantes son incapaces de adherirse al epitelio gástrico (Odenbreit et al., 2002), por lo que se ha propuesto que estas adhesinas son determinantes del tropismo de la bacteria hacia el epitelio gástrico (Odenbreit et al., 1999). El receptor de estas adhesinas es la laminina (Odenbreit., 2005) y se sabe que pueden inducir la expresión de IL-6 e IL-8 (Lu et al., 2007).

Estas adhesinas pueden alterar el epitelio gástrico puesto que pueden modular las vías de señalización pro-inflamatorias del hospedero mediante la activación de señales dependientes de quinasas, c-Fos y cAMP (Lu et al., 2007). La forma recombinante de AlpA tiene la capacidad de estimular la producción de anticuerpos en el suero de pacientes generando una respuesta humoral específica en ellos (Xue et al., 2005); es por este motivo que esta adhesina ha sido propuesta como candidato para el desarrollo de vacunas. A pesar de que todos grupos poblacionales de la bacteria presentan la adhesina de manera constitutiva, en un estudio reciente se evidenció que existe diferenciación genética entre aislamientos bacterianos de América y Europa para los alelos *alpA* y *alpB* indicando que existen procesos evolutivos más profundos que ameritan ser estudiados (Gutiérrez-Escobar, 2013).

5.3.2. HorB

Esta adhesina está codificada por el gen *horB* y tiene un tamaño de 30 kDa (Alm et al., 2000) y tienen un papel importante en la supervivencia de la bacteria puesto que mutantes en este gen que no expresan la proteína, y muestran una reducción significativa en su capacidad de adherirse al epitelio (Snelling et al. 2007). Además, se reportó que el gen se encuentra de manera persistente en aislamientos provenientes de pacientes con gastritis (Taghvaei et al., 2012). Las cepas deficientes en la proteína HorB muestran una reducción significativa en la producción de O-polisacáridos en su membrana externa, lo que ocasiona una disminución de los antígenos Le^x y Le^y y disminuye su capacidad de adhesión (Snelling et al. 2007; Edwards et al., 2000; Mahdavi et al., 2003). A pesar de su importancia funcional, HorB ha sido poco estudiada; sin embargo, algunos estudios sugieren que fuera de su papel como adhesina también podría participar para el paso de moléculas hidrofílicas a través de la membrana externa de la bacteria (Bina et al., 2000).

5.4. Tipα

El 75% de los casos de cáncer gástrico a nivel mundial son atribuibles a la infección por *H. pylori* (Marshall y Warren, 1984; IARC, 1994; Polk y Peek, 2010; Kim et al., 2011). La inflamación crónica de la mucosa gástrica debido a la infección por la bacteria induce la expresión de un perfil de citocinas proinflamatorias como: el factor de necrosis tumoral alfa (TNF- α), la interleucina 1 (IL-1) y la interleucina 6 (IL-6), y varias quimioquinas (Polk y Peek, 2010; Kim et al., 2011; Senthilkumar et al., 2011; Hannelien et al., 2012).

En modelos animales se ha demostrado que la infección con *H. pylori* induce la promoción tumoral en el estómago iniciado por varios carcinógenos (Sugiyama et al., 1998; Shimizu et al., 1999). Sin embargo, los mecanismos sobre cómo la inflamación generada por la bacteria promueve el desarrollo de cáncer en el estómago humano no se conocen y por lo tanto lo hace un blanco de extrema importancia investigativa. Los experimentos realizados con ratones deficientes de TNF- α han mostrado que esta citocina es esencial no solo para la promoción de los tumores, sino también que da origen a la secuencia de citocinas que promueve la promoción tumoral que lleva a la expresión de IL-1 y IL-6 (Suganuma et al., 1999, 2002).

La proteína inductora de TNF- α , denominada Tipα de *H. pylori* se asocia con la capacidad de la bacteria para inducir la actividad del TNF- α y se ha sugerido que juega un papel importante en la promoción del cáncer gástrico (Suganuma et al., 2005). En líneas celulares de cáncer gástrico de modelo murino MGT-40, el tratamiento con Tipα recombinante induce una fuerte expresión del gen codificante para TNF- α mediante la activación del factor de transcripción NF- κ B. También se ha mostrado que esta proteína induce la transformación de

células Bhas 42 y de líneas celulares humanas (Suganuma et al., 2005; Watanabe et al., 2014). Así, Tip- α se considera como un factor carcinogénico de la bacteria.

El gen *tipa* es también especie específico y único en el genoma de *H. pylori*. Este factor de virulencia no presenta similitud con ningún otro de la bacteria y su modo de acción es independiente de la isla de patogenicidad Cag, de la citotoxina VacA o de la ureasa (Suganuma et al., 2005, 2008, 2006). Se identificó la nucleolina como receptor para este factor de virulencia en la célula epitelial. Una vez, Tip- α se une a la nucleolina es trasladado al citoplasma y de allí al núcleo donde activa el factor de transcripción NF- κ B (Watanabe et al., 2010). La nucleolina es una proteína mayor de los nucléolos y tiene múltiples funciones como el remodelado de la cromatina, la recombinación del ADN, el procesamiento del ARNr y la estabilización del ARNm (Ginisty et al., 1999; Stork et al., 2007).

5.5. Respuesta inmune frente a la infección por *Helicobacter pylori*

La respuesta inmune contra *Helicobacter pylori* se caracteriza por presentar infiltrados de linfocitos y monocitos y un incremento significativo de la expresión de IL-1, IL-8 y IL-6 en el antro gástrico, además por la inducción de linfocitos T tipo CD4+ y CD8+ (Nurgalieva et al., 2005). También se caracteriza por la proliferación de linfocitos B y la presencia de células dendríticas en folículos linfoides (Terrés et al., 1998) y por la producción de anticuerpos tipo IgM e IgG contra la bacteria (Graham et al., 2004; Dixon et al, 2011).

La infección también induce niveles aumentados de citocinas como interferón- γ (IFN- γ), factor de necrosis tumoral α (TNF- α), IL-1, IL-6, IL-7, IL-8, IL-10, e IL-18 (Lindholm et al.,

1998). Sin embargo, no se induce la expresión de la IL-4 en la mucosa gástrica de la mayoría de los pacientes (D'Elíos et al., 1997). Por consiguiente, se concluye que la infección induce una respuesta polarizada tipo Th1, aunque también se ha asociado con la expresión de IL-17A en la mucosa gástrica (Luzza et al., 2000).

La IL-17A es el miembro más estudiado de la familia de citocinas IL-17 (IL-17A-F) y es secretada por linfocitos T CD4+ tipo Th17, así como por otros tipos de células del sistema inmune (Cua et al., 2010). Las bacterias de tipo extracelular y hongos inducen una fuerte expresión de IL-17A que estimula el estroma y las células epiteliales para secretar citocinas proinflamatorias y quimosinas tales como: TNF- α , IL-1 β , IL-6, CXCL1, CXCL2, CCL2, CCL7, CCL20, las cuales reclutan neutrófilos, macrófagos y linfocitos al sitio de la infección (Cooper et al., 2009). Sin embargo, también se ha descrito que la infección induce la producción de linfocitos T reguladores (Treg) (Lundgren et al., 2003; Raghavan et al., 2003; Harris et al., 2008) y la depleción de este linaje celular mediante inyección de anticuerpos anti-CD25 promueve el desarrollo de gastritis y una baja carga bacteriana (Rad et al., 2006).

Cuando la infección cursa en niños se induce una baja respuesta de tipo Th17 y Th1, y se produce una respuesta mediada por linfocitos T reguladores (Treg) produciendo una gastritis leve y una infección persistente (Harris et al., 2008; Serrano et al., 2013). *H. pylori* puede inhibir la proliferación de linfocitos T. Usando líneas celulares de linfocitos T se observó que la citotoxina VacA interfiere con los eventos de señalización de calcio en el interior de estas células y previene la activación de la calcineurina (Boncristiano et al., 2003; Gebert et al., 2003). La inhibición subsecuente por defosforilación del factor de transcripción NFAT resultó en la supresión de la expresión de la IL-2 y la proliferación de los linfocitos.

Un resultado similar fue observado por el factor inmunosupresivo γ -glutamyl transpeptidasa (γ -GGT) capaz de inhibir la proliferación de células T al detenerlas en fase G (1) del ciclo celular (Schmees et al., 2007). Otro factor, la arginasa de *H. pylori* puede perjudicar la función de los linfocitos T durante la infección. En modelo celular se observó que cepas de la bacteria no mutantes para la arginasa inhibieron la proliferación de los linfocitos, depletaron la L-arginina y redujeron la expresión de la cadena CD3 del receptor de la célula T (Zabaleta et al., 2004).

Cerca del 80% al 90% de las cepas de *H. pylori* presentan el antígeno de Lewis en sus lipopolisacáridos muy similares a los antígenos Lewis que se expresan sobre la superficie del estómago humano (Simoons-Smit et al., 1996). Las variantes positivas de la bacteria para este antígeno son capaces de unirse a las lectinas tipo C de en las células dentríticas gástricas (DC-SIGN) y bloquear el desarrollo de linfocitos Th1 (Bergman et al., 2004). Además de suprimir la activación de los linfocitos T, se ha demostrado que la bacteria regula la función de la inmunidad innata. Por ejemplo, la fagocitosis de la bacteria es prevenida en parte por la isla de patogenicidad (Allen et al., 2000; Ramarao et al., 2000) y aunque induce la producción de radicales de oxígeno evita su acción mediante su catalasa (Ramarao et al., 2000). Tanto la flagelinas como los LPS de la bacteria no activan adecuadamente la presentación de antígeno vía receptores tipo Toll (Gewirtz et al., 2004; Suda et al., 2001).

Asociada con la respuesta de tipo celular, la respuesta humoral se ha detectado en casi todos los pacientes infectados por *H. pylori* (Perez-Perez et al., 1988). En suero se ha detectado que anticuerpos tipo IgA e IgG en personas con infección crónica están dirigidos contra diferentes antígenos de la bacteria (Perez-Perez et al., 1988). También se ha descrito una

respuesta local contra antígenos de la bacteria en pacientes con infección crónica. Estos pacientes tienen una frecuencia muy alta de células secretoras de IgA e IgM totales mucho mayor que personas no infectadas, mientras que la frecuencia de células secretoras de IgG fue igual en los dos grupos (Mattsson et al., 1998). De manera notoria, la infección también induce la secreción de auto-anticuerpos contra células del epitelio gástrico que podría generar gastritis (Claeys et al., 1998). Estos auto-anticuerpos podrían ser citolíticos para la célula epitelial mediante la activación del complemento, induciendo apoptosis o disparando una reacción citotóxica dependiente de anticuerpos que produce la destrucción del tejido.

5.6. Patologías gástricas asociadas con la infección por *Helicobacter pylori*

Helicobacter pylori coloniza el estómago del 50% de la población mundial (Perez-Perez et al., 2004; Khalifa et al., 2010) y es el agente etiológico de la gastritis, la úlcera péptica y el cáncer gástrico (Amieva y El-Omar, 2008; Cover y Blaser 2009). Generalmente la infección por *H. pylori* es transmitida de manera intrafamiliar, se adquiere durante la niñez y se establece a lo largo de la vida del hospedero (Fujimoto et al., 2007). En la mayoría de individuos infectados, la bacteria causa gastritis asintomática, pero entre un 10 a 15 % de los casos progresa a úlcera péptica, gastritis atrófica, metaplasia intestinal, displasia y en mucho menor porcentaje a cáncer gástrico (Correa y Piazuelo., 2012).

Aunque la bacteria es el principal factor de riesgo para el desarrollo del cáncer gástrico, otros factores como el grado de inflamación y el patrón histológico que presente la gastritis inicial en el paciente también influyen en el riesgo de este cancer. La gastritis de predominio antral se asocia con el desarrollo de úlceras duodenales y la gastritis de predominio en cuerpo o

pan-gastritis se asocia con el desarrollo de úlceras gástricas y cáncer gástrico (Atherton 2006).

Se ha propuesto que una vez adquirida la infección en la niñez, si el individuo infectado no desarrolla gastritis atrófica, los altos niveles de ácido en la mucosa gástrica favorecen la gastritis antral y la aparición de úlceras duodenales en la segunda o tercera década de su vida. Por el contrario, si se desarrolla gastritis atrófica tempranamente, estos individuos con infección persistente por *H. pylori* y bajos niveles de ácido, tendrán mucho mayor riesgo de cáncer gástrico cuando sean adultos mayores (Ubukata et al. 2011).

En la década de los 70s se propuso como modelo una cascada gástrica precancerosa que describe la sucesión de cambios morfológicos en la mucosa relacionados con el desarrollo del cáncer gástrico (Correa et al., 1975). Esta secuencia empieza con una mucosa gástrica normal que se inflama produciendo una gastritis no atrófica; si esta inflamación es persistente en respuesta se generara una gastritis multifocal atrófica sin metaplasia intestinal que luego puede evolucionar a una metaplasia intestinal completa y a una incompleta, lo que posteriormente genera una displasia de bajo grado que puede cursar a una de alto grado para finalizar en un adenocarcinoma de tipo invasivo (Correa y Piazuelo, 2012).

5.6.1. *Helicobacter pylori*: de gastritis a cáncer gástrico

La superficie de la mucosa gástrica está cubierta por una capa de moco que la bacteria debe atravesar para evadir las condiciones adversas del estómago. *Helicobacter pylori* interactúa con las mucinas empleando una serie de adhesinas como por ejemplo BabA, SabA, LabA (Ilver et al. 1998; Mahdavi et al. 2002; Rossez et al. 2014). Cuando la bacteria está cerca del epitelio gástrico activa la expresión de genes proinflamatorios en la célula hospedera mediante receptores reconocedores de patrones moleculares como son: receptores tipo Toll 2 y NOD1 (Viala et al. 2004), y el inflamosoma (Kim et al. 2013; Vanaja et al. 2015). Por otro lado, y como ya se mencionó, la bacteria presenta una serie de factores de virulencia que activan las vías de señalización asociadas con la inflamación en las células del epitelio gástrico lo que da como resultado la secreción de citosinas y quimioquinas como son: IL-8, IL-1 β , TNF- α , IL-6, IL-12, CCL2-5, CCL20, y CXCL1-3 (Cook et al. 2014).

Esta secreción de quimioquinas induce el reclutamiento de neutrófilos, macrófagos, mastocitos, células detriticas, linfocitos hacia la lámina propia del epitelio gástrico produciendo su inflamación o gastritis (Atherton y Blaser 2009). A su vez, estas células producen moléculas que lesionan el tejido como son las especies reactivas de oxígeno y nitrógeno (Italiani y Boraschi 2014), y enzimas como la perforina y granzimas (Yun et al. 2005) mientras montan sus sistemas de defensa. Cuando el proceso inflamatorio del epitelio es muy prolongado se produce atrofia. Esta se define como una reducción en el número y en la forma del tejido glandular de la mucosa gástrica. De hecho, las glándulas se aplanan y se van perdiendo paulatinamente. Este cambio en el tejido posteriormente dará origen a una

gastritis atrófica multifocal cuyos focos ocurren principalmente en el antro y en el cuerpo y cuya extensión avanza con el tiempo (Correa et al., 1970).

La metaplasia intestinal se define como el cambio fenotípico de un epitelio gástrico a un epitelio de tipo intestinal. Esta lesión se considera como un estado avanzado de la gastritis atrófica ya que las glándulas metaplásicas reemplazan a las glándulas originales a medida que estas se van perdiendo. Esta lesión se clasifica como completa o intestinal y como incompleta o colónica (Matsukura et al., 1980). La de tipo incompleta generalmente se considera como una condición que predispone el desarrollo del cáncer (Filipe et al., 1994; González et al., 2010; Rokkas et al., 1991; Tava et al., 2006) y algunos la consideran como una forma suave de displasia (Tosi et al., 1993).

La displasia gástrica es el penúltimo estadio de la secuencia de la carcinogénesis gástrica y se define por la presencia de epitelio neoplásico sin evidencia de invasión. Esta lesión se caracteriza por la presencia de núcleos alargados, hipercromáticos que han perdido su ubicación basal en las células y se ven desplazados; además, la mitosis es frecuente. No se observan glándulas ordenadas, más bien son irregulares, bifurcadas o con ramificaciones. Esta lesión se clasifica como de bajo o alto grado indicando su asociación al desarrollo de cáncer gástrico (Correa y Piazuelo, 2012).

5.6.2. *Helicobacter pylori* y adenocarcinoma gástrico

El cáncer gástrico se ubica en la quinta posición en número de casos a nivel mundial con aproximadamente 100000 nuevos casos por año (Colquhoun et al. 2015). De estos la mayoría

se presenta en Asia, particularmente en la China, donde se presenta el 66% de los casos del continente. (Ferlay et al. 2015) y en Colombia es el que mas muertes produce (Sierra et al. 2016). El cáncer gástrico ocupa la tercera posición en mortalidad a nivel mundial, debido en gran medida a que su diagnóstico es tardío (Herrero et al. 2014). Este cáncer se puede dividir en dos subtipos dependientes de la localización: cardial que emerge de las células de la unión gastro-esofágica y el no cardial que emerge en la zona distal del estómago. El cáncer gástrico cardial presenta factores de riesgo similares al adenocarcinoma esofágico y al esófago de Barret y en términos generales se acepta que no están relacionados con la infección por *Helicobacter pylori* (Colquhoun et al. 2015).

Por el contrario, el cáncer gástrico no cardial está fuertemente asociado con la infección. De hecho, un 89% de los casos se le atribuyen a la bacteria. Es por este motivo que la bacteria ha sido clasificada como un carcinógeno tipo 1, el riesgo de desarrollar cáncer gástrico en un individuo infectado oscila entre el 1 y el 2% (Plummer et al. 2015). Los dos tipos histológicos de cáncer gástrico no cardial son: intestinal y difuso. El tipo intestinal se desarrolla de manera progresiva y dependiente de la inflamación. La infección por *H. pylori* conlleva a una inflamación crónica de la mucosa gástrica, después de varias décadas puede producir una atrofia de las glándulas gástricas, posteriormente a una metaplasia intestinal, displasia y finalmente el adenocarcinoma gástrico (Correa y Piazuelo, 2012). Se ha observado que la erradicación de la bacteria reduce la incidencia de gastritis atrófica, pero no reduce el riesgo de cáncer gástrico a menos que la terapia de erradicación sea antes de que ocurran los cambios pre malignos (Malfertheiner et al. 2006) ver figura 2.

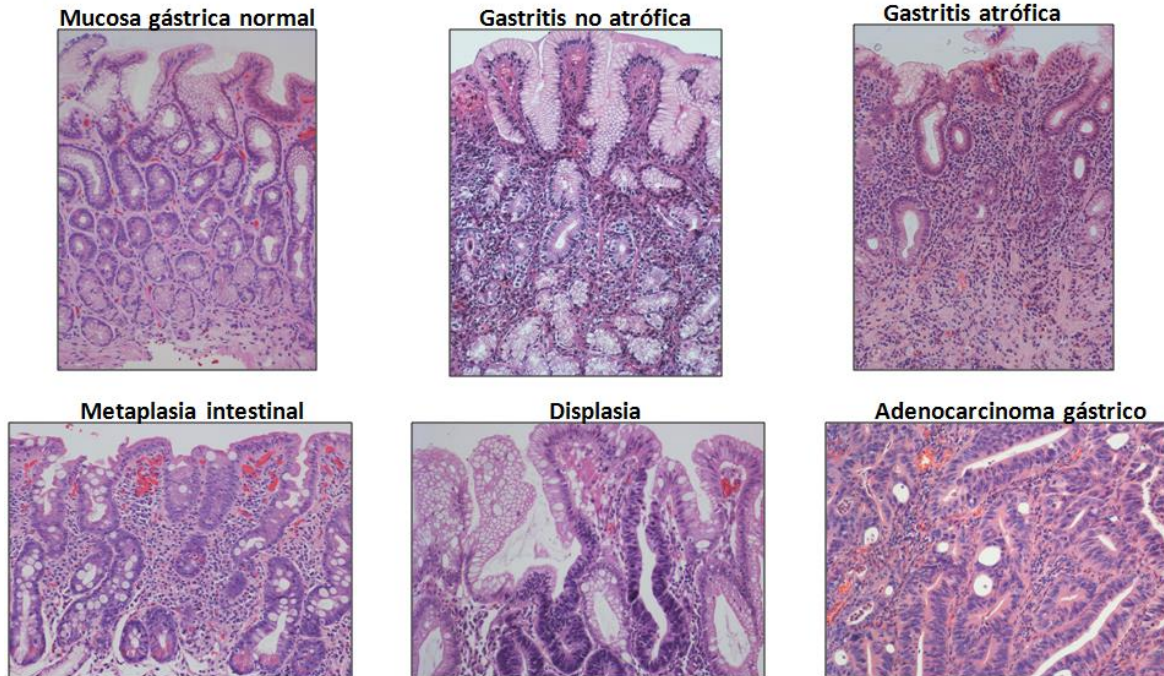


Figura 2. Características histopatológicas de la mucosa gástrica generadas por la infección crónica de *Helicobacter pylori*. **Mucosa normal:** Se observa un alto número de glándulas bien diferenciadas y sin infiltrado inmune en lámina propia. **Gastritis no atrófica:** Se observa infiltración de células inmunes (neutrófilos y leucocitos) pero el número y la forma de las glándulas se conserva. **Gastritis atrófica:** Se observa infiltrado abundante en lámina propia y una pérdida tanto en número como en forma de las glándulas. **Metaplasia intestinal:** Se observa recambio de epitelio en las glándulas estomacales con la presencia de células de tipo caliciforme y también infiltrado inmune. **Displasia:** Se observa que los núcleos de las células del epitelio gástrico están hiper cromáticos, acigarrados y con pérdida de su polaridad. **Adenocarcinoma gástrico:** No se observan tejidos diferenciados, probablemente con presencia de células en división celular y núcleos mitóticos. Tomado y modificado de: Correa P, Piazuelo MB. The gastric precancerous cascade. J Dig Dis 2012 Jan;13(1):2-9.

5.6.3. *Helicobacter pylori* y úlcera péptica

H. pylori es la principal causa de las úlceras pépticas. El 95% de las úlceras duodenales y el 70% de las úlceras gástricas se atribuyen a la infección con la bacteria (Ford et al. 2016). Este tipo de lesiones se asocia con una mortalidad significativa debido a hemorragias o perforaciones de los órganos. La inflamación crónica del antro asociada a la infección conlleva a la destrucción de las células delta y a la subsecuente reducción de la somatostatina secretada. La somatostatina inhibe la producción de gastrina de las células G vecinas lo que conduce a una hipergastrinemia. En el caso de una gastritis de tipo antral predominante, los

niveles elevados de gastrina sobre estimulan la producción de ácido por parte de las células parietales del cuerpo gástrico que no ha sido lesionado resultando en una hiperclorhidria (Robinson et al., 2017).

En algunos pacientes con este tipo de gastritis la producción elevada de ácido puede resultar en metaplasia del epitelio duodenal. Esto le permite a la bacteria colonizar y causar la inflamación y la posible ulceración del duodeno (Robinson et al., 2017). Por el contrario, las úlceras gástricas se pueden desarrollar en aquellos individuos con gastritis del cuerpo gástrico o pan-gastritis en los que la producción de ácido es normal o reducida. La baja producción de ácido debido a la pérdida de células parietales resulta en cambios atróficos de la mucosa de corpus inducida por la inflamación crónica de esta parte del estómago. A pesar del aumento en la producción de gastrina por parte del antro cuando está infectado se establece un estado de hipoclorhidria. Esto previene el desarrollo de úlceras duodenales pero debido al daño de la mucosa gástrica puede resultar en el desarrollo de úlceras gástricas. También se ha descrito que a partir de las úlceras gástricas pueden emerger lesiones pre malignas y cáncer gástrico (Atherton and Blaser 2009; Malfertheiner 2011).

5.6.4. *Helicobacter pylori* y linfoma tipo MALT

El linfoma gástrico tipo MALT se asocia a la infección por *Helicobacter pylori*. Este tipo de cáncer es poco frecuente (aproximadamente 0.8 por 100,000 pacientes por año). Un 10% de los casos son independientes de la infección, aunque quizás esto se deba a infecciones no diagnosticadas o a infecciones por otros miembros del género *Helicobacter*. La inflamación mediada por la bacteria induce la formación de folículos linfoides en la mucosa gástrica. (Genta et al. 1993). La expansión no contralada de las zonas marginales de linfocitos B en

estos folículos ocurre como consecuencia de la inflamación crónica y la estimulación antigénica continua (Du y Atherton 2006). Las células tumorales por lo general se diseminan por la mucosa gástrica y en un 40% de los casos ocurre una migración a nodos linfáticos regionales y sitios distales de la mucosa. Los linfomas tipo MALT de bajo grado pueden transformarse en linfomas de células B grandes que son más agresivos. Esto ocurre aproximadamente en la mitad de los casos y presentan un pronóstico desfavorable (Du y Atherton 2006).

5.7. Aspectos evolutivos moleculares y genéticos poblacionales de bacterias

Las variantes fenotípicas y genotípicas observadas en los distintos organismos tienen un ancestro común; además estas variaciones heredadas son mantenidas o eliminadas por acción de la selección natural. La selección natural opera bajo ciertas circunstancias ambientales en las que se induce la reproducción diferencial de unos organismos con una mayor adaptación relativa sobre otros generando un cambio en las frecuencias de los fenotipos/genotipos (Futuyma, 1986), este tipo de selección (selección positiva) induce la fijación de mutaciones benéficas según las características del entorno. Este postulado evolutivo dio pie a la generación de dos escuelas de pensamiento, en la primera se intentó dar un recuento histórico evolutivo de los organismos y en la segunda, se buscó identificar las fuerzas selectivas responsables de la variación adaptativa observada. El papel que juega la selección natural en la determinación de las características fenotípicas en bacterias esta bien documentado (Zhang, 2002). Sin embargo, el papel de la misma a nivel molecular es un tema de continuo debate.

Con el auge de las técnicas en biología molecular que permitieron el acceso a la información genética se descubrió que las proteínas y los genes presentan una variación tan alta que es bastante complejo explicarla totalmente por acción de la selección natural, lo que dio origen a la teoría neutral de la evolución molecular (Kimura, 1968, 1983). Esta teoría postula que la mayor parte de la variación observada a nivel molecular es mantenida por efectos de la deriva génica y mutaciones neutrales o casi neutrales y no solo por la selección positiva (Kimura, 1983; Ohta, 1992). Es decir, la mayor parte de la variación a nivel molecular no tiene una función adaptativa.

La discusión contemporánea sobre el tema de la selección natural a nivel molecular postula la existencia de dos tipos: i- La selección positiva que puede ser balanceadora y adaptativa de tipo darwiniano, y ii- La selección negativa o purificadora (Li, 1997). La clave para entender la selección positiva es entender que esta induce un incremento en la frecuencia de mutantes con cambios benéficos o adaptativos en las siguientes generaciones (Li, 1997). Mientras que en la selección purificadora las mutantes con cambios similares o menores al promedio son eliminadas gradualmente en las siguientes generaciones (Li, 1997). Mientras que la primera promueve la diversidad genética en las poblaciones, la segunda purga o purifica la población al eliminar algunas de esas variantes.

Con la teoría neutral o casi neutral se pueden establecer los patrones de cambio a nivel de secuencia producto de la deriva génica y la mutación sin acción de la selección natural, es decir, que cuando los postulados de la teoría no se cumplen es por acción de la selección natural. La teoría indica que el ritmo de sustitución de un nucleótido es igual a su tasa mutacional y no depende del tamaño poblacional, es decir, que el número de mutaciones se

encuentra en balance y por ende la probabilidad de su fijación en una población grande es baja. Además, predice como debe ser el polimorfismo en secuencias de la misma especie. La probabilidad de que dos secuencias elegidas al azar difieran en una posición para una población haploide es $\theta = 2N_e\mu$ donde θ es la diversidad genética, N_e es el tamaño poblacional y μ es la tasa mutacional. Indicando que mientras más grande sea la población o más alta sea la tasa mutacional mayor será la probabilidad de que las dos secuencias sean diferentes (Kimura, 1968, 1983). Entonces, indirectamente cualquier desviación de los postulados neutrales indicará la acción de la selección natural ya sea positiva o purificadora.

Para estudiar la adaptación a nivel molecular se puede optar por dos aproximaciones diferentes: i- Estudiar la distribución del polimorfismo en las secuencias de ADN y/o ii- Estudiando las sustituciones sinónimas (sin) y no sinónimas (no-sin) en las secuencias asumiendo al codón como unidad evolutiva mínima. Con el fin de entender los métodos basados en polimorfismos se necesita primero claridad en algunos conceptos básicos. El primero es la diversidad genética que se puede describir por el indicador pi (π) o el número de nucleótidos diferentes por sitio entre dos secuencias tomadas al azar (Nei y Li, 1997) y por el indicador teta (θ) que se obtiene indirectamente mediante el cálculo del número de sitios segregativos o que difieren entre el número de secuencias $\theta = K/a$ (K número de sitios segregantes y a número de secuencias), pi es afectada por los alelos de mayor frecuencia y es independiente del tamaño de la muestra, mientras teta si es afectada por el tamaño de la muestra y por los alelos poco frecuentes (Tajima, 1983, 1989).

Con estos dos estimadores se puede evaluar si un grupo de secuencias se ajusta al modelo neutral o si están desviadas. Por ejemplo, si existe selección positiva aumentará la frecuencia

alélica y por tanto se incrementará π ; pero si existe un aumento de alelos deletéreos se incrementará θ (Watterson, 1975; Tajima 1983, 1989). Entonces, al determinar las diferencias entre ambos parámetros se puede identificar si la selección es positiva o purificadora. Una de las pruebas más usadas para entender si la selección actúa sobre una muestra de secuencias, es la prueba Tajima (D) que evalúa precisamente la diferencia entre π y θ a partir de la siguiente expresión $D = \frac{\pi - \theta}{\sqrt{(\pi - \theta)}}$; cuando D es negativa, θ es mayor que π , lo que indica la presencia de mutaciones deletéreas. Por el contrario, si D es positivo θ es menor que π indica que algunos alelos se pueden encontrar bajo selección positiva incrementando sus frecuencias. Si los dos estimadores son iguales es porque la población se encuentra en equilibrio neutral (Tajima, 1983, 1989).

Existen otros métodos que cumplen el mismo propósito como por ejemplo el Fu – Li en el que consideran la distribución de las mutaciones en una genealogía de muestras en una población dada. Su lógica indica que las mutaciones más antiguas estarán con mayor frecuencia en las ramas más antiguas (internas) del árbol y aquellas más recientes por ende se ubicarán en las ramas más nuevas (externas). Si se observa acumulación de mutaciones en las ramas externas se podría pensar en un evento de selección purificadora, pero si se observan mutaciones en las ramas internas se podría deducir un evento de selección positiva (Fu y Li, 1993). Sin embargo, se considera que la prueba Tajima es más efectiva (Simonsen et al., 1995).

En los métodos basados en sustituciones el codón es la unidad mínima evolutiva. Cada codón se corresponde con un aminoácido y en cada codón se encuentran tres posiciones

nucleotídicas afectadas por tasas mutacionales diferentes por lo que ni la mutación ni la selección son uniformes en los codones. Un cambio en la primera y en la tercera posición se considera por lo general como sinónimo y no cambia el aminoácido, pero un cambio en la segunda posición generará un cambio en la codificación y por lo tanto esta posición es un punto de acción de la selección natural.

Existen numerosos métodos para la estimación de las tasas de sustitución, entre ellos los evolutivos de los que el de Nei-Gojobori y su modificación son los más ampliamente usados. En este método las sustituciones sinónimas y no sinónimas se determinan calculando el número observado de sustituciones sin y no-sin y el número de sitios potenciales con sustituciones sin y no-sin. Los sitios potencialmente sin y no-sin son descritos a partir de cada codón asumiendo igual probabilidad de cambio para todos los nucleótidos (Nei y Kumar, 2000). En el método modificado se tienen en cuenta las transversiones y las transiciones para hacer el cálculo de las sustituciones.

5.8. Aspectos topológicos de la evolución a escala de secuencia y estructura proteica

Los análisis de tasas de sustitución pueden ser realizados directamente a nivel de la secuencia de la proteína y así determinar el tipo de selección operante (positiva y/o purificadora) sitio por sitio de la secuencia y así establecer un razonamiento evolutivo para entender el ritmo de cambio (Kosakovsky y Frost, 2005). Este ritmo de cambio evolutivo a escala proteica es el objetivo de estudio de varias disciplinas como la evolución molecular, la genómica comparativa y la proteómica estructural (Pal et al., 2006).

El ritmo de cambio evolutivo sitio específico a nivel de secuencia proteica es considerado como una herramienta poderosa para la identificación de sitios que evolucionan bajo un tipo de selección particular y son funcionalmente importantes (Sauna y Kimchi Sarfaty, 2011). En muchos estudios se han encontrado correlaciones entre el ritmo de cambio evolutivo en proteínas con algunas de sus propiedades como por ejemplo, el número de residuos totales o la longitud (Ingvarsson, 2007), como se distribuyen en la célula (Wall et al., 2005), el número de proteínas que pueden interactuar (Fraser et al., 2002), su función y el nivel de expresión genética (Pagan et al., 2012; Wall et al., 2005), inclusive la posición de la proteína en una ruta de señalización celular (Ramsay et al., 2009).

También se ha descrito que la topología y las propiedades geométricas de las proteínas, la densidad de contactos realizados por residuo en la estructura, la estructura secundaria o la exposición a solventes juegan un papel importante en la evolución proteica (England y Shakhnovich, 2003; Bloom et al., 2006; Bustamante et al., 2000; Marsh and Griffiths, 2005; Rorick and Wagner, 2011; Zhou et al., 2008). Las proteínas están conformadas por uno o más polímeros lineales que se pliegan en diversidad de conformaciones espaciales cruciales para su función. Por ende, la topología y la simetría son aspectos importantes en su organización estructural. A pesar de su alta complejidad y dinámica la estructura proteica adopta topologías discretas que pueden ser clasificadas como globulares, cilíndricas o toroidales.

Una ventaja de que la estructura proteica sea discreta en términos topológicos es que puede ser tratada como un grafo desde la teoría de redes, ya que los residuos pueden ser considerados como nodos y las interacciones entre ellos como conexiones entre dichos nodos.

Las redes de interacción de residuos (RIR) pueden ser empleadas entonces para la caracterización estructural de las proteínas y para describir su proceso evolutivo (Bode et al., 2007; Krishnan et al., 2008; Sun y He, 2011; Greene, 2012; Taylor, 2013). La teoría de grafos permite enlazar los procesos evolutivos moleculares con la estructura tridimensional de las proteínas de una forma simple.

El comportamiento de los parámetros topológicos depende directamente de la forma en como ocurren las interacciones de los nodos y las redes de interacción de residuos no son la excepción. Se procederá a dar una explicación sobre los parámetros usados para su caracterización topológica. En un grafo del tipo $G = (N, E)$ donde N es el grupo de nodos (residuos) y E es el grupo de interacciones (enlaces) sus parámetros topológicos básicos serían:

- El grado de un nodo (k) es la medida topológica más simple y representa el número de conexiones que tiene dicho nodo. El valor k de un nodo i se representa así: $k_i = \sum_j a_{ij}$ mientras más alto sea este valor para un nodo mayor será el número de conexiones del mismo.
- El coeficiente de agrupamiento es una medida de que tan cohesiva es la red localmente (C_i). El C_i para un nodo i representa la proporción entre el número total de conexiones que lo conecta con sus vecinos inmediatos y el número total de conexiones posibles entre los vecinos inmediatos del nodo i . Este parámetro se define así: $C_i = \frac{2e_i}{k_i(k_i-1)}$ donde e_i es el número total de conexiones que presenta un nodo i con sus vecinos más cercanos y k_i es el número de vecinos del nodo i . En una red de interacción de residuos significa que entre mayor el número de conexiones que tiene

un residuo con sus vecinos, más alta conectividad de la región en la red en la red y mejor estabilidad local a la proteína.

- La intermediación (B_i) es una medida del número de rutas posibles que cruzan a través del nodo de interés. Para un nodo m , este parámetro se computa como la suma del número de rutas entre pares de nodos (i, j) que pasan a través de m dividido por el número total de de rutas entre los nodos. Este parámetro topológico se expresa así:

$$B(m) = \sum_{i \neq m \neq j} \frac{d_{ij}(m)}{d_{ij}}$$

donde d_{ij} es el número total de rutas que van del nodo i al nodo j y $d_{ij}(m)$ el número de rutas que pasan a través de m . La intermediación es normalizada al dividir por el número de pares de nodos que no incluyen a m , es decir, $\frac{(n-1)(n-2)}{2}$ donde n es el número de nodos en la red. Este parámetro indica que tanto

control tiene un nodo sobre el flujo de información en una red. En una red de interacción de residuos señala cuáles nodos son los que forman y fijan la estructura terciaria y por tanto gobiernan la topología de la red.

- La cercanía (Cli) mide la centralidad de un nodo basándose en que tan cercano está de otros nodos en la red y también es una medida de cuántos pasos se requieren para acceder un nodo a partir de otro. La distancia entre dos nodos se define como la ruta más corta que los conecta. Tradicionalmente, este parámetro topológico para un nodo i , se expresa así: $Cli = \frac{g-1}{\sum_{j=1}^g d_{ij}}$ donde d_{ij} es la distancia más corta entre los nodos i -

j y g es el número total de nodos en la red. Este parámetro sirve para identificar residuos de importancia funcional en las proteínas y residuos evolutivamente conservados. Aquellos residuos que presentan un valor alto para este parámetro topológico interactúan directamente o por medio de muy pocos pasos con todos los demás residuos de la red de interacción de residuos (Dhriti y Sudip, 2012).

5.9. Aspectos básicos de genética de poblaciones bacterianas

Los cambios a nivel de secuencia también se ven afectados por el flujo de genes y la estructura poblacional. En bacterias, el flujo genético es diferente al de los eucariotas y la estructura poblacional está representada más por líneas clonales que por organismos con reproducción sexual. Un proceso esencial en la genética de poblaciones bacterianas es la transferencia lateral de información genética mediante dos procesos: la recombinación homóloga y la recombinación sitio específica en la que los fragmentos de ADN son sustituidos por otros (Snyder y Champness, 1997; Gogarten et al., 2002). Estos dos procesos tienen consecuencias importantes tanto en la estructura poblacional como en la microevolución en bacterias. Las poblaciones bacterianas se reproducen por fisión binaria en la que de una célula madre surgen dos células hijas. La diversidad en las poblaciones se debe en parte al surgimiento de mutaciones que son transferidas a las células hijas que a su vez acumulan mutaciones y así durante generaciones sucesivas hasta el punto de generar linajes independientes (Spratt y Maiden, 1999; Gogarten *et al.*, 2002).

Las bacterias presentan tres mecanismos de transferencia lateral del material genético: conjugación, transformación y transducción. Estos procesos son diferentes a los sistemas de reproducción sexual presentes en los eucariotas ya que se presentan en una sola dirección, es decir de donador a receptor, se transfieren fragmentos de secuencias que pueden ser desde algunos cientos de pares de nucleótidos hasta miles, y no están vinculados con ningún proceso reproductivo. En términos generales se considera que una frecuencia de recombinación de 10^{-5} por gen por generación es muy alta (Cohan, 1994a; Maynard Smith, 1995; Gogarten et al., 2002).

En poblaciones bacterianas se ha encontrado que una mayor tasa de recombinación está asociada a menor clonalidad de la población. Los procariotas presentan un amplio espectro de estructuras poblacionales, unas altamente clonales como *Salmonella enterica*, *Pseudomonas syringae* y *Borrelia burgdorferi* y otras no clonales como *Neisseria gonorrhoeae*, *N. meningitidis* y *Helicobacter pylori*. Actualmente se acepta que la mayoría de poblaciones presenta elementos clonales y no-clonales (Dykhuizen et al., 1992; Souza et al., 1992; Maynard Smith et al., 1993; Gordon et al., 1995; Holmes et al., 1999; Silva et al., 1999; Spratt y Maiden, 1999; Maggi Solcà et al., 2001; Sakar y Guttman, 2004; Vinuesa et al., 2005c).

En una población estrictamente clonal, los linajes originados por efecto de la mutación serían independientes, estarían influidos por efectos selectivos esporádicos y la filogenia de sus genes presentaría la misma topología del linaje haciendo más pertinente su estudio desde el cladismo que desde la genética de poblaciones (Maynard Smith, 1995; Levin y Bergstrom, 2000). Por el contrario, en poblaciones no clonales se observaría una alta diversidad genética, muy poco ligamiento génico y aunque los supuestos de la genética poblacional podrían ponerse a prueba, los análisis filogenéticos de los genes no reconstruirían la historia evolutiva de los linajes (Maynard Smith, 1995; Gogarten et al., 2002).

Explicar cómo la variación genética es generada por mutaciones puntuales y recombinación, y es guiada por la selección y la deriva génica, es un gran reto para la genética poblacional y la evolución molecular bacteriana. La enorme diversidad encontrada en bacterias a nivel de su genoma y entre especies hace esta tarea bastante compleja (Li, 1997; Nei y Kumar, 2000; Lawrence, 2002). Los estudios evolutivos y genéticos en bacterias, se han enfocado en los

patrones de variación en genes específicos inducidos por la recombinación, con el fin de encontrar eventos de recombinación intragénica y ligamientos en desequilibrio entre alelos ubicados en diferentes regiones del genoma bacteriano. Estos fenómenos a su vez se ven reflejados en las reconstrucciones filogenéticas de genes particulares al presentar topologías propias y algunas veces bastante complejas (Maynard Smith, 1995; Spratt y Maiden, 1999; Feil y Spratt, 2001).

Otro proceso esencial para la evolución molecular bacteriana es la capacidad que tienen las especies de intercambiar fragmentos génicos. Los linajes bacterianos comparten información de esta manera independientemente de las distancias evolutivas implicadas (Cohan, 1994a; Feil y Spratt, 2001). En diversos genomas bacterianos es posible encontrar regiones mosaico, producto de eventos de transferencia lateral de fragmentos génicos; *Helicobacter pylori* no es la excepción (Nelson et al., 1999; Galibert et al., 2001; Perna et al., 2001; Wood et al., 2001; Welch et al., 2002). Ejemplos de este tipo de eventos son la transferencia de plásmidos, y las islas de patogenicidad (Sullivan y Ronson, 1998; Wood *et al.*, 2001; Welch *et al.*).

Una ventaja evolutiva de las bacterias sobre los eucariotas es que gracias a la transferencia de genes inducen procesos adaptativos rápidos, que pueden generar nuevos linajes adaptados a nuevos nichos y nuevas condiciones ecológicas a través de pocos eventos evolutivos (Ward, 1998; Lawrence, 1999, 2000; Gogarten *et al.*, 2002). Sin embargo, es claro que las bacterias forman especies con delimitaciones fenotípicas y genéticas discretas y conforman linajes bien definidos (Cohan, 1994b; Maynard Smith, 1995; Wernegreen y Riley, 1999).

5.10. Genómica evolutiva y estructura poblacional de *Helicobacter pylori*

En 1977 se reportó por primera vez la secuencia de un genoma de *Helicobacter pylori* a partir de un aislamiento obtenido de un paciente con gastritis crónica. Esta cepa se denominó 26695 y hoy se considera la cepa de referencia (Tomb et al., 1997). El genoma de referencia presenta un tamaño de 1.67 Mb, un 38.9% de G+C, 1445 proteínas y 1555 genes, 7 ARNr, 36 ARNt y 65 pseudogenes (www.Pubmed.com). A la fecha, en el Genbank se encuentran 878 genomas de *H. pylori* con una calidad de anotación de secuencia variable. De estos, 101 genomas están completos, de 28 está el cromosoma completo, 138 de ellos están a nivel de scaffolds y 611 a nivel de contigs. A partir de toda la información en genbank se puede inferir que en promedio el tamaño del genoma de la bacteria es de 1.63 Mb, tiene 1452 proteínas, 1621 genes y un 38.9% de G+C. Solo en 52 proyectos se han reportado secuencias de plásmidos. A partir del análisis de 75 genomas completos se determinó que la especie tiene 1173 familias proteicas conservadas y 673 dominios proteicos conservados (Cao et al., 2016).

H. pylori ha coevolucionado con los humanos desde su migración fuera de África hace unos 60.000 años (Linz et al., 2007). Las poblaciones de la bacteria se han distribuido globalmente siguiendo patrones similares al de los humanos, tanto de migración como de establecimiento en diversos territorios. Se han determinado siete poblaciones ancestrales para *H. pylori* según su origen geográfico: hpEurope, hpNEAfrica, hpAfrica1, hpAfrica2, hpAsia2, hpSahul y hpEastAsia. Esta última población se ha subdividido en las subpoblaciones hspEAsia, hspMaori y hspAmerind (Yamaoka, 2009; Falush et al., 2003; Achtman et al., 1999; Moodley et al., 2003; Devi et al., 2007). La subpoblación hspAmerind descende de cepas asiáticas

que llegaron a América a través del estrecho de Bering hace 12.000 años (Kersulyte et al., 2010).

Un evento reciente de migración humana fue la colonización de América por parte de la corona española hace 500 años. Durante este evento, migraron al continente americano tanto conquistadores europeos como esclavos africanos con su respectiva carga de nuevos patógenos. Esto produjo por un lado la desaparición de casi el 80% de la población humana nativa gracias a la nueva carga de enfermedad y por otro el desplazamiento de las cepas HspAmerind de *H. pylori* por cepas tipo HpEurope (Biachine y Ruso, 1992; Parrish et al., 2008; de ST et al., 2011; Kodaman et al., 2014; Shiota et al., 2014; Dominguez-Bello et al., 2008; Suzuki et al., 2012). En Colombia se ha demostrado que las cepas que infectan a la población mestiza son en su mayoría de tipo hpEurope y en menor proporción cepas tipo hpAfrica1 detectadas, estas últimas, principalmente en población afroamericana en la costa pacífica del país (de Sablet et al., 2011; Kodaman et al., 2014, Shiota et al., 2014). Recientemente, se ha reportado que en Colombia y otros países de la región como México y Nicaragua, existen líneas evolutivas independientes de la bacteria (Shiota et al., 2014; Thorell et al., 2017) sugiriendo eventos locales de evolución rápida.

Helicobacter pylori presenta una muy alta tasa mutacional y recombinogénica (Suerbaum, 2005), lo que facilita que sufra procesos de evolución rápida tanto a nivel del genoma como a nivel de genes especie específicos (Furuta et al., 2015; Cao et al., 2015). Por ejemplo, se ha observado que la topología de las reconstrucciones filogenéticas de algunas proteínas de membrana externa como HomB (Oleastro et al., 2009), SabB y Omp27 (Talarico et al., 2012) o para la toxina VacA (Gangwer et al., 2010) difiere de las observadas en análisis multilocus

o en análisis filogenómicos, por lo que el estudio de estos genes puede contribuir a revelar información importante de eventos evolutivos locales en aislamientos obtenidos en la población latinoamericana.

6. Metodología

Los genomas analizados en el presente trabajo fueron secuenciados a partir de 115 aislamientos de *H. pylori* obtenidos entre 1998 y 2007, de pacientes procedentes de las ciudades de Bogotá, Tunja y pueblos aledaños que pertenecen al cepario del Instituto Nacional de Cancerología, en Bogotá. Los aislamientos fueron obtenidos de pacientes con diferentes tipos de patología gástrica: 30 casos de gastritis (G), 20 casos de adenocarcinoma gástrico (GA), 28 casos de gastritis atrófica (AG), 30 casos de metaplasia intestinal, 5 casos de gastritis concomitante con úlcera duodenal (G-DU) y 2 de metaplasia intestinal concomitante con úlcera duodenal (IM-DU). Todos los procedimientos microbiológicos y moleculares se llevaron previamente a cabo en las instalaciones del Instituto Nacional de Cancerología por parte del Grupo de Investigación en Biología del Cáncer. Todos los reportes histopatológicos de los pacientes reposan en los archivos del mismo grupo de investigación.

6.1. Ensamble y anotación

A partir de las lecturas crudas se realizó el ensamble de los genomas mediante el flujo de trabajo denominado A5 (Coil et al., 2015). Este flujo de trabajo fue desarrollado con el objetivo de ensamblar genomas pequeños, de organismos haploides y cuyas lecturas se hubieran generado con tecnología illumina. El pipeline A5 integra todos los pasos necesarios para el ensamble de genomas. Brevemente, primero las lecturas crudas fueron sometidas a un proceso de remoción de adaptadores, regiones de baja calidad y corrección de errores mediante la aplicación del algoritmo SGA (Simpson y Durbin, 2012). Posteriormente, las lecturas corregidas fueron sometidas a un primer ensamblaje de contigs mediante el programa IDBA-UD (Peng et al., 2012). Posteriormente, los contigs fueron empleados en la

construcción de scaffolds con el programa SSPACE (Boetzer et al., 2011). Los scaffolds que no pasaron los criterios de calidad por defecto fueron sometidos a un reensamblaje. Finalmente, los scaffolds definitivos fueron anotados usando la tecnología RAST (Overbeek et al., 2014) y la herramienta de anotación para procariotas del NCBI. De los 115 genomas, 12 presentaron baja calidad de secuencia por tanto fueron excluidos para los análisis posteriores. Sin embargo, la información correspondiente a los marcos de lectura con calidad de secuencia suficiente fue tomada en cuenta en análisis posteriores.

6.2. Determinación del contenido de genes

El contenido de genes de los genomas ensamblados se determinó usando la herramienta GET_HOMOLOGUES (Contreras-Moreira y Vinuesa, 2013). Tres esquemas de comparación fueron empleados para el análisis: OrthoMCL (OMCL), COGtriangles (COG) (Kristensen et al., 2010) y bidirectional best hit (BDBH). Los genomas fueron agrupados de la siguiente manera: i- 36 genomas de referencia extraídos de la base de datos NCBI y ii- 139 genomas representando el total. Solo se escogieron genomas de referencia que estuvieran anotados como completos y los únicos “drafts” empleados fueron los generados para los genomas de aislamientos colombianos. Seguidamente, se definieron cuatro compartimientos para el contenido de genes: i - el core estricto, que representa el número de genes conservados por todos los genomas; ii- el core suave, que representa a los genes conservados en por lo menos el 95% de los genomas; iii- shell, representa los genes conservados en un porcentaje de similaridad menor al 95%; y vi- cloud, que representa los genes que son aislamiento o cepa específicos (Koonin y Wolf, 2008; Kaas et al., 2012; Snipen et al., 2009). Se determinó el número de genes por compartimiento y se definió el core y el pangenoma (Tettelin et al., 2005).

6.3. Análisis de locus múltiples (MLST)

Los genomas de los aislamientos colombianos de la bacteria fueron anotados en la base de datos PUBMLST (<http://pubmlst.org/helicobacter/>) y los fragmentos concatenados de los genes *atpA*, *efp*, *mutY*, *ppa*, *trpC*, *ureI*, y *yphC* comúnmente usados para el análisis de locus múltiples para las 103 secuencias fueron guardados en archivo de texto. Además, se descargaron 137 fragmentos concatenados para cepas de referencia distribuidas poblacionalmente así: hpEurope: 82; hpAfrica1: 16; hpWestAfrica: 23; hspSouthIndia: 2; hpEastAsia: 8; and hspAmerind: 7. En total, 240 secuencias fueron usadas para realizar el análisis de locus múltiples. Las 240 secuencias fueron alineadas en Muscle (Edgar, 2004) y la reconstrucción filogenética se llevó a cabo usando el modelo T92+G+I (Modelo Tamura con función gamma y sitios invariantes). La robustez estadística se determinó mediante 1000 iteraciones bootstrap usando el programa Mega V6 (Tamura et al., 2013). El árbol fue editado en la herramienta iTOL (Letunic y Bork, 2016).

6.4. Análisis filogenómicos: métodos basados en alineamiento de secuencias y métodos libres de alineamiento.

Con el fin de disminuir redundancia y optimizar los tiempos de cómputo, se excluyeron dos genomas de referencia de los análisis. Se usaron 103 genomas correspondientes a los aislamientos colombianos más 36 cepas de referencia de *H. pylori* extraídas de la base de datos del NCBI. Para la reconstrucción filogenómica se emplearon dos estrategias basadas en alineamiento de secuencias y una estrategia basada en un método libre de alineamiento.

6.4.1. Análisis basados en alineamientos de fragmentos de secuencia.

Con el fin de optimizar los tiempos de cómputo el algoritmo de Gegenees permite crear una base de datos de fragmentos cuyo tamaño es definido por el usuario y luego realizar un alineamiento de dichos fragmentos todos contra todos usando Blastn (Agren et al., 2012; Kuman et al., 2012). Para el análisis se determinaron condiciones de alta astringencia, el tamaño del fragmento fue de 200 pb elegidos cada 100 pb y se eligió un promedio de similitud del 90% para generar una matriz de similaridad.

6.4.2. Análisis basados en kmeros sin alineamiento de secuencias.

La segunda estrategia fue detectar los PSN (polimorfismos simples de nucleótidos) del core genómico y después realizar una reconstrucción filogenética con esta información usando el programa KSNP v3.0 (Gardner et al., 2015), esta suite ya ha sido empleada anteriormente para análisis filogenómicos de *H. pylori* (Van Vliet et al., 2015). Los PNS son presentados en forma de secuencias cortas o *k*meros que se van agrupando en clústeres a medida que se va realizando la comparación de secuencias entre ellos. Un paso importante, es determinar el tamaño de este *k*mero porque de él depende la calidad y la cobertura de los genomas que serán usados para la reconstrucción filogenómica. La suite KSNP v3.0 presenta un plugin denominado K-chooser el cual fue empleado para determinar el valor por *k*mero necesario para la reconstrucción. El valor promedio para el tamaño del *k*mero encontrado fue de 21 y así se obtuvieron 141 PNS para el core genómico. Cada análisis generó una matriz de similaridad entre otros formatos. Estas matrices, fueron analizadas en Splitstree4 (Huson, 1998) y visualizadas en iTOL (Letunic y Bork, 2016).

6.5. Análisis filogenético y poblacional para factores de virulencia

Se realizó un análisis evolutivo y poblacional para factores de virulencia con el objetivo de identificar si estos se han diferenciado en población colombiana y explorar la asociación de polimorfismos proteicos con el diagnóstico patológico. Los genes a estudiar fueron seleccionados bajo los siguientes criterios: 1- presentes en todos los aislamientos de la bacteria (por este motivo el factor de virulencia mayor CagA no fue incluido en el análisis), 2- Ser reconocidos como factores de virulencia, y 3- ser específicos de la especie.

Los genes *alpA*, *vacA* y su dominio *vacA/p55*, *horB* y *tip-a* fueron seleccionados para el estudio de eventos locales de evolución. Las 137 secuencias de nucleótidos y de proteínas para cada factor de virulencia fueron alineadas con el programa Muscle (Edgar, 2004) y se guardaron en archivos independientes. Los archivos de los alineamientos fueron llevados al programa MEGA v7.0 (Tamura et al., 2013) donde se determinaron los modelos evolutivos y se realizaron las reconstrucciones filogenéticas con 1000 repeticiones bootstrap usando el algoritmo NJ (Saitou y Nei, 1987). Adicionalmente, las desviaciones del modelo neutral se midieron mediante la prueba D de Tajima (Tajima, 1989) y la prueba de selección basada en codones Z usando MEGA v6.0 (Tamura et al., 2013). Seguidamente, los genes se agruparon en dos poblaciones: una de los aislamientos colombianos y otra de las cepas de referencia. Con estos archivos se determinó la diversidad genética y se obtuvieron los siguientes estadísticos poblacionales: número de haplotipos, diversidad haplotípica (Hd), diversidad nucleotídica (Pi) y el promedio de las diferencias nucleotídicas (k). También se aplicaron pruebas de heterogeneidad genética y de flujo genético poblacional mediante los métodos

Snn, GammaST y FsT, y se determinaron los eventos de recombinación y conversión génica usando el programa DnaSP 5.10 (Librado y Rozas, 2009).

6.6. Análisis de divergencia funcional y selección positiva.

El programa DIVERGE v3 (Gu et al., 2013) fue empleado para determinar eventos de divergencia funcional de tipo I con el fin de detectar cambios funcionales a nivel de proteínas asociados con procesos de divergencia evolutiva (Gu, 1999). Los eventos de selección positiva fueron evaluados como los ritmos de cambios de sustituciones sinónimas y no sinónimas y se aplicó la prueba FEL usando el servidor datamonkey (Kosakovsky Pond y Frost, 2005). Solo los resultados con $p < 0.05$ fueron considerados como estadísticamente significativos. Con la intención de predecir el potencial inmunológico de los cuatro factores de patogenicidad estudiados se realizaron dos predicciones inmunoinformáticas empleando las secuencias de proteínas completas. La primera se realizó con el servidor <http://tools.iedb.org/bcell/> (Jens et al., 2006) para predecir epítomos lineales para linfocitos B y el segundo usando la base de datos SYFPEITHI para predecir epítomos presentados específicamente por el complejo mayor de histocompatibilidad tipo II a linfocitos CD4⁺ (Schuler et al., 2007).

Para determinar la existencia de polimorfismos proteicos o carga mutacional significativamente diferentes entre los grupos de patología gástrica, se tomaron las secuencias de proteínas de los aislamientos colombianos y se agruparon según el diagnóstico clínico de origen de la siguiente manera: G gastritis, GA gastritis atrófica, MI metaplasia intestinal y AG adenocarcinoma gástrico. Seguidamente, las secuencias fueron alineadas en

Muscle (Edgar, 2004) y posteriormente, se cruzó el dato de la variación polimórfica de cada residuo ubicado en cada columna del alineamiento con el dato de patología y se realizó un conteo de cambios. Para determinar si el polimorfismo proteico observado se diferenciaba por grupos de patología se realizó una prueba chi-cuadrado y una prueba chi-cuadrado de Pearson para obtener un valor p usando el servidor Meta-CATS (Pickett et al., 2013).

6.7. Predicción de la estructura proteica y generación de redes de interacción de residuos

El servidor I-TASSER fue empleado para predecir las estructuras de las proteínas AlpA, VacA/p55, HorB y Tip- α de aislamientos colombianos (Yang et al., 2015). Las siguientes estructuras nativas fueron descargadas desde el banco de datos de proteínas PDB: 2QV3 (Kelly et al., 2007) (VacA/p55) y 3GIO (Jang et al., 2009) (Tip- α), y usadas como moldes específicos, mientras que para HorB y para AlpA se partió desde la secuencia de proteínas de la cepa de referencia 26695. La desviación del valor del cuadrado medio (RMSD), la superimposición y el análisis de la superficie de las estructuras proteicas fueron analizados usando Chimera v1.11.2 (Pettersen et al., 2004).

Se uso el programa Cytoscape (Shannon et al., 2003) para la reconstrucción de las redes de interacción de residuos. Brevemente, desde Cytoscape se invocó el programa Chimera V1.2 (Pettersen et al., 2004) y mediante el plugin StructureViz se realizó la transformación de las coordenadas de los archivos PDB correspondientes a las estructuras control y a los modelos predichos a un grafo indirecto. En este grafo los aminoácidos α correspondieron con los nodos y los contactos entre ellos como las conexiones. El criterio para establecer un contacto significativo entre nodos fue una distancia de máximo 5.0 Å ya que este valor se aproxima

al límite superior de las fuerzas atractivas de London–van der Waals (Greene y Higman, 2003).

Cada red de interacción de residuos fue analizada usando el plugin NetworkAnalyzer para determinar los parámetros topológicos: k , C_i , C_{li} , B_i , Strees, coeficiente topológico y conectividad (Nadezhda et al., 2012). Los parámetros topológicos asociados al flujo de información y flexibilidad de la red fueron obtenidos mediante el plugin RINalyzer (Nadezhda et al., 2012). Las redes fueron comparadas mediante la prueba no paramétrica de Mann Whitney para determinar si su distribución era diferente usando la suite R (R Core Team, 2013).

7. Resultados

7.1. Contenido de genes

En el estudio se incluyeron 115 aislamientos de *H. pylori* provenientes de 30 pacientes colombianos con diagnóstico de gastritis (G), 28 pacientes con gastritis atrófica (GA), 30 pacientes con metaplasia intestinal (MI), 20 pacientes con adenocarcinoma gástrico (AG), 5 pacientes con úlcera duodenal y gastritis (G-DU), 2 pacientes con úlcera duodenal y metaplasia intestinal (IM-UD), procedentes de una zona de alto riesgo para cáncer gástrico conformada por Bogotá, Tunja y poblaciones aledañas, ciudades colombianas ubicadas en la cordillera de los Andes. De estos 12 no se incluyeron debido a baja calidad de secuencia.

Los 103 genomas de los aislamientos colombianos de *H. pylori* mostraron en promedio 85 contigs, una cobertura de 127x, un tamaño de 1.65MB, 39% de G+C y 1647 genes (ver anexo 1). A partir del análisis realizado en los 136 genomas, se obtuvieron 490 clústeres con el método BDBH, 7920 con OMCL, y 8845 COG. El consenso de los tres métodos presentó 416 proteínas esenciales (Figura 3A) que son consideradas parte del core universal en bacterias (Charlebois y Doolittle, 2004). A partir de los métodos COG y OMCL se construyó una matriz de similaridad que permitió determinar el pangenoma y el core genoma de la bacteria. A partir de esta matriz se usaron los métodos propuestos por Charlebois and Doolittle (Charlebois y Doolittle, 2004) y Tettelin (Tettelin et al, 2005) para determinar el contenido de genes. Se encontraron 537 genes en el core estricto, 943 en el core suave, 1347 para el shell y 4734 para el cloud (Figura 3 B, C).

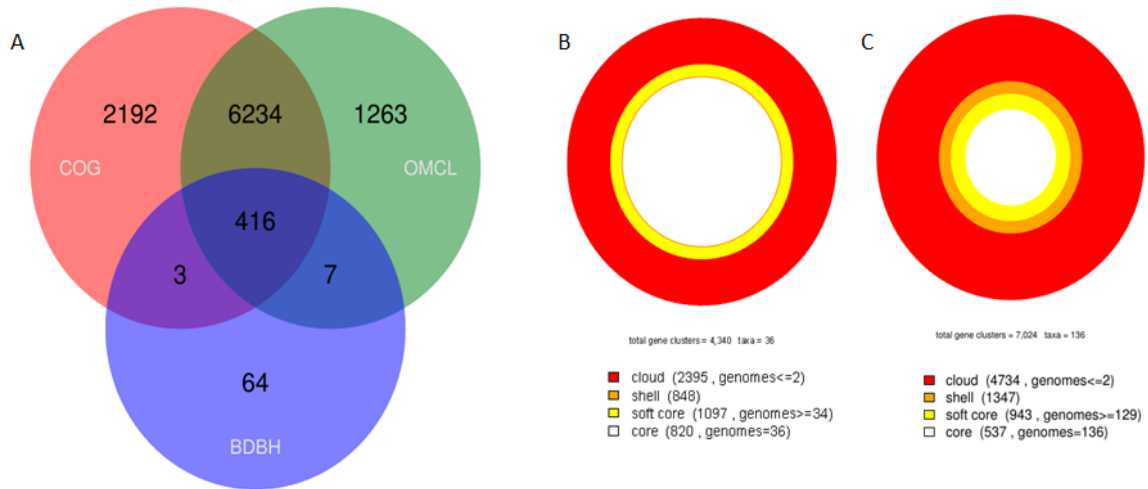


Figura 3. Análisis de contenido de genes. A) Se observa un total de 416 proteínas comunes para todos los genomas con los tres algoritmos empleados. B) Distribución de categorías de genes para los genomas de referencia y para toda la población.

El árbol pangenómico derivado de este análisis mostró una distribución geográfica acorde con lo reportado para *Helicobacter pylori* (Achtman et al., 1999). Sin embargo, también se observaron clados específicos conformados exclusivamente por aislamientos colombianos, indicando la existencia de posibles líneas evolutivas independientes para *Helicobacter pylori* en esta zona del país (Figura 4).

observó un clado exclusivamente conformado por aislamientos colombianos (líneas punteadas); también se observó que ningún aislamiento formó clados con los grupos asiáticos o amerindios. Este resultado es concordante con el análisis del contenido de genes, ver Figura 5.

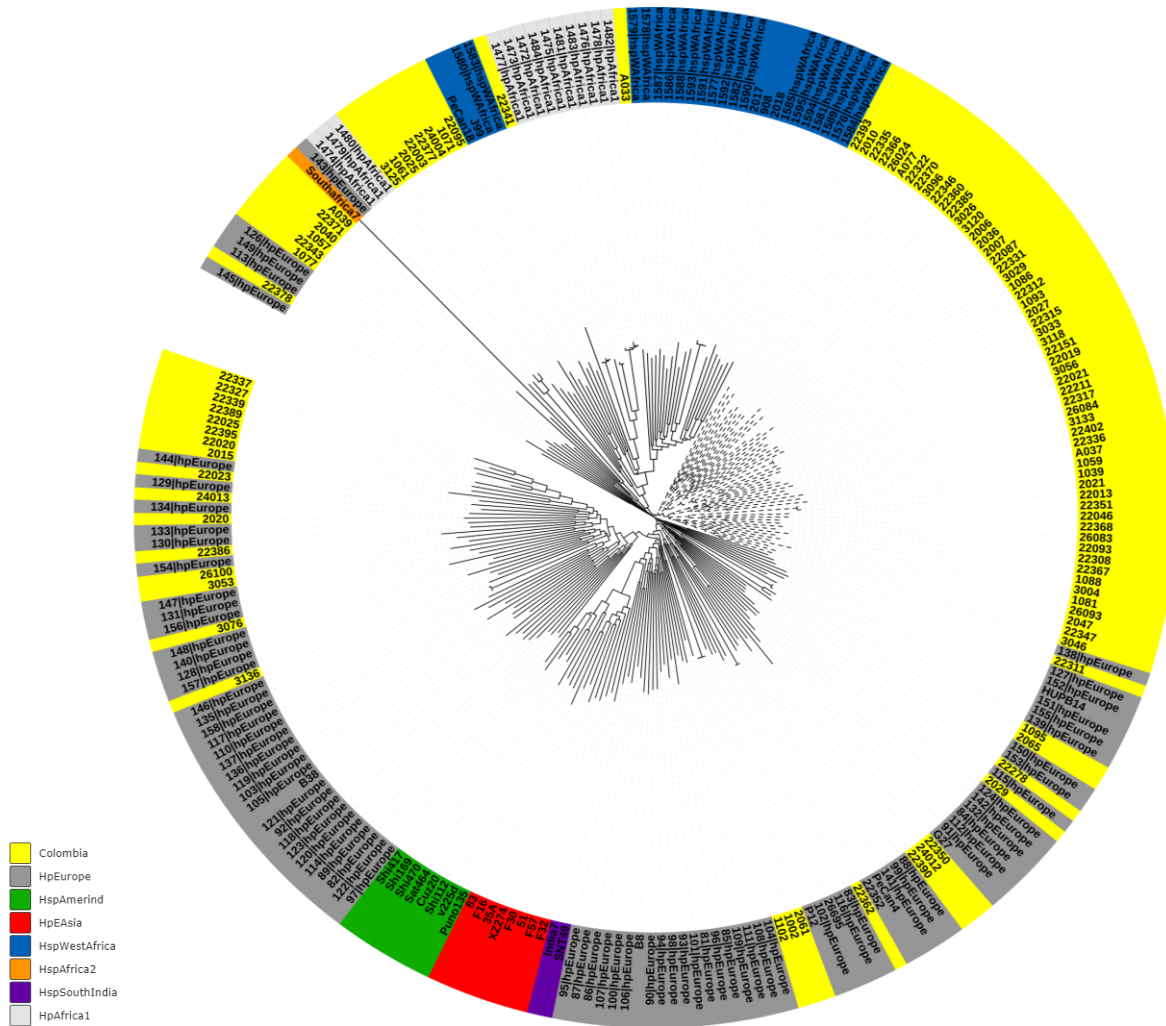


Figura 5. Análisis de locus múltiples. Los genomas de los aislamientos colombianos fueron anotados usando <http://pubmlst.org/helicobacter/>. Los archivos con las secuencias concatenadas con los fragmentos de los genes *atpA*, *efp*, *mutY*, *ppa*, *trpC*, *ureI*, y *yphC* de 103 aislamientos colombianos y 163 secuencias de referencias fueron descargados de la base de datos PubMLST. El número de secuencias y los referentes poblacionales fueron: hpEurope: 82; hpAfrica1: 16; hpWestAfrica: 23; hspSouthIndia: 2; hpEastAsia: 8; y hspAmerind: 7. Las secuencias concatenadas para todas las muestras fueron alineadas en Muscle. La reconstrucción filogenética fue realizada empleando el modelo T92+G+I (Modelo de Tamura con función gamma y sitios invariables) y la robustez estadística se logró mediante 1000 repeticiones Bootstrap usando Mega V6. El árbol resultante fue editado en iTOL v3. Las líneas punteadas son los clados conformados exclusivamente por aislamientos colombianos.

7.3. Análisis filogenómicos de aislamientos colombianos de *H. pylori*

Con el fin de contrastar los resultados obtenidos en el análisis de contenido de genes y en el MLST, se emplearon dos técnicas computacionales de última generación y de alta resolución para la reconstrucción filogenómica; una basada en alineamiento de secuencias y la otra libre de alineamiento abarcando la totalidad del genoma, es decir, una matriz de 225639 genes conteniendo todos los genes de cada genoma.

Aunque el análisis filogenómico basado en alineamiento de secuencias fue computacionalmente intensivo, reveló una estructura filogeográfica concordante con la previamente reportada para *H. pylori* (Achtman et al., 1999). Cuando los genomas de aislamientos colombianos fueron incorporados, en concordancia con el MLST anteriormente mostrado, ninguno se agrupó con las poblaciones hspAmerind o hpAsia y, un grupo de 15 genomas de aislamientos colombianos conformó un clado bien diferenciado con los genomas de la subpoblación HspWestAfrica. En este clado los genomas colombianos se ubicaron más cerca al origen que las cepas africanas de referencia, lo que podría indicar una asimilación de las cepas tipo africano en la zona de estudio.

Seguidamente, se observó que un grupo de 19 genomas colombianos conformó un clado con una cepa de la población HpEurope. Sin embargo, la mayoría de genomas de la población HpEurope formó un clado definido y sin presencia de genomas colombianos. Algunos aislamientos colombianos y HpEurope se ubicaron dispersos en el árbol, lo que podría señalar el papel de la recombinación y la conversión génica en la evolución de los genomas. Aproximadamente el 50% de los genomas de aislamientos colombianos se agruparon de

manera exclusiva en cinco clados independientes que podrían estar reflejando un proceso microevolutivo rápido. (Figura 6).

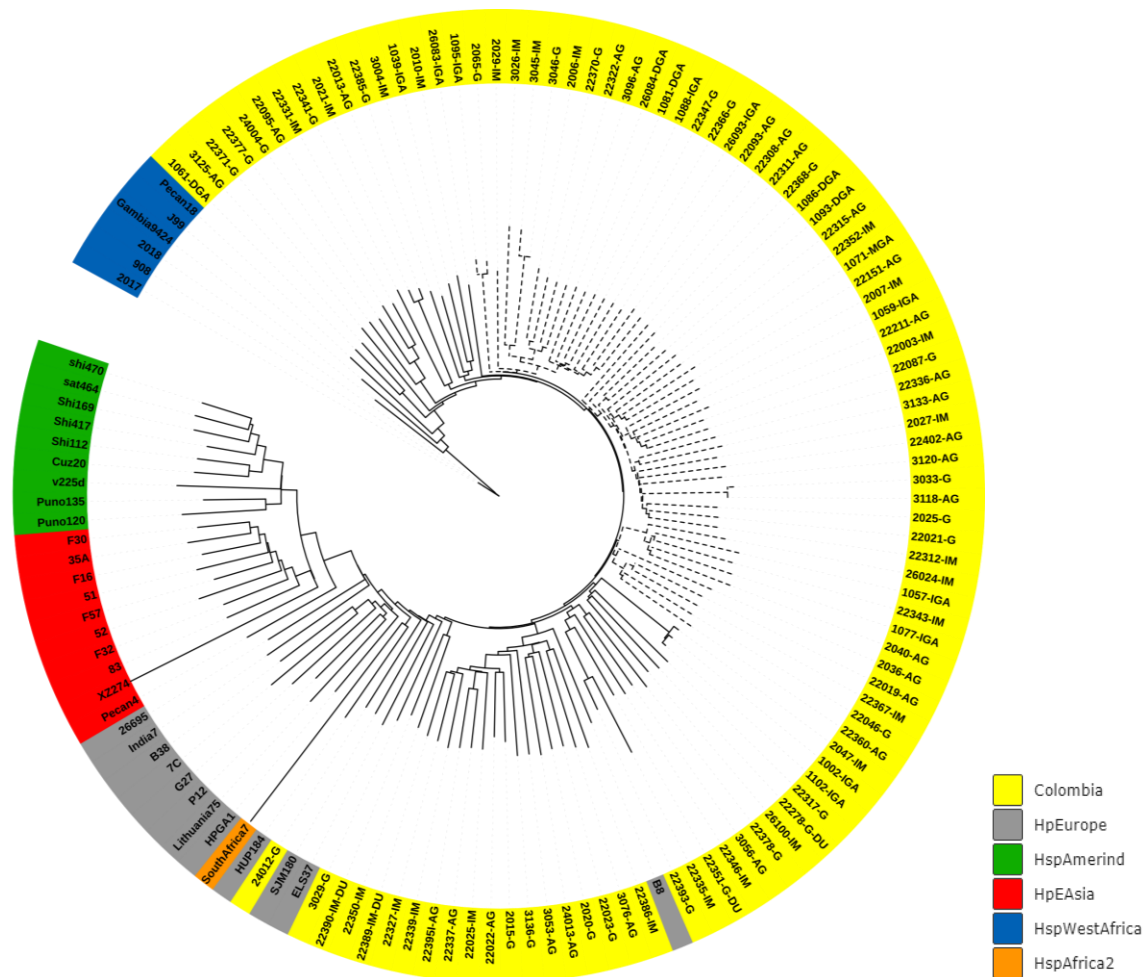


Figura 6. Filogenómica de aislados colombianos de *Helicobacter pylori*. Poblaciones control usados en la reconstrucción filogenómica hpEurope (gris) genomas: 26695, B8, B38, ELSE37, G27, HPAG1, Lithuania75, P12 y SJM180; hpWestAfrica (azul) genomas: 908, 2017, 2018, Gambia94-24, J99 y PeCan18; hpAfrica1 (gris) genomas: Southafrica 7, Southafrica 20; hspSouthIndia (magenta) genomas: India7 y SNT49; hpEastAsia (rojo) genomas: 35A, 51, 83, F16, F30, F32, F57 y XZ274, hspAmerind (verde) genomas: Cuz20, PeCan4, Puno135, Sat464, Shi112, Shi169, Shi417, Shi470 y v225d y genomas Colombia (Amarillo). La filogenómica obtenida a partir de 137 genomas completos de *Helicobacter pylori* ver Tabla S1 para información de los genomas colombianos.

El análisis filogenómico basado en un método libre de alineamiento fue menos intensivo computacionalmente y mostró resultados con mucha mejor resolución topológica del árbol. Se observaron genomas colombianos relacionados con los grupos poblacionales HpEurope y HspWestAfrica, pero ninguna relación con los grupos poblacionales HspAmerind o

HspAsia, y se determinó la presencia de cinco a siete clados exclusivamente conformados por genomas de aislamientos colombianos (Figura 7).

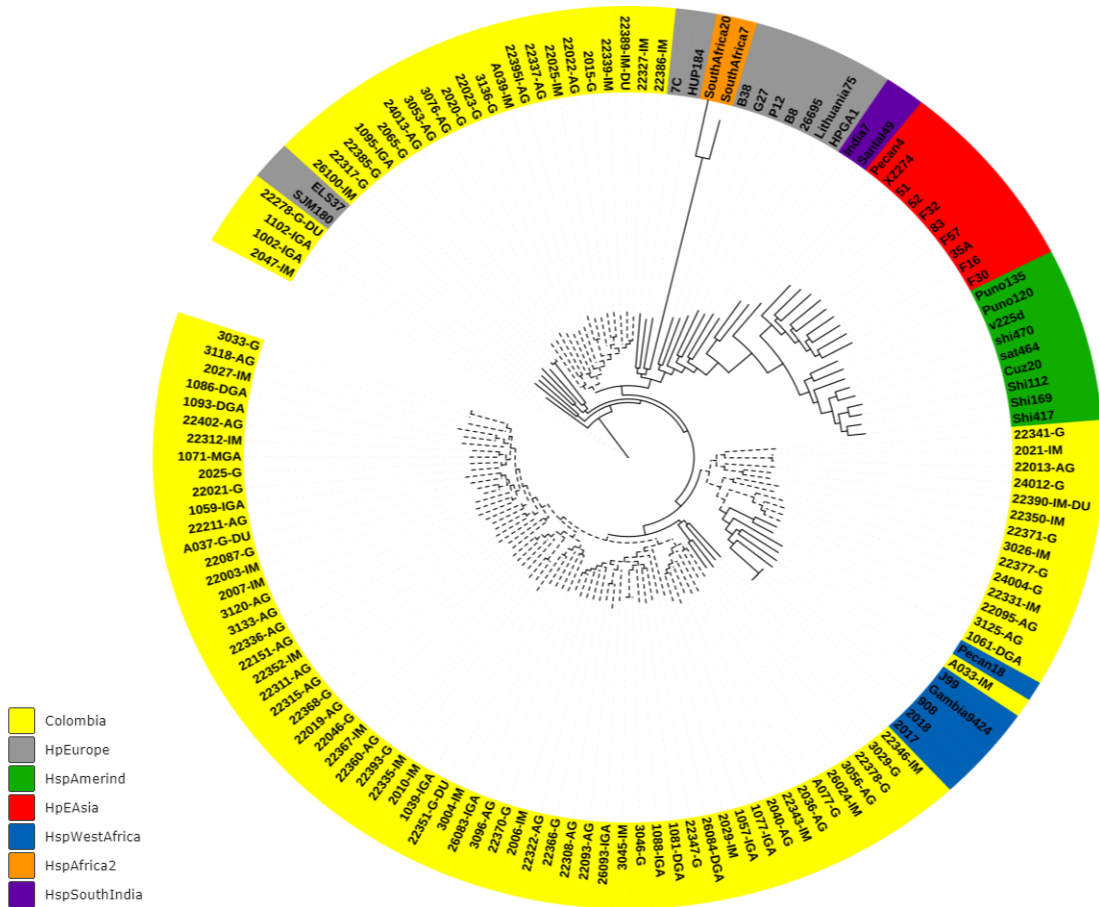


Figura 7. Árbol filogenético para los PNS conservados analizados con kSNP v3.0. Las líneas puntuadas representan los clados colombianos.

7.4. Diversidad genética de factores de virulencia

Con el fin de identificar si en esta región del país la bacteria está bajo un proceso evolutivo de carácter local, se estudiaron los patrones de variación genética y evolutiva en genes codificantes para cuatro factores de virulencia: *AlpA*, *VacA*, *HorB* y *Tipα*. La elección de estos genes se basó en tres criterios principales: i) Ser especie específicos, ii) Tener un papel importante en el proceso infeccioso de la bacteria y iii) Estar presentes en todos los aislamientos estudiados. Con el fin de entender mejor la diversidad de los cuatro genes, los grupos de secuencias de referencia y colombianas fueron comparados separadamente y en conjunto. Los siguientes parámetros genéticos poblacionales básicos fueron determinados: el número de haplotipos (H), la diversidad haplotípica (H_d), diversidad nucleotídica (π), número promedio de diferencias nucleotídicas entre haplotipos (k), la heterogeneidad y flujo génico en la población global (prueba S_{nn}) y por pares de poblaciones (F_{st} y Γ_{St}), recombinación (RM) y conversión génica (CG).

El gen *alpA* presentó el mayor número de haplotipos con 134 y el gen *tipα* el menor número con 104 para el total de secuencias. En los aislamientos colombianos estos mismos genes también mostraron el mayor y el menor número de haplotipos. Sin embargo, en el grupo de referencia no hubo variaciones, siendo el número de haplotipos de 33 y 32 respectivamente.

Con respecto a la diversidad nucleotídica (π), en el grupo de referencia el gen *tipα* fue el más diverso con un valor de 0,382 y el gen *horB* presentó el menor valor, 0,104 y los genes *alpA* y *vacA* mostraron valores intermedios de 0,152 y 0,171. Los aislamientos colombianos presentaron una menor diversidad nucleotídica con respecto al grupo de referencia, con

valores que oscilaron entre 0,061 y 0,031. En los haplotipos (k), se observó un comportamiento contrario al observado en diversidad nucleotídica, lo que indicó que estos genes tienen el potencial de presentar cargas mutacionales diferenciadas en la zona de estudio. En los cuatro genes se observaron eventos de recombinación y conversión génica, siendo los genes *vacA* y *alpA* los que presentaron el mayor número de eventos mientras que *horB* y *tipa* los menores. Esto significa que las reorganizaciones y los cambios sitio específicos a nivel de secuencia pueden ser una de las principales causas del proceso evolutivo de estos genes. Una característica común para todos los grupos evaluados fue la alta diversidad haplotípica que en promedio fue de 0,994. El gen *vacA* representó un desafío para el análisis evolutivo debido a que es muy polimórfico y ha sufrido múltiples eventos de recombinación, inserciones y deleciones. Aunque los análisis presentados hasta el momento son coherentes entre los cuatro genes, debe notarse que, para este gen en particular, solo se seleccionaron los sitios parsimoniosos.

Con el propósito de realizar un análisis evolutivo más exhaustivo de la citotoxina VacA en los aislamientos colombianos, se realizó un análisis sobre el dominio p55 de esta proteína. Para este análisis se emplearon solo las secuencias proteicas correspondientes al dominio que se denominó como VacAp55col y seis secuencias correspondientes a los tipos m1, m2 y a la quimera m1/m2 que fueron usadas como referencias. Los alineamientos mostraron que el 84,4% de las secuencias provenientes de los aislamientos colombianos era de tipo m1, mientras que el 15,2% de tipo m2. La región codificante para el dominio VacAp55Col presentó 431 sitios segregantes y 95 haplotipos con una diversidad de 0,9999. La diversidad nucleotídica (π) fue de 0,0741, el valor para teta (θ_w) fue 0,0619 y el número promedio de diferencias nucleotídicas (k) fue 99,523. Finalmente, la prueba Tajima (D) fue negativa para

los cuatro genes, sugiriendo que alelos raros para estos genes se han fijado en la población y a su vez, que una rápida expansión poblacional ocurrió después de un efecto cuello de botella, ver tabla 1.

Tabla 1. Análisis de diferenciación y flujo genético de tres alelos de aislamientos colombianos de *H. pylori*.

	<i>D</i>	Estadísticos poblacionales					Diferenciación genética		Flujo genético	
		<i>H</i>	<i>Hd</i>	<i>Pi</i>	<i>k</i>	<i>RM</i>	<i>CG</i>	<i>Snn</i>	<i>GammaST</i>	<i>Fst</i>
VacA <i>Colombian and reference</i>	-0.401	126	0.999	0.106	64.83	92	44	0.910***	0.165 (2.52)	0.074 (6.19)
<i>Colombian</i>		99	0.999	0.061	40.77	98				
<i>Reference</i>		33	0.997	0.171	120.284	112				
HorB <i>Colombian and reference</i>	-1.999	120	0.996	0.051	37.21	48	26	0.921***	0.053 (8.92)	0.077 (5.98)
<i>Colombian</i>		91	0.995	0.031	25.751	43				
<i>Reference</i>		33	0.995	0.104	76.4	34				
AlpA <i>Colombian and reference</i>	-1.476	134	0.999	0.067	81.411	98	54	0.870***	0.178 (2.30)	0.120 (3.63)
<i>Colombian</i>		102	0.998	0.039	53.335	91				
<i>Reference</i>		33	0.998	0.152	200.868	70				
Tipa <i>Colombian and reference</i>	-0.168	104	0.984	0.1	42.423	27	18	0.929***	0.370 (0.85)	0.284 (1.25)
<i>Colombian</i>		73	0.974	0.043	12.728	25				
<i>Reference</i>		32	0.995	0.382	104.057	23				

***, P<0.001

Se muestran parámetros para diversidad genética, D: prueba Tajima, P: Poblaciones comparadas, H: número de haplotipos, Hd: Diversidad haplotípica, Pi: Diversidad de nucleótidos, k: Carga genética, Rm: recombinación y CG: conversión génica. Para diferenciación genética se usó la prueba Snn y para flujo génico los test GammaST y Fst. El paréntesis para las pruebas de flujo genético es el número Nm para cada una. Los análisis fueron desarrollados en DnaSP. (Valores de P *: 0.05, **: 0.01 – 0.05, ***: 0.001).

Con el fin de explorar posibles eventos selectivos y el balance entre cambios neutrales se realizó para cada gen un análisis de cambios sinónimos y no sinónimos ($\omega=K_a/K_s$) a lo largo de la secuencia. Tanto en los cuatro genes completos como en la región codificante para el dominio VacAp55Col se observaron regiones con valores superiores a 1 indicando posibles eventos de selección positiva y/o divergencia funcional. (Figura 8).

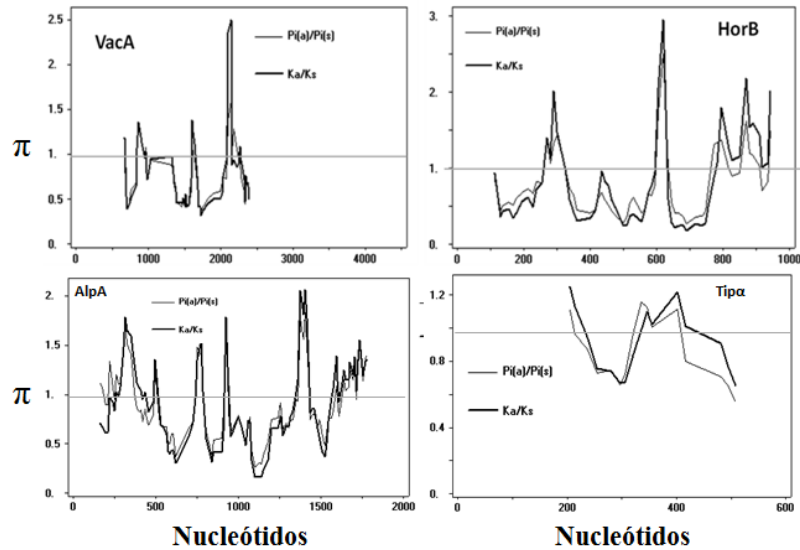


Figura 8. Análisis de Ka/Ks contra posición nucleotídica. Se emplearon 137 secuencias. Los análisis fueron desarrollados en DnaSP.

El árbol filogenético para AlpA, usando solo las secuencias proteicas de los genomas de referencia, fue caracterizado en un estudio previo (Gutiérrez-Escobar, 2013). En dicho estudio se determinó que esta filogenia presentó una clara subdivisión entre grupos poblacionales de la bacteria del este y oeste. En un clado se observaron las poblaciones HspAmerind y HspAsia divididas en dos subclados, en el otro las poblaciones HpEurope y HspSouthIndia. Finalmente, en el último grupo solo se observaron secuencias de tipo HspWestAfrica y HpAfrica2. Una característica de este árbol es que no sugirió eventos de recombinación entre el este y el oeste.

Cuando las secuencias de los aislamientos colombianos fueron incluidas en la filogenia se identificaron 5 clados mayores que siguen un intrincado patrón de ramas. 1- Un grupo con las poblaciones HspWestAfrica, Colombia y HpEurope; 2- Un grupo con las poblaciones HpEurope, HspSouthIndia y otras secuencias de la población Colombia, 3- Un grupo único

de la población Colombia; 4- Un grupo representado por las poblaciones HspAsia y HspAmerind, y finalmente 5- Un grupo con representantes de HpEurope, HspAsia y Colombia. La introducción de las secuencias de los aislamientos colombianos modificó la estructura bilobular del árbol, indicando que las secuencias colombianas presentan componentes ancestrales de varias poblaciones, lo que les permite asociarse de forma diferencial con cualquier grupo geográfico de la bacteria.

La reconstrucción filogenética de los aislamientos colombianos para AlpA mostró una clara diferenciación en tres clados principales. Los tres clados fueron denominados Col1, Col2 y Col3. De hecho, un análisis más detallado permitió identificar siete subclados denominados del 1 al 7 que sugieren que un intenso proceso microevolutivo está ocurriendo en la población colombiana estudiada para esta proteína y quizás, eventos de divergencia funcional que han ocurrido recientemente. (Figura 9AyB).

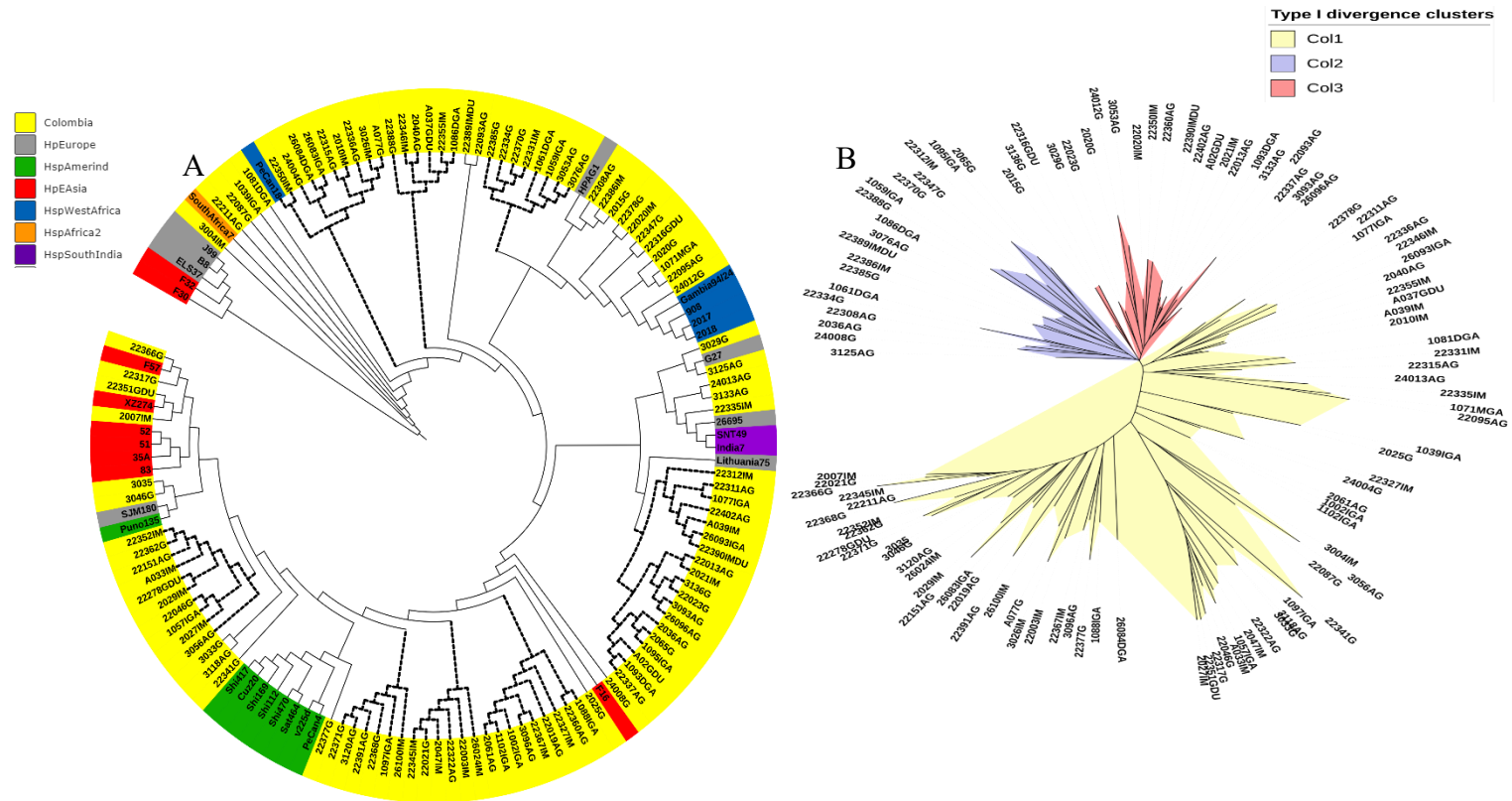


Figura 9. Reconstrucción filogenética para la adhesina AlpA de *Helicobacter pylori*. A) Se incluyeron 137 secuencias; las filogenias fueron inferidas usando el algoritmo NJ y la distancia fue computada usando el método 2 de Kimura con distribución gamma de 1 y todos los gaps fueron eliminados. B) Análisis filogenético exclusivo de aislamientos colombianos revelando tres clados bien diferenciados. Para todas las reconstrucciones se emplearon 1000 repeticiones bootstrap como soporte estadístico y solo se mostraron los nodos significativos.

En el mismo estudio que se reportó la filogenia de AlpA también se analizó y describió el proceso evolutivo de la adhesina HorB (Gutiérrez-Escobar, 2013). La filogenia de HorB también se caracterizó por ser bilobular y dividir las poblaciones en dos grandes grupos. Sin embargo, para este gen las poblaciones HpEurope y HspAsia se ubicaron en un mismo clado y las poblaciones HspAmerind, HspSouthIndia, HspWestAfrica y HpAfrica2 en el otro clado y curiosamente las poblaciones de tipo africano no presentaron un clado diferenciado.

Al incorporar las secuencias colombianas se observaron cuatro clados relativamente bien definidos. El primer clado agrupó las poblaciones HspWestAfrica y Colombia con un subclado muy bien definido para esta última subpoblación, el segundo mostró grupos de las poblaciones HpEurope, HspAmerind y Colombia, el tercero grupos mixtos HpEurope/Colombia y HpAsia/Colombia y el último conformado por secuencias dispersas y otros subclados tipo Colombia (Figura 10A). El árbol filogenético desarrollado exclusivamente con aislamientos colombianos mostró cinco clados bien diferenciados, quizás indicando eventos de divergencia funcional (Figura 10B).

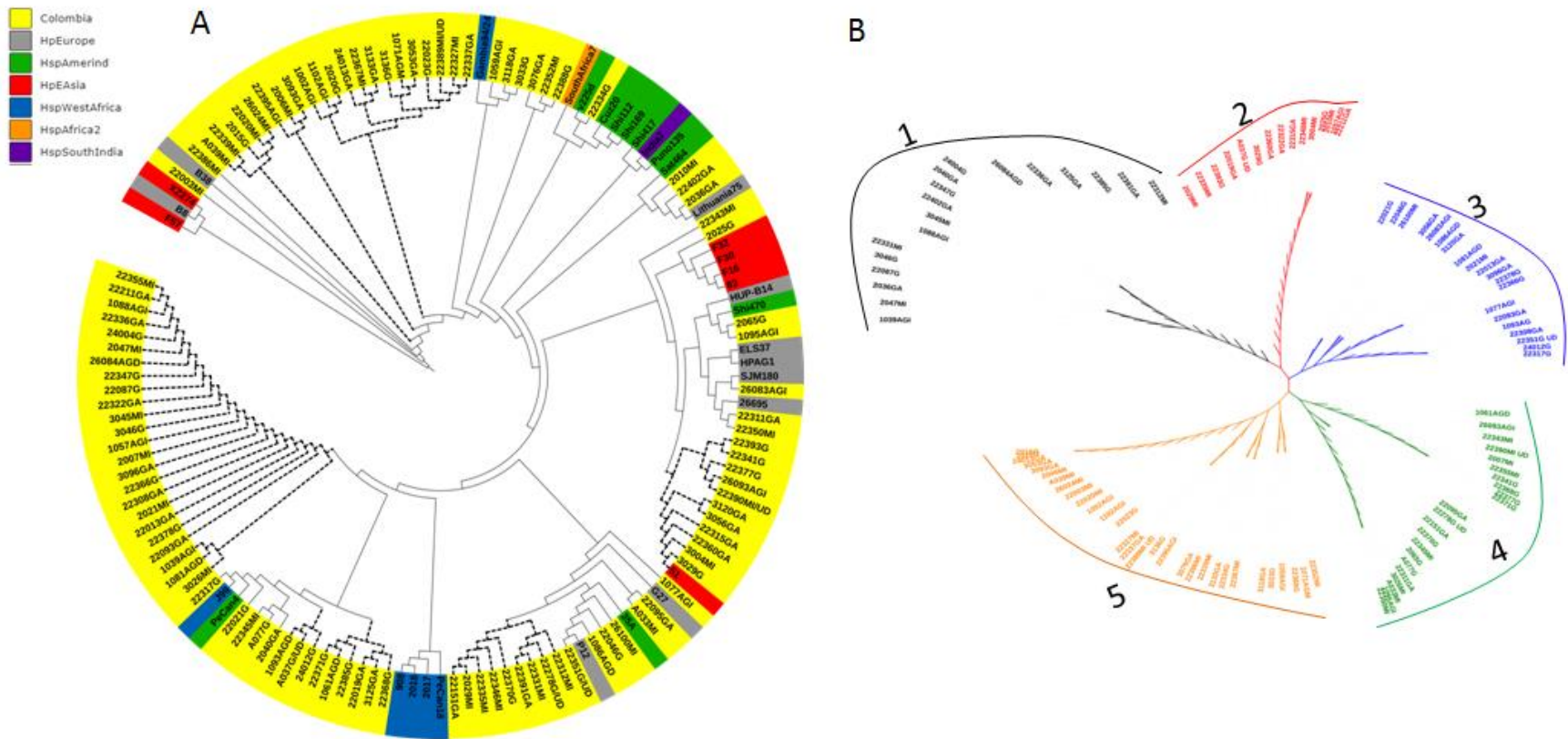


Figura 10. Reconstrucción filogenética para la adhesina HorB de *Helicobacter pylori*. A) Se incluyeron 137 secuencias; las filogenias fueron inferidas usando el algoritmo NJ y la distancia fue computada usando el método 2 de Kimura con distribución gamma de 1, los sitios invariantes y todos los gaps fueron eliminados. B) Análisis filogenético exclusivo de aislamientos colombianos revelando cinco clados bien diferenciados. Para todas las reconstrucciones se emplearon 1000 repeticiones bootstrap como soporte estadístico y solo se mostraron los nodos significativos.

La filogenia para VacA ha sido ampliamente estudiada, no solo a nivel global sino también para los dominios p33 y p55 (Gangwer et al., 2010). Cuando se empleó la secuencia completa o casi completa de la proteína se identificaron tres grupos principales correspondiente a cepas asiáticas, no asiáticas y un grupo mixto y se observó una distribución coherente con las regiones m1, m2 y quimérica; el mismo patrón se observó con el dominio p55 (Gangwer et al., 2010). (Figura 11A). El árbol filogenético con las secuencias de los aislamientos colombianos mostró un patrón de distribución similar, pero más disperso que el de las otras proteínas. De izquierda a derecha se lograron identificar 4 grupos: El primero conformado por las poblaciones Colombia, HspAmerind, HspWestAfrica y HpEurope; el segundo por 2 subclados Colombia; el tercero, varios subclados con miembros de las poblaciones HpAsia, HspAmerind, HpEurope y Colombia; y el último un gran clado mixto con miembros de todos los grupos de poblaciones. (Figura 11B).

Debido al excesivo polimorfismo y la baja resolución del árbol con las secuencias completas se procedió a realizar la reconstrucción filogenética usando el dominio p55. El árbol filogenético para VacAp55col con las secuencias de aislamientos colombianos mostró claramente dos clados principales, uno con secuencias del tipo m2 y otro con secuencias tipo m1. Se observaron ramas muy cortas y cinco subclados muy bien diferenciados para el tipo m1. También se encontraron ramas dispersas que corresponden a las secuencias quimera m1/m2. (Figura 11B).

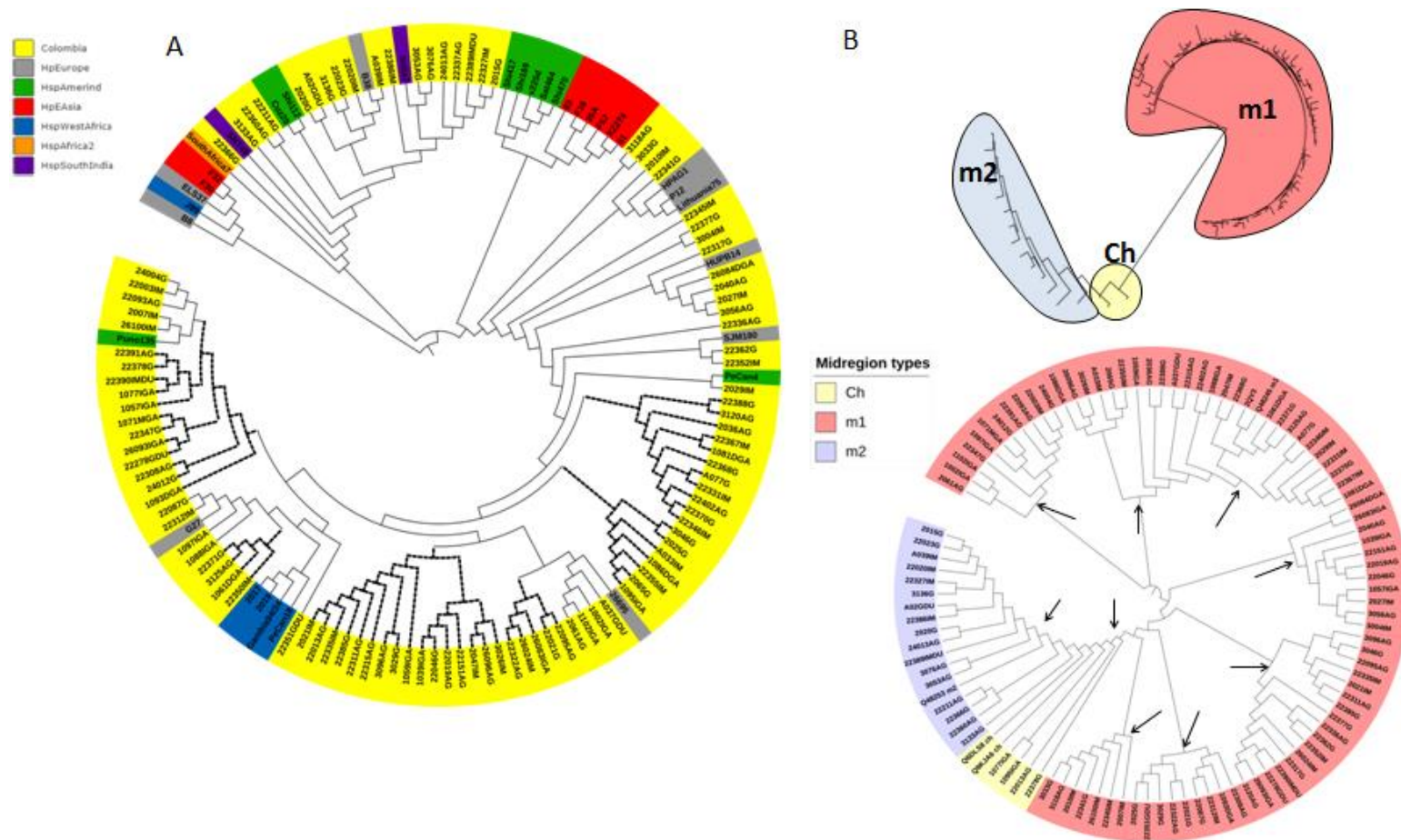


Figura 11. Reconstrucción filogenética para la toxina VacA de *Helicobacter pylori*. A) Se incluyeron 137 secuencias completas de la citotoxina; las filogenias fueron inferidas usando el algoritmo NJ y la distancia fue computada usando el método GTR con distribución gamma, todos los gaps fueron eliminados. B) Análisis filogenético exclusivo del dominio VacAP55Col de los aislamientos colombianos revelando tres clados bien diferenciados panel superior y por lo menos cinco subclados para el tipo m1. Para todas las reconstrucciones se emplearon 1000 repeticiones bootstrap como soporte estadístico y solo se mostraron los nodos significativos.

La filogenia para la proteína Tip α para las 34 secuencias de referencia mostró una división poblacional acorde a lo descrito para *H. pylori*. Se identificaron clados para cada grupo poblacional: HspWestAfrica, HspAmerind, HpAsia, un clado mixto HpEurope/HspSouthIndia y un clado para HpAfrica2. (Figura 12A). Al adicionar las secuencias de los aislamientos colombianos a la filogenia, se observó un patrón similar al observado previamente con las otras proteínas, se conformaron múltiples clados mixtos y clados con solo presencia de la población Colombia. Una vez más, la recombinación y la convergencia génica probablemente están afectando la distribución observada en el árbol control para este factor de virulencia. (Figura 12AB). El análisis filogenético usando solo secuencias de los aislamientos colombianos mostró una distribución muy similar a la observada para AlpA donde se obtuvieron tres clados bien diferenciados. (Figura 12C).

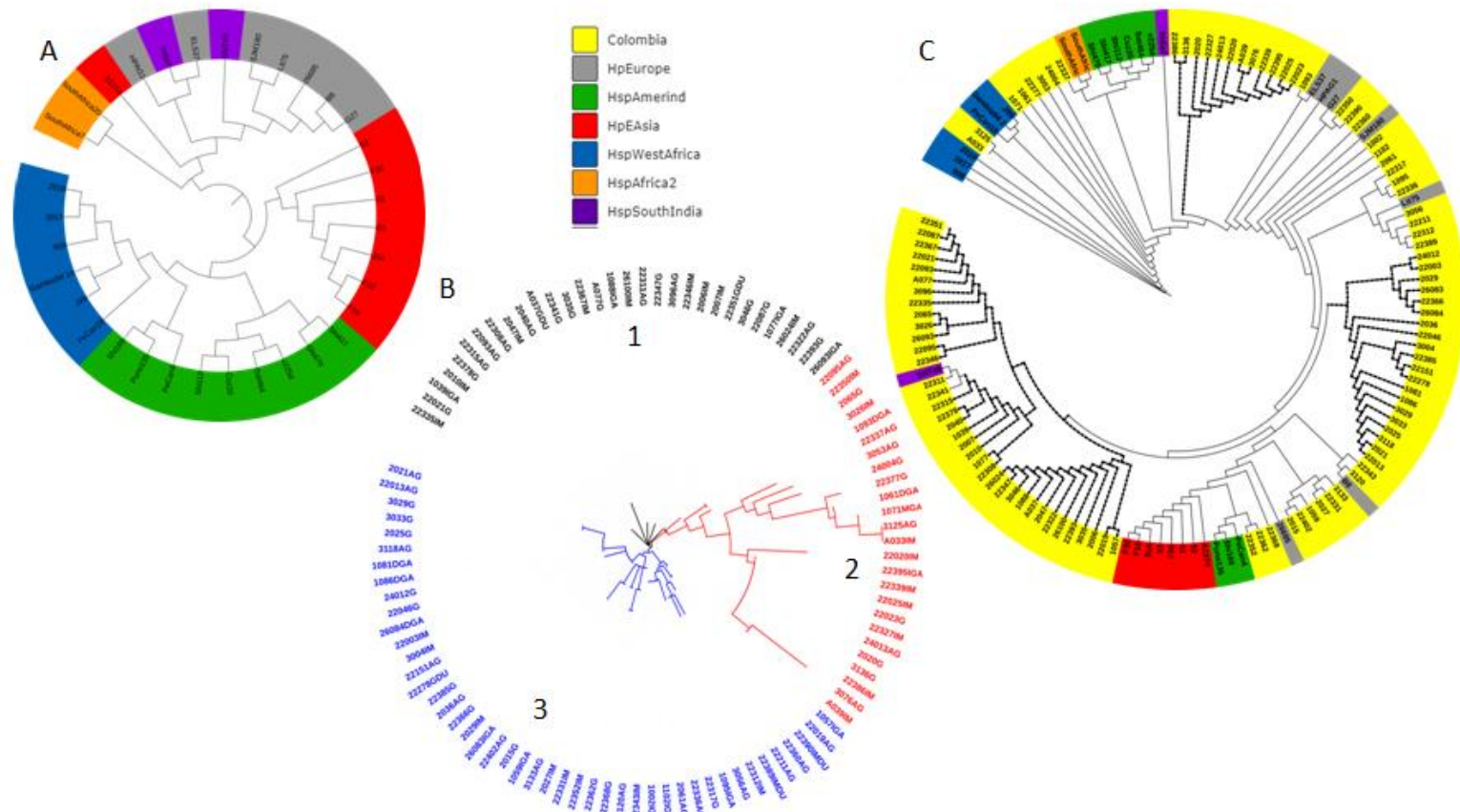


Figura 12. Reconstrucción filogenética para la adhesina Tipa de *Helicobacter pylori*. A) Se incluyeron 37 secuencias completas del factor de patogenicidad como referencia. Las filogenias fueron inferidas usando el algoritmo NJ y la distancia fue computada usando el método LG con distribución gamma. B) Análisis filogenético con todas las secuencias proteicas. C) Árbol filogenético para la proteica usando solo los aislamientos colombianos. Para todas las reconstrucciones se emplearon 1000 repeticiones bootstrap como soporte estadístico y solo se mostraron los nodos significativos. Todos los gaps fueron eliminados.

Los árboles filogenéticos usando solo las secuencias de proteínas de los aislamientos colombianos, dieron cuenta de la generación de clados independientes densamente ramificados, recombinantes y bien definidos. Además, los resultados obtenidos para el parámetro ω indicaron posibles eventos de divergencia funcional y selección positiva que dieron forma a las variantes alélicas circulantes para estos factores de virulencia en la zona de estudio.

7.5. Determinación de eventos de divergencia funcional y selección positiva para los factores de virulencia

El análisis de divergencia funcional de la proteína AlpA realizado a partir del árbol filogenético de los aislamientos colombianos mostró que los residuos A219, G446, Q190, V/S65, V23 presentaron un $\theta_D > 0,8$ entre los clados Col1/Col3, señalando ritmos evolutivos diferentes entre los aislamientos colombianos. Para el análisis de selección positiva y negativa de AlpA, se analizaron 540 sitios empleando el algoritmo FEL. Este algoritmo mostró que el 3.3% de los sitios de la proteína están bajo selección positiva y que el 21.4% evolucionó bajo selección negativa. Se identificó que los residuos L52, S65, Y103, M131, S206, V226, S233, Q234, T244, M294, I455 y V471 presentaron eventos de selección positiva. La predicción de epítomos lineales para linfocitos B mostró que 45.1% de la secuencia proteica de este factor tiene potencial de ser reconocida por anticuerpos. Se evidenció que el residuo V65 bajo divergencia funcional y selección positiva podría también estar bajo selección inmune, al encontrarse en péptidos predichos a ser reconocidos por el sistema inmune del hospedero. Una situación similar se observó para los residuos T244 e I455 bajo selección positiva, ver tabla 2 y tabla 3. La predicción para epítomos del complejo principal de histocompatibilidad tipo II (CPH II) dió como resultado, que los residuos bajo

divergencia funcional V65 y V23, y todos los residuos bajo selección positiva se ubicaron en péptidos con el potencial de ser presentados a linfocitos CD4⁺.

Aunque la filogenia para HorB de los aislamientos colombianos mostró cinco clados bien diferenciados, no se identificaron sitios bajo divergencia funcional. Sin embargo, se determinó que el 1.3% de los sitios de la proteína HorB evoluciona bajo selección positiva, siendo los residuos I10, L192, N253 y T266 los más significativos y que el 29.86% de los sitios evoluciona bajo selección purificadora, ver tabla 2 y tabla 3. La predicción para epítomos B mostró que el 48.25% de la secuencia de proteína podría ser reconocida por anticuerpos, y se identificó que el residuo T266 se encontró en zona de reconocimiento para linfocitos B. Con respecto a la predicción de epítomos CPH II, se observó que los residuos L192 y N253 bajo selección positiva se ubicaron en péptidos con puntajes significativos para ser reconocidos por los linfocitos CD4⁺ del hospedero.

La reconstrucción filogenética para el dominio VacAp55Col de los aislamientos colombianos mostró cinco subclados bien diferenciados: dos de tipo m1, dos para quimeras m1/m2 y uno para el tipo m2. Esta topología fue usada para la identificación de eventos de divergencia funcional que fueron mapeados contra la estructura 2QV3; sin embargo, no se encontraron resultados significativos, ver tabla 2. Se lograron identificar eventos de selección positiva y negativa para el dominio VacAp55Col, de los 491 sitios estudiados usando la prueba FEL, se logró identificar que el 3.7% está bajo selección positiva y que el 22.2% está bajo selección negativa o purificadora. Los sitios N11, Y21, H33, V39, K206, A241, E246, L645, Y321 y P443 evolucionan bajo un modelo de selección positiva, ver tabla 3.

La predicción de epítomos lineales tipo B mostró que 48.3% de la secuencia proteica del dominio tiene potencial de ser reconocida por anticuerpos y los residuos N11, Y21, H33, V39, K206, A241, P443 y E246 bajo selección positiva fueron reconocidos en regiones de reconocimiento. La predicción para epítomos para linfocitos T mostró que los residuos N11, Y21, H33, K206, A241 y E246 bajo selección positiva se ubicaron en péptidos predichos para ser reconocidos por linfocitos CD4⁺ del hospedero.

La filogenia para la proteína Tipα de los aislamientos colombianos mostró tres clados. Las comparaciones entre los tres clados mostraron cinco sitios con un $\theta D > 1$ indicando que los residuos L1, N127 S135 y N146 mostraron importantes cambios en los ritmos evolutivos. El algoritmo FEL identificó 2 sitios bajo selección positiva V145 y V150, y 34 sitios bajo selección purificadora. Se determinó que el 47,7% de la secuencia presentó potencial de ser reconocida por anticuerpos, y los residuos bajo divergencia funcional L1, S62, N127, N146 y los dos residuos bajo selección positiva se encontraron en zonas de la proteína predichas como epítomos lineales tipo B. La predicción de epítomos tipo T mostró que los residuos S62, N127 y D135 se ubicaron en péptidos reconocidos significativamente, ninguno de los residuos bajo selección positiva fue reconocido, ver tabla 2 y tabla 3.

Tabla 2. Análisis de divergencia funcional en factores de patogenicidad de aislamientos colombianos de *H. pylori*.

Proteína	Clados	Residuo	θ	LRT	Valor P
AlpA	Clado1/Clado3	446	0.96	17.54	0.000028
		219	0.90		
		65	0.89		
		190	0.85		
		23	0.80		
		187	0.78		
		197	0.78		
		101	0.62		
		181	0.62		
		132	0.61		
		58	0.61		
		110	0.58		
		113	0.57		
		253	0.56		
285	0.52				
282	0.52				
Tipa	Clado 2/Clado 4	1	1.00	18.18	0.000020
		62	1.00		
		127	1.00		
		135	1.00		
		146	1.00		
VacAP55Col	Clado 1/Clado 2	454	0.81	3.12	0.077813
		342	0.72		
HorB	Clado 1/Clado 3	95	0.5	3.09	0.078774

Los coeficientes de divergencia (θ D) y sus valores LRT fueron obtenidos a partir de comparaciones de los clados detectados en los análisis filogenéticos con aislamientos colombianos de cada factor de patogenicidad estudiado. El valor LRT θ D es el puntaje logarítmico de la prueba del radio de verosimilitud contra la hipótesis nula (θ D = 0) (Gu, 1999). Este es el valor que arroja DIVERGE como resultado y sigue una distribución chi cuadrada con un grado de libertad, entonces valores superiores a 3,84 son indicadores de divergencia funcional entre clados.

Tabla 3. Análisis de selección positiva en factores de patogenicidad de aislamientos colombianos de *H. pylori*.

	Codon	dS	dN	dN/dS	dN-dS	p-value
AlpA	52	0.255	2.085	8.166	1.084	0.0069
	65	0.000	5.375	Infinite	3.185	0.0000
	99	0.000	2.165	Infinite	1.283	0.0060
	103	0.000	1.665	Infinite	0.987	0.0017
	131	0.000	2.584	Infinite	1.531	0.0001
	206	0.000	5.397	Infinite	3.197	0.0000
	226	0.000	1.860	Infinite	1.102	0.0071
	233	0.410	5.838	14.238	3.216	0.0003
	234	0.000	4.229	7.795E+16	2.505	0.0003
	244	0.000	1.949	Infinite	1.155	0.0010
	294	0.291	4.265	14.679	2.355	0.0040
	303	0.000	1.393	Infinite	0.825	0.0052
	310	0.000	3.836	Infinite	2.272	0.0000
	314	0.000	21.309	7.5008E+15	12.625	0.0062
	318	0.286	6.533	22.843	3.701	0.0000
	332	0.000	1.339	Infinite	0.793	0.0092
	455	0.000	2.537	Infinite	1.503	0.0003
471	0.000	3.228	Infinite	1.912	0.0001	
Tpa	145	0.000	1.103	2.6934E+16	1.761	0.0480
	150	0.000	3.259	Infinite	5.202	0.0065
VacAP55Col	11	0.000	2.043	Infinite	2.135	0.0151
	21	0.000	4.172	Infinite	4.361	0.0017
	33	0.000	3.374	Infinite	3.527	0.0037
	39	0.000	2.787	Infinite	2.913	0.0007
	206	0.000	2.307	Infinite	2.411	0.0150
	241	0.000	1.811	Infinite	1.893	0.0241
	246	0.000	1.013	Infinite	1.059	0.0309
	321	0.000	3.523	Infinite	3.683	0.0277
	443	0.000	3.849	2.42E+15	4.024	0.0009
HorB	10	0.000	0.056	0.152	5.696	0.017
	192	0.000	2.607	1.87	9.208	0.002
	253	0.000	1.174	0.823	7.012	0.008
	266	0.000	2.142	1.477	11.760	0.001

Este método emplea una aproximación por máxima verosimilitud para inferir los ritmos de cambio sinónimos dS y no-sinónimos dN a partir del alineamiento proteico y su árbol filogenético. Después de la optimización del largo de las ramas del árbol y de los parámetros de sustitución de nucleótidos, se fija el modelo MG94xREV para cada codón para inferir los ritmos de cambio sinónimos y no-sinónimos. La prueba de hipótesis sitio específica se realiza mediante la prueba LRT para asegurar que dN es significativamente más grande que dS.

7.6. Carga mutacional diferencial para los factores de virulencia por grupos de patología gástrica y análisis topológico

Con el fin de identificar si existían posiciones polimórficas en el alineamiento de proteínas de los factores de virulencia significativamente diferentes entre las patologías gástricas de la población de estudio o una carga mutacional diferencial, se realizó el siguiente análisis: Las secuencias se agruparon de la siguiente manera: 1- G gastritis, 2- GA gastritis atrófica, 3- MI metaplasia intestinal y 4- AG adenocarcinoma gástrico para cada factor de virulencia.

Solo las proteínas AlpA y VacAp55Col mostraron resultados significativos en este análisis. En la proteína AlpA se obtuvieron resultados positivos para el sitio N370 con un valor $p < 0,04$, que mostró correlaciones significativas para diferenciar los grupos: G/GA ($R=0,999$), G/MI ($R=0,642$), MI/AG ($R=0,63$) y GA/MI ($R=0,63$), la predicción para epítomos de la inmunidad humoral mostró que este residuo se ubicó en un péptido con el potencial de ser presentado por el CMH II, ver tabla 4.

El dominio VacAp55Col mostró dos sitios, el primero I261 $p < 0,002$ con el potencial de diferenciar los grupos: G/MI ($R=1$), G/AG ($R=1$), MI/AG ($R=1$) y el segundo sitio K321 $p < 0,03$ con correlaciones significativas para diferenciar los grupos: GA/MI ($R=1$), G/MI ($R=0,999$) y G/GA ($R=0,992$). Los dos residuos fueron reconocidos en péptidos para ser reconocidos por el CPH II y el residuo Y321 también se ubicó en un péptido predicho a ser reconocido por la inmunidad humoral, ver tabla 4.

Tabla 4. Determinación de residuos polimórficos significativamente diferentes entre grupos de patología gástrica mediante el uso del servidor Meta-CATS

Proteína	Sitio	Grupos diferenciados	Correlación*	Valor p
AlpA	370	1 y 2	0.999	0.04
		1 y 3	0.642	
		3 y 4	0.63	
		2 y 3	0.63	
VacAp55Col	261	1 y 3	1	0.002
		1 y 4	1	
		3 y 4	1	
	321	2 y 3	1	0.03
		1 y 3	0.999	
		1 y 2	0.992	

Las secuencias fueron alineadas usando el programa Muscle y las columnas significativamente diferentes entre grupos se determinaron mediante una chi cuadrada de Person. Grupos: 1- Gastritis, 2- Gastritis atrófica, 3-

Metaplasia intestinal, 4- Adenocarcinoma gástrico. Las proteínas HorB y Tip α no mostraron resultados significativos.

Las estructuras proteicas predichas y los análisis topológicos posteriores para los aislamientos colombianos se eligieron según la posible diferenciación entre patologías gástricas. Es decir, que para este análisis se mostraron los resultados del dominio VacAp55Col con las mutaciones I261_K321_1071MGA, I261_Y321_26093IGA y V261_R321_22360AG y para la proteína AlpA con las mutaciones N370_22341G y S370_1071MGA. Los modelos para VacAp55col de los mutantes colombianos fueron reconstruidos a partir de la estructura 2QV3. Se observó una alta similitud estructural entre el modelo y el templado, tanto la región N como la C terminal del modelo para el dominio mostraron las láminas β y las hélices α previamente descritas en la estructura 2QV3. (Figuras 13 A y B).

Para AlpA los modelos se reconstruyeron *de novo* debido a que hasta el momento no se ha cristalizado esta proteína. La estructura que mayor similitud presentó fue la pilina adhesiva GBS104 del grupo B de *Streptococcus agalactiae* (Krishnan et al., 2013) con un C-score de 0,62 y 0,313 Å, curiosamente esta proteína también tiene como ligando la laminina. (Figura 13 E, D).

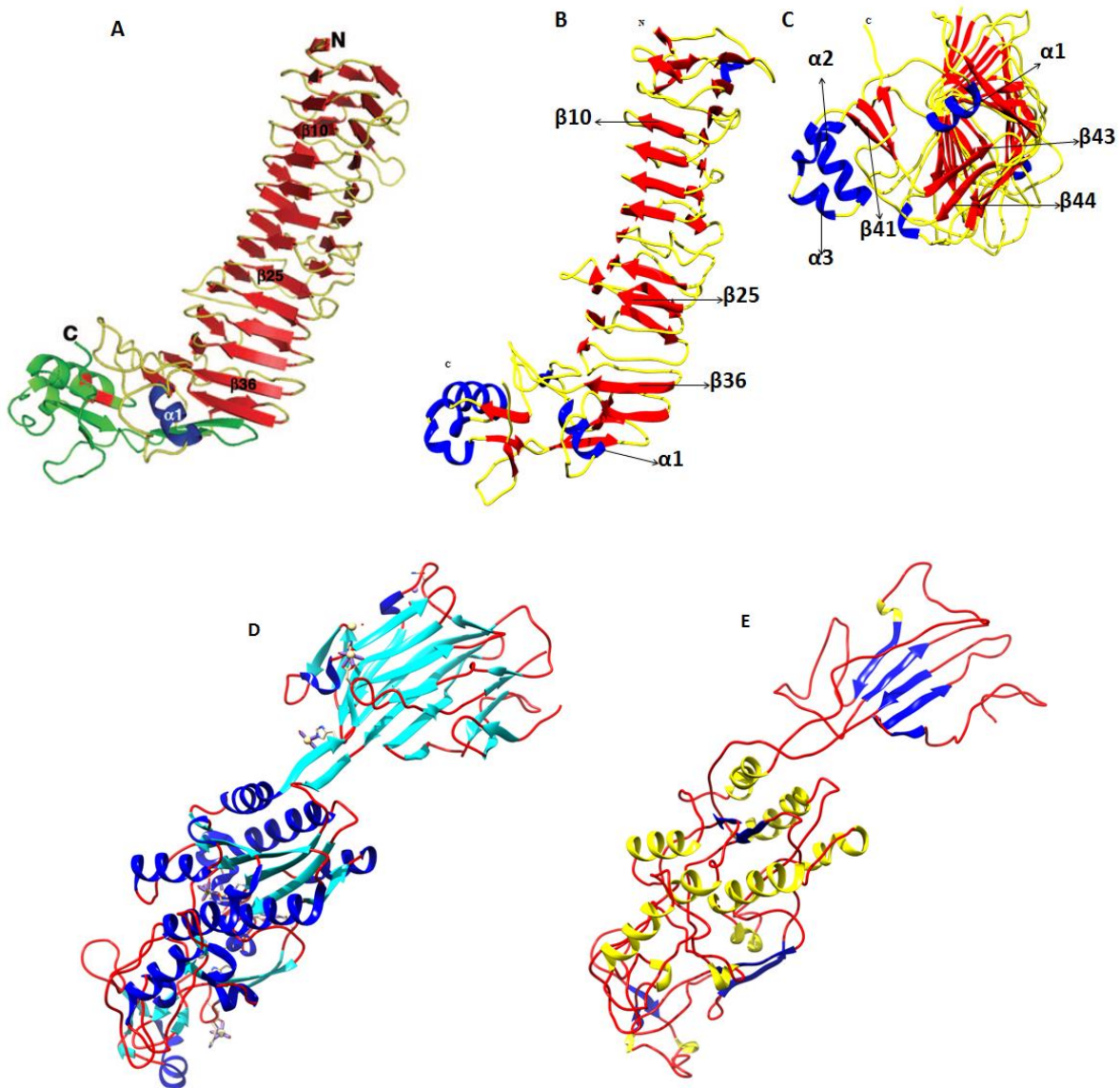


Figura 13. Modelo del dominio VacAp555Col usando el servidor I-TASSER. A) El modelo fue predicho a partir de la estructura con código PDB 2QV3. B) vista lateral y basal del modelo predicho para el dominio a partir de la variante I261_K321_1071MGA. El análisis por RMSD entre el modelo y la estructura fue de 0.182 Å° para los carbonos α y 0.61 Å° para el esqueleto. Se observa alta similitud en la composición de hélices y laminas. C) Estructura de la pilina GBS104 correspondiente con el código PDB 3TXA. E) Modelo de la variante S370_1071MGA de AlpA de aislamientos colombianos de la bacteria.

Los plots de Ramachandran para VacAp55Col, indicaron que para el modelo I261_K321_1071MGA el 81,5% de los residuos se ubicó en zonas favorecidas, el 12,5% en regiones permitidas y el 5,9% en regiones no permitidas; para el modelo I261_Q321_26093IGA el 81,5% de los residuos se ubicó en regiones favorecidas, el 9,7%

en regiones permitidas y el 5,7% fuera de las regiones permitidas; y finalmente el modelo para V261_R321_22360AG mostró que el 74,3% de los residuos se ubicó en zonas favorecidas, el 16,8% en regiones permitidas y el 8,8% en regiones no permitidas y finalmente. Para la estructura control 2QV3 se observó que el 93,7% de los residuos se ubicó en regiones favorecidas, el 5,0% en regiones permitidas y el 1,4% en regiones no permitidas, teniendo en cuenta que se considera que una estructura de alta calidad presenta un 90% de residuos en la región favorable, se consideró que la calidad de los modelos fue suficiente para los análisis propuestos.

Los plots de Ramachandran para AlpA mostraron que para el modelo 26695, el 70,2% de los residuos se ubicó en regiones favorables, el 19,7% en regiones permitidas y el 10,1% en regiones no permitidas. Para la variante N370_22341G el 81,5% de los residuos se ubicó en zonas favorables, el 13,3% en regiones permitidas y el 5,2% en zonas no permitidas, y para la variante S370_1071IM, el 78,4% de los residuos se ubicó en regiones favorables, el 15,8% en regiones permitidas y el 5,9% en regiones no permitidas. Para cada modelo, si se suman las zonas favorables y las regiones permitidas, se puede observar que están cerca de la regla del 90% y pueden ser considerados para análisis posteriores.

Las estructuras predichas fueron entonces transformadas a redes indirectas en donde se conservaron las interacciones entre los carbonos α . Los parámetros topológicos en promedio para cada red fueron los siguientes: Para B_i fue 0.01, para C_i 0.140, para Cl_i 0.420 y para k fue de 16.70. Individualmente, se observó que cada red presentó residuos diferentes para cada parámetro evaluado como sus principales nodos topológicos, indicando que cada proteína se comportó como una unidad discreta a pesar de pertenecer a la misma especie. Lo anterior

quizás sea el efecto de la recombinación y la conversión génica a escala topológica, ver tabla 5.

Tabla 5. Residuos topológicos esenciales para cada parámetro topológico estudiado

Modelos	Residuos			
VacAp55Col	Bi	Cli*	Ci	K
2QV3 (m1)	A619 (0.124)	G648, R552, S553, G527, E600, G709, A671, N715, N711, N806, G432 y N358.	I569 (1.174)	N408 (36)
I261_K321_1071AGM (m1)	L254 (0.107)	E246, G294, S199, A317, G355, N361, N357.	V234, F240 (0.181)	D216, R371, F272 (36)
I261_Y321_26093AGI (m1)	T237 (0.181)	G294, S199, S316, G355, N361 y N357.	T237 (0.181)	R371 (30)
V261_R321_22360GA (m2)	R212 (0.182)	G198, P224, D319, G271, G344, S341, G78, S388.	R212 (0.182)	R111, N109 (39)
AlpA				
26695	T102 (0.151)	Q93, Q289, V290, G246, T48, A301, A276, G390, K320, T460, Q350.	N29 (0.1909)	N6 (33)
N370_22341G	K410 (0.333)	A95, S463, Y461, L202, N317, P303, N263.	T418 (0.163)	R110 (31)
S370_1071AGM	H406 (0.300)	N352, L198, L199, G86, T240, S88, G168, T125, N126, P467, G324, V225, R305, T460, D281, V278, E18, P7, G8.	R181 (0.155)	Y373 (30)

El grado de un nodo (k), el coeficiente de agrupamiento (Ci), la intermediación (Bi) y la cercanía (Cli). Los rangos de todos los parámetros oscilan entre 0 y 1. * Todos los valores iguales a 1. Entre paréntesis, el correspondiente valor numérico para cada parámetro evaluado.

Con el fin de determinar si las redes de interacción de residuos correspondientes a las variantes de aislamientos colombianos eran diferentes de la estructura control, se realizó un análisis de perfiles topológicos para cada parámetro evaluado. Los perfiles topológicos indicaron que las variantes colombianas para VacAp55Col evidenciaron diferenciación a nivel topológico para la mayoría de los parámetros evaluados a excepción de *Bi*. Se identificó que otros parámetros topológicos también mostraron diferenciación. Por ejemplo, la conectividad mostró una diferenciación muy significativa (2QV3 vs I261_K321_1071MGA,

$p = 4.792e-39$, 2QV3 vs I261_Y321_26093IGA, $p = 1.976e-38$ y 2QV3 vs V261_R321_22360AG, $p = 2.458e-42$), el coeficiente topológico (2QV3 vs I261_K321_1071MGA, $p = 0.003$, 2QV3 vs I261_Y321_26093IGA, $p = 0.002$ y 2QV3 vs V261_R321_22360AG, $p = 0.0008$). Estos resultados sugieren, primero que los nodos tienden a estar correlacionados en grupos definidos por valores similares para el parámetro k haciendo que las redes sean asortivas y, segundo que la manera en como los nodos están asociados y el número de dichas asociaciones es diferente entre las redes.

La caracterización de los perfiles topológicos para AlpA mostró diferencias significativas para casi todos los parámetros evaluados excepto Bi . Estas redes mostraron diferencias topológicas muy significativas. No solo los parámetros básicos mostraron diferenciación, sino también, por ejemplo, la excentricidad (26695 vs N370_22341G, $p = 1.57148e-149$ y 22695 vs S370_1071AGM, $p = 6.615197e-90$) indicando que los nodos están diferencialmente conectados en las redes; la conectividad (26695 vs N370_22341G, $p = 3.425284e-07$ y 22695 vs S370_1071AGM, $p = 1.998464e-10$) indicando que los nodos se relacionan diferencialmente entre las variantes colombianas y la estructura control; y la radialidad (26695 vs N370_22341G, $p = 1.848185e-32$ y 22695 vs S370_1071AGM, $p = 5.263628e-28$) indicando que existen nodos en las variantes colombianas con una función de regulación de la red que son diferenciales con respecto a la estructura de control, ver tabla 6.

Tabla 6. Comparación de los parámetros topológicos básicos entre las redes de referencia y las redes de los modelos.

Parámetros topológicos	Referencia	Modelo	Valor P
<i>k</i>	226695	N370_22341G	0.027
		S370_1071MGA	1.00E-07
	2QV3	I261_K321_1071MGA	1.00E-05
		I261_Y321_26093IGA	7.00E-06
		V261_R321_22360AG	0.003
<i>Ci</i>	226695	N370_22341G	1.57E-67
		S370_1071MGA	6.61E-90
	2QV3	I261_K321_1071MGA	1.00E-05
		I261_Y321_26093IGA	2.83E-07
		V261_R321_22360AG	4.40E-05
<i>Cli</i>	226695	N370_22341G	1.57E-67
		S370_1071MGA	6.61E-90
	2QV3	I261_K321_1071MGA	0.026
		I261_Y321_26093IGA	0.022
		V261_R321_22360AG	0.021
<i>Bi</i>	226695	N370_22341G	0.841
		S370_1071MGA	0.988
	2QV3	I261_K321_1071MGA	0.282
		I261_Y321_26093IGA	0.457
		V261_R321_22360AG	0.073

Parámetros topológicos: grado (*k*), coeficiente de agrupamiento (*Ci*), intermediación (*Bi*) y cercanía (*Cli*). Redes correspondientes a modelos: AlpA: N370_22341G y S370_1071MGA. VacAp55Col: I261_K321_1071MGA, I261_Y321_26093IGA y V261_R321_22360AG. El valor p se obtuvo mediante la aplicación de la prueba Mann Whitney para datos no pareados ($p > 0.05$).

Con el fin de establecer si existió una relación evolutiva entre el reconocimiento inmune, la selección natural y la topología de la red, se mapearon los residuos más significativos para cada parámetro topológico, contra las regiones de reconocimiento de epítomos y contra las que están bajo selección positiva o purificadora para cada proteína, ver tabla 7. Todos los residuos para cada parámetro topológico estudiado se ubicaron en regiones de los dos factores de patogenicidad con potencial de ser reconocidos por el sistema inmune. Específicamente, las predicciones indican que podrían ser reconocidos por péptidos del CMH II. Dos residuos, P303 de la variante N370_22341G de AlpA y E246 de la variante I261_K321_1071MGA de VacAp55col topológicamente esenciales según el parámetro *Ci* también se encontraron bajo

selección positiva; estos residuos fueron los más interconectados en las redes de interacción para cada variante. Sin embargo, se observó que la selección purificadora fue dominante para seleccionar las variantes topológicas en las poblaciones de aislamientos colombianos, ver tabla 7.

Tabla 2. Relación entre los sitios topológicos esenciales con tipo de selección natural y predicción inmune.

Ptop	Residuos	Modelo	Sel. Purificadora	Sel. positiva	Epitopes
k	R110	N370_22341G	2.54E-06	ns	CMH tipo II
	Y373	S370_1071MGA	ns	ns	CMH tipo II
	D216	I261_K321_1071MGA	0.0008	ns	CMH tipo II
		I261_Y321_26093IGA	ns	ns	ns
	R111, N109	V261_R321_22360AG	6.94212E-06, 0.01	ns	CMH tipo II
Ci	T418	N370_22341G	ns	ns	CMH tipo II
	R181	S370_1071MGA	1.71E-07	ns	CMH tipo II
	F240	I261_K321_1071MGA	1.65E-07	ns	CMH tipo II
	T237	I261_Y321_26093IGA	ns	ns	CMH tipo II
	R212	V261_R321_22360AG	ns	ns	CMH tipo II
Cli	A95, S463, Y461, N317** , P303* , N263	N370_22341G	1.49873E-05**	0.005*	CMH tipo II
	N352** , G8** , V278** , G86, 288, G168, T125, N126, N352, P467, E18, P7; (L198, L199, G168, G324, V225, R305, D281, V278, G294, 2199, G355, N361, N357, E246*)	S370_1071MGA	0.0002**, 1.25433E-11**, 0.00001**	ns	Lineal tipo B; (CMH tipo II)
	G294, 2199, G355, N361, N357, E246*	I261_K321_1071MGA	ns	0.03*	CMH tipo II
	G294, 2199, G355, N361 y N357	I261_Y321_26093IGA	ns	ns	CMH tipo II
	P244** , D319** , G344** , S341** , G78** , G198, S388	V261_R321_22360AG	0.002**, 6.90141E-07**, 4.1321E-05**, 0.00004**, 3.42663E-05**	ns	CMH tipo II

Ptop: parámetros topológicos: grado (k), coeficiente de agrupamiento (Ci), intermediación (Bi) y cercanía (Cli). **son residuos de un grupo topológico que están bajo selección purificadora y además bajo reconocimiento inmune al mismo tiempo, lo valores p se dan de acuerdo al orden de aparición de los residuos resaltados en negrita. *son residuos bajo selección positiva que además están bajo posible reconocimiento inmune. Los valores mostrados para los tipos de selección corresponden a los valores p determinados por el algoritmo FEL.

Otra característica identificada en algunos de los residuos considerados como ejes topológicos esenciales para las redes, fue su utilidad para diferenciar las patologías y los referentes poblacionales estudiados. Las propiedades topológicas de AlpA mostraron ser diferentes no solo entre distintas patologías, sino también con respecto a las referencias. Por ejemplo, el parámetro *k* logró diferenciar los grupos gastritis y metaplasia intestinal y gastritis atrófica de adenocarcinoma gástrico. Sin embargo, este patrón también se observó con otros residuos identificados desde los otros parámetros topológicos estudiados, ver tabla 8. Este

resultado es importante en la medida que correlacionó la dinámica topológica de la red de interacción de residuos de la adhesina con la diferenciación entre grupos por patologías.

Tabla 8. Utilidad de los sitios topológicos esenciales para la diferenciación de grupos de patología gástrica.

Ptop	Residuos	Modelo	Gdif	Correlación*	Valor p
k	R110	N370_22341G	1 y 3	0.9986	6.77E-06
			2 y 4	0.7256	
Ci	R181	S370_1071MGA	2 y 5	1	0.04471
			3 y 4	1	
			4 y 5	0.9998	
			2 y 4	0.9993	
			3 y 5	0.9987	
			1 y 3	0.996	
			2 y 3	0.996	
			1 y 5	0.955	
			1 y 2	0.8909	
			1 y 4	0.7987	
Cli	N126	S370_1071MGA	1 y 2	1	1.26E-04
			1 y 3	1	
			2 y 3	1	
			4 y 5	0.9987	
			3 y 5	0.8982	
			2 y 5	0.8495	
			3 y 4	0.8054	
			2 y 4	0.7256	

Ptop: parámetros topológicos: grado (k), coeficiente de agrupamiento (Ci), intermediación (Bi) y cercanía (Cli). Las secuencias fueron alineadas con el programa Muscle y las columnas significativamente diferentes entre grupos se determinaron mediante una chi cuadrada de Person. Grupos: 1- Gastritis, 2- Gastritis atrófica, 3- Metaplasia intestinal, 4- Adenocarcinoma gástrico y 5- Referentes. Solo la proteína AlpA mostró resultados significativos. Los análisis se llevaron a cabo en usando el servidor Meta-CATS.

8. Discusión

H. pylori ha acompañado a la humanidad desde hace aproximadamente 116.000 años, antes de su salida de África y ha estado en un constante proceso coevolutivo con su hospedero desde entonces (Moodley et al. 2012). La bacteria ha acompañado las migraciones realizadas por el hombre, inicialmente fuera de África y luego al resto del planeta y, es tan estrecha la relación con su hospedero que los análisis por MLST de la bacteria sirven para trazar rutas migratorias humanas tanto ancestrales como recientes (Backert et al., 2016). Es decir, la bacteria refleja en su genoma los cambios geográficos de su hospedero.

La microevolución de *H. pylori* depende estrechamente de lo que ocurre con la ecología de su hospedero: si este explora nuevos territorios, se aísla geográficamente y se establece en grupos, la bacteria generará cambios adaptativos que terminarán en la evolución de linajes discretos, independientes y locales que imitarán el proceso de asentamiento de su hospedero en un proceso coevolutivo (Moodley et al. 2012). Recientemente, este fenómeno se ha descrito en poblaciones de África, el este asiático (Backert et al., 2016), Arabia (Kumar et al., 2017), Malasia (Kumar et al., 2015), México, Nicaragua (Thorell et al., 2017) y ahora en Colombia (Gutiérrez-Escobar AJ et al., 2017).

Este proceso coevolutivo está regulado por la capacidad patogénica de la bacteria y por la respuesta inmune del hospedero. Durante la fase aguda de la infección, la bacteria pasa por un intenso estrés selectivo debido a la inmunidad del hospedero. Como respuesta se produce un estallido mutacional, en donde solo las variantes poblacionales de la bacteria más fuertes sobreviven y logran establecer una infección crónica que acompañará a su hospedero a lo largo

de su vida (Linz et al., 2014b). Mientras la bacteria permanezca en su hospedero en forma crónica, realizará rearrreglos genómicos para mantener su estado coevolutivo, pero si entra en contacto con un nuevo hospedero, como por ejemplo una nueva población, se dispararán estallidos mutacionales mediante selección positiva en adhesinas y proteínas que están en contacto con el sistema inmune del hospedero, que le confieran ventajas reproductivas a la bacteria y mediante selección purificadora se eliminarán los alelos que le resten ventaja selectiva a las variantes poblacionales (Linz et al., 2013). De esta manera, es como las poblaciones de la bacteria han podido reflejar la distribución poblacional humana y se han establecido sus diferentes poblaciones a nivel global (Backert et al., 2016; Moodley et al. 2012).

El continente americano fue poblado por primera vez hace aproximadamente 15.000, años al final del último periodo glacial, por personas que migraron desde el norte de Asia y cruzaron por el estrecho de Bering hasta Alaska. Fue así como surgió el haplotipo Amerindio en el humano y la subpoblación HspAmerind de la bacteria a lo largo del territorio (Cavalli-Sforza et al., 1994; Dillehay, 2009; Goebel et al., 2008; Meltzer, 2009). Sin embargo, la ancestría humana y de *Helicobacter pylori* fue modificada profundamente en el continente americano, gracias a la colonización europea. Los colonizadores ibéricos arribaron al continente en misiones comandadas por Cristóbal Colón que atracaron en las Bahamas en 1492 y trajeron principalmente inmigrantes españoles y portugueses (Fernández – Armesto, 2003; Kamen, 2002; McAlister, 1984). Se cree que en el momento de su arribo la población de nativos americanos (haplotipo Amerindio) era aproximadamente de 100 millones de habitantes (Denevan, 1992; Sanchez-Albornoz, 1974; Thornton, 1987).

Al comienzo de la colonización, el 80% de los migrantes fueron hombres (Boyd-Bowman, 1976; Ribeiro, 1995) y el proceso se caracterizó primero, por la brutalización y exterminio masivo (cerca del 95%) de la población nativa (Kamen, 2002; Morner, 1967; Montenegro, 2006; Thornton, 1987); segundo, por el extensivo mestizaje entre los ibéricos y las mujeres nativo americanas (Moreno-Estrada et al., 2013; Burkholder y Johnson, 2003; Fernández – Armesto, 2003; Sánchez-Albornoz, 1974; Lavrin, 1989; Adhikari et al., 2017); y tercero, por el tráfico de esclavos que también entró a ser parte importante del proceso de mestizaje a medida que la población nativa colapsaba (Curtin, 1969; Thomas, 1997). Se sabe que, durante este periodo, cerca de 10 millones de africanos llegaron al continente americano (Curtin, 1969; Thomas, 1997).

En Colombia, Ossa y colaboradores en 2016 (Ossa et al., 2016) encontraron que actualmente la ancestría humana predominante en el Caribe, los Andes y la Orinoquía es de tipo europeo, mientras que en las regiones Pacífico y Amazonia predominan los tipos africano y nativo americano. La región andina es la que más ancestría de tipo europeo presenta, aunque también tiene un porcentaje nativo americano importante. La región andina y la región Amazónica fueron las que menor ancestría de tipo africano mostraron. La región de la Orinoquía mostró un patrón de ancestría similar a la región andina, aunque su componente africano y nativo americano fue un poco más elevado. La alta ancestría de tipo europeo en los Andes se debió a que, durante la colonización, el río Magdalena, ubicado entre las cordilleras central y occidental, sirvió como ruta principal de transporte al interior del país (Ossa et al., 2016).

Desde una perspectiva genómica poblacional, el proceso de mestizaje se dio entre poblaciones humanas que evolucionaron de forma independiente por miles de años en Europa, Africa y América (Bryc et al., 2010; Montinaro et al., 2015; Ruiz-Linares et al., 2014; Wang et al., 2008). Durante este tiempo cada población humana se adaptó a condiciones medioambientales localmente específicas (Grossman et al., 2013; Sabeti et al., 2006) y se generaron secuencias genómicas novedosas con haplotipos combinados que nunca coexistieron previamente en la historia humana (Jordan IK, 2016).

Antes de la colonización, existía una subpoblación de *H. pylori* en la población nativa americana denominada hspAmerind, que evolucionó a partir de las cepas del norte de Asia (O'Rourke, 2009). Esta subpoblación ha sido aislada de nativos americanos en el Perú (Kersulyte et al., 2010), México, Venezuela y Colombia (Camorlinga-Ponce et al., 2011), que aún viven en aislamiento geográfico. Sin embargo, debido al intenso proceso de mestizaje y a la casi desaparición de su hospedero nativo en zonas urbanas, esta subpoblación fue prácticamente reemplazada por una del tipo HpEurope (de ST et al., 2011; Kodaman et al., 2014; Shiota et al., 2014; Scinschi et al., 2010; Thorell et al., 2016; Yamaoka et al., 2002). Se sugiere que cepas del tipo HpEurope coevolucionaron acompañando el proceso de mestizaje y así, se generaron los linajes independientes de *Helicobacter pylori*, identificados en el presente trabajo y otros (Thorell et al., 2017; Gutierrez-Escobar et al., 2017).

Desde una mirada histórico-evolutiva, es posible plantear que mientras los grupos poblacionales humanos se mezclaban por el proceso colonial y posterior a él, las cepas tipo HpEurope se recombinaron, intercambiaron porciones de genes, compitieron entre ellas y además, realizaron su respectivo estallido mutacional, coevolucionado con su nuevo

hospedero mestizo en los diferentes países de América Latina (Linz et al., 2014b; Linz et al., 2013; Muñoz-Ramírez et al., 2017). Estas cepas recombinantes debieron adaptarse a condiciones medioambientales típicas del territorio tropical, como son períodos de lluvia-sequía, gradientes altitudinales abruptos, patrones de alimentación, una microbiota nueva y un sistema inmune con una nueva carga de polimorfismos, que indujeron la microevolución y adaptación de estos nuevos linajes de *H. pylori* en mestizos de la zona estudiada (Martínez Palomino et al., 2008; Martínez et al., 2011; Yang et al., 2016; Torres et al., 2013, Adamsson et al., 2013; Gutiérrez-Escobar et al., 2017).

El análisis del pangenoma de los aislamientos secuenciados en la zona de estudio no solo evidenció la presencia de linajes independientes de la bacteria, sino que también aportó nuevos datos sobre el contenido de genes para *H. pylori*, apoyando lo previamente mencionado. Un estudio reportó que la bacteria contiene 1193 familias de genes conservados a partir de 39 genomas completos estudiados (Ali et al., 2015). Sin embargo, en el presente estudio se empleó una técnica para el análisis del pangenoma de la bacteria con mayor poder de resolución que permitió identificar nuevas capas de genes conservados (Contreras-Moreira y Vinuesa, 2013).

A partir de 36 genomas completos de la bacteria se logró identificar que el core estricto de la especie está constituido por 820 clústeres o grupos de genes, y el softcore por 1093; siendo este último valor comparable con los datos reportados en (Ali et al., 2015). Sin embargo, se reporta por primera vez, un valor para el Shell de 848 clústeres y para el cloud de 2395, ambos compartimentos representando genes accesorios de la bacteria. Cuando el pangenoma de todas las cepas y aislamientos fueron analizados (136 genomas en total) se observó una

disminución en 283 genes en el core estricto y de 154 genes en el softcore. Pero, por otro lado, se observó un aumento de 499 genes para el Shell y de 2339 para el cloud. Lo que sugiere que la recombinación, la conversión génica y la evolución concertada (Santoyo & Romero, 2005) han jugado un papel importante en la microevolución rápida de familias de genes que se han adaptado localmente a las variaciones genéticas del hospedero mestizo.

H. pylori es una especie con características idóneas para generar este tipo de cambios adaptativos y microevolutivos a escala local. Posee gran diversidad genética (Sawyer, 1989), es naturalmente competente, presenta una altísima tasa de recombinación y mutación (Sawyer, 1989); de hecho, podría ser capaz de intercambiar la mitad de su genoma por medio de recombinación en un periodo de 40 años (Falush et al., 2001) y genera estallidos mutacionales comparables a los observados en virus (Linz et al., 2014b). Esta plasticidad genómica, le permite a la bacteria generar nuevos subtipos poblacionales rápidamente como los que se han reportado en Senegal, Madagascar (Linz et al., 2014a), Malasia (Kumar et al., 2015), Arabia (Kumar et al., 2017) y Colombia. Esto significa que no solamente fue el paisaje genético y la ancestría poblacional de los humanos la que cambió, sino también los de la bacteria, la cual experimentó un intenso proceso de mezcla genética durante y después de la conquista (de ST et al., 2011; Kodaman et al., 2014).

Los análisis filogenéticos y poblacionales efectuados en los factores de virulencia, mostraron un patrón microevolutivo que permitió dar soporte a la hipótesis de los linajes independientes en la zona de estudio, de hecho, mostraron claramente las diferencias preconquista y postconquista de América. Las filogenias usando solo secuencias de los referentes poblacionales, mostraron en menor o mayor grado una distribución bilobular o trilobular,

separando el grupo entre cepas del este y cepas del oeste, posiblemente un reflejo de la época de la preconquista. Este tipo de segregación poblacional ya ha sido reportada previamente para otros genes de la especie, en los que no se han empleado secuencias de aislamientos similares a los del presente estudio (Cao P et al., 2005; Maeda S et al., 1998; Oleastro M et al., 2009a; Van Doorn LJ et al., 1999).

Al incluir las secuencias de los aislamientos colombianos en todos los análisis filogenéticos para cada factor de virulencia, se observó que la topología de los árboles estuvo influenciada probablemente por la recombinación y la conversión génica, lo cual concuerda con la postconquista, en la que entraron en contacto poblaciones, tanto humanas como de las bacterias que llevaban miles de años divergiendo en forma individual. Una consecuencia directa de la postconquista y el mestizaje, fue la inducción de eventos de divergencia funcional y diversificación, entre los miembros parálogos/homólogos para cada factor de virulencia estudiado (Santoyo G and Romero D, 2005) y, por ende, la generación de variantes locales con ritmos de evolución diferenciales.

La recombinación, la conversión y la diversificación funcional, son tres procesos importantes para la inducción de eventos microevolutivos en *H. pylori*. Se han reportado previamente procesos evolutivos mediados por recombinación genética para la bacteria, por ejemplo, en las adhesinas *homB* y *homA* (Dorer et al., 2013; Dubois et al., 1999; Oleastro et al., 2009a; Oleastro et al., 2009b; Yamaoka et al., 2002), *sabB* y *omp27* (Talarico et al., 2012), y *babA* (Hennig et al., 2006, Thorell et al., 2016). Y también, eventos de diversificación alélica generadores de innovación biológica (Ohno, 1970; Zhang, 2003; Lynch y Conery, 2000), como por ejemplo, los observados en el gen *vacA*, que codifica para la citotoxina

vacuolizante de *Helicobacter pylori*, en donde se han reportado tanto mutantes no funcionales (Ito et al., 1998), así como eventos de innovación funcional, al expandir su capacidad de reconocer receptores en la célula hospedera como por ejemplo: LRP1 (Yahiro et al., 2012), EGFR (Seto et al., 1998), la esfingomielina (Utt et al., 2001; Gupta et al., 2008), glicosfingolípidos (Gupta et al., 2008) y herapan sulfato (Roche et al., 2007), mostrando novedad evolutiva después de un evento de diversificación, como una señal de adaptación funcional al sistema inmune del hospedero (Dubois A et al., 1999; Oleastro M et al., 2009b).

Los alelos codificantes para los factores de virulencia estudiados en la zona de estudio, mostraron claras señales de recombinación, conversión genética y diversificación funcional. Es así como en todas las filogenias se observaron puntos de divergencia para cada proteína en forma de clados bien diferenciados. Por ejemplo, para AlpA se observaron tres clados (Col1, Col2 y Col3), para HorB cinco clados, para Tip α tres clados y para VacAp55Col tres clados (Ch, m1 y m2). Estas topologías sugirieron que en la zona de estudio ocurrió un proceso selectivo de naturaleza divergente de nacimiento y muerte de genes (Nei y Rooney, 2005) para estos factores de patogenicidad.

Seguidamente, los valores negativos para la prueba Tajima sugieren que la población sufrió una rápida expansión después de un cuello de botella y que existen alelos raros con alta frecuencia (Tajima, 1989). Este panorama es coherente con dos planteamientos que se basan en los resultados de este estudio. El primero, a escala filogenómica que apunta al proceso coevolutivo que sufrió la bacteria con su nuevo hospedero mestizo; y el segundo, a nivel filogenético y topológico para los factores de patogenicidad, que sugiere el papel de la selección inmune. Aunque es claro que los genes evolucionan de forma diferencial y no

reflejan la distribución poblacional de la bacteria (Yahara et al., 2012; Yahara et al., 2013), sí es posible sugerir a partir de los análisis evolutivos, que las poblaciones de la bacteria en la zona de estudio surgieron a partir de una población pequeña que paulatinamente se expandió hasta dar origen a líneas evolutivas independientes con altos niveles de adaptación local y diferenciación genética.

Es probable que en la zona de estudio exista un potencial microevolutivo importante para la bacteria. Esto se puede apreciar en el alto número de haplotipos encontrados en los aislamientos colombianos, además, en su extrema diversidad, lo que significa que sí se incluyen más secuencias de estos genes en la filogenia, es probable que se observen más cladogramas debido a una mejora en la resolución filogenética. Como se mencionó previamente, la recombinación y la conversión génica son procesos que también contribuyen al potencial evolutivo en la zona de estudio. Se observaron eventos de recombinación y conversión génica similares entre los aislamientos colombianos y la referencia. Esto significa que los haplotipos de los subtipos colombianos de la bacteria en la zona de estudio se han mezclado localmente contribuyendo al flujo genético, y por tanto, contribuyendo a la ganancia de diversidad alélica y a la consecuente diferenciación poblacional de los linajes bacterianos (Israel et al., 2000; Pinto et al., 2005).

Otra línea de evidencia que soportó los procesos evolutivos locales sugeridos en este estudio fue la diferenciación genética observada entre los aislamientos colombianos y la referencia. Teniendo en cuenta que no se contó con una muestra de miles de secuencias, los métodos para determinar la diferenciación genética basados en frecuencias haplotípicas, no tendrían suficiente poder de discriminación debido a que muchos haplotipos solo aparecerían una sola

vez. Para evitar este sesgo, se usó la prueba de diferenciación genética *Snn* (*nearest-neighbor Statistics*) que emplea el número de diferencias nucleotídicas entre haplotipos para determinar el espacio de secuencias y determinar las diferencias entre ellas. Esta prueba fue estadísticamente significativa ($p > 0.001$) para todos los genes, indicando que los aislamientos colombianos están diferenciados con respecto a la referencia en la zona de estudio, aunque conservando el flujo genético.

Los factores de virulencia de las poblaciones estudiadas en la zona, también mostraron evidencia de regiones con valores superiores a 1 en el análisis Ka/Ks, eventos de divergencia funcional, selección positiva y purificadora. La divergencia funcional y la selección positiva, ocurren en una proteína después de un proceso evolutivo mayor, como lo es la diversificación génica e indica que algunos de los residuos en las proteínas presentan ritmos de cambio evolutivo diferenciales entre los clados de una filogenia y algunos de ellos confieren ventajas reproductivas (Gu, 1999, 2001). Estos resultados sugieren que la diferenciación funcional observada en las variantes alélicas de la zona de estudio, se debió en parte a un proceso de adaptación que favoreció la selección de alelos beneficiosos para los aislamientos mediante selección positiva (Wertheim JO et al., 2015) y por otra, a eventos de selección purificadora (Lynch M, Conery JS, 2000) que purgaron los alelos menos beneficiosos de la población, siguiendo precisamente el proceso coevolutivo descrito por Linz y colaboradores (Linz et al., 2013) en el hospedero mestizo de la zona de estudio.

Definir por qué ocurren estos cambios evolutivos en proteínas es una tarea muy compleja (Echave et al., 2016); sin embargo, los análisis inmunoinformáticos realizados en el presente trabajo podrían, servir como guía para describir los procesos evolutivos de los factores de

patogenicidad estudiados. Del total de sitios identificados bajo divergencia funcional y bajo selección positiva el 65.9% se ubicó en epítomos tipo B y tipo T. Se podría sugerir que en la zona de estudio se puede estar presentando un bucle selectivo, en el que por un lado el sistema inmune del hospedero induce la evolución adaptativa observada en los factores de patogenicidad (Aguilera et al, 2009) y por el otro, es posible que la respuesta inmune del hospedero mestizo se intensifique en respuesta a la presencia de nuevas variantes alélicas. Lo anterior podría estar intensificando el proceso inflamatorio del hospedero en la zona (Tu et al., 2008; Sagaert et al., 2010; Herrera y Parsonnet, 2009). Algo muy notable de la población mestiza es que a nivel interespecífico sus ancestrías son altamente variables, es decir, no existen dos poblaciones mestizas con los mismos porcentajes de ancestría en Latinoamérica (Bryc et al., 2010). Lo que se sugiere en el presente estudio es que los estallidos mutacionales mediante selección positiva en factores de virulencia, deberían ser más frecuentes en poblaciones de la bacteria que coevolucionan con un hospedero mestizo, que con un hospedero de haplotipo simple, ya que cada vez que la bacteria se trasmite, debe adaptarse a un hospedero con ancestría variable.

H. pylori se transmite en forma vertical, sin embargo, mucha evidencia sugiere que la bacteria también se trasmite de forma horizontal con bastante frecuencia, especialmente en países en vías de desarrollo (Delport et al, 2006; Schwarz et al., 2008; Vale y Vitor, 2010). Si la bacteria se trasmite de forma estrictamente vertical, la evidencia presentada sobre eventos de divergencia funcional y selección positiva para los factores de patogenicidad, es contradictoria porque cuando la transmisión es vertical se supone que el patógeno ha coevolucionado con el hospedero para no dañarlo (Week et al, 2007), lo cual al menos en la zona de estudio no es cierto; mientras que, sí el patógeno se trasmite de forma horizontal, sí

existe una explicación evolutiva para el incremento de patogenicidad y lesión al hospedero (Lenski y May, 1994). Esta idea también refuerza el bucle selectivo propuesto para la zona de estudio.

Una pregunta evolutiva fundamental con respecto a *Helicobacter pylori*, es si existe algún factor selectivo que permita diferenciar las cepas por grupos o tipos de patología gástrica (Comunicación personal con Ichizo Kobayashi). Aunque esta pregunta debe ser resuelta desde una mirada transdisciplinaria, los resultados del presente trabajo mostraron que la carga mutacional a escala proteica para dos de los cuatro factores de virulencia estudiados fue diferente entre los grupos de patologías gástricas. De forma interesante, se identificó que el residuo N370S/V para la proteína AlpA fue significativamente diferente ($p < 0,04$) entre los grupos gastritis/gastritis atrófica, gastritis/metaplasia intestinal, metaplasia intestinal/adenocarcinoma gástrico y gastritis atrófica/metaplasia intestinal. Y para VacAp55Col se identificaron dos residuos, el primero I261V ($p < 0,002$) que fue diferente entre los grupos gastritis/metaplasia intestinal, gastritis/adenocarcinoma gástrico, metaplasia intestinal/adenocarcinoma gástrico. Y el segundo sitio, K321Y/R ($p < 0,03$) diferente entre los grupos gastritis atrófica/metaplasia intestinal, gastritis/metaplasia intestinal y gastritis/gastritis atrófica. Sugiriendo, que posiblemente los eventos de divergencia funcional y selectivos observados para estos dos factores de virulencia puedan pertenecer a un haplotipo nuevo que quizás se relacione con las patologías en la zona de estudio.

Desde una perspectiva evolutiva, fue muy significativo encontrar el residuo Y321 del dominio VacAp55Col como parte de esta carga mutacional diferencial entre patologías, ya que, al estar bajo un patrón de selección positiva, significa que este mutante puede estar

confiriendo una ventaja selectiva a las variantes que lo tengan y que quizás, se relacione con el mejoramiento de la capacidad patogénica de la bacteria. En la misma línea de razonamiento, se encontró que el residuo V65 de AlpA identificado tanto bajo divergencia funcional como bajo selección positiva, puede estar relacionado con un evento de selección episódica y divergente en la zona de estudio, lo cual podría ser un indicativo de innovación funcional para esta adhesina, y quizás exploración de nuevos receptores en la célula hospedera. Es importante enfatizar que no se está proponiendo que estos residuos se asocian clínicamente con la diferenciación de las patologías gástricas, sólo se proponen como componentes de una carga mutacional, que sí es diferente entre las secuencias obtenidas a partir de pacientes con diferentes patologías gástricas en la zona de estudio, por lo tanto, se limitará con proponer que al ser nuevos hallazgos, sería importante explorar la utilidad de estas variantes alélicas, como marcadores diagnósticos en futuros estudios correlacionales de tamizaje alélico poblacional en esta zona de Colombia.

Con lo discutido hasta el momento, se podría plantear una justificación teórica suficiente sobre la existencia de linajes independientes de la bacteria en esta zona del país. Sin embargo, todos los análisis desarrollados implicaron solo información a nivel de secuencia, pero no sobre la información que se puede extraer a partir de la estructura de dichos factores. Entonces, la pregunta planteada fue, si la diferenciación observada a nivel genómico y genético también generó diferenciación topológica en la estructura de factores de patogenicidad con respecto a sus referentes globales. Para esto se reconstruyeron modelos protéicos para las proteínas VacAp55Col y AlpA de los aislamientos de la zona que fueron elegidos con base a la presencia de residuos bajo selección positiva, divergentes y los residuos identificados por diferenciar los grupos de patología.

Todas las redes de interacción de residuos estudiadas, presentaron características topológicas diferenciadas entre los controles y los aislamientos que apuntaron a que los factores de patogenicidad están diferencialmente conectados en la zona de estudio. La comparación de la distribución del parámetro k entre las redes control y los modelos para ambas proteínas de los aislamientos colombianos mostraron ser significativamente diferentes ($p < 0,05$) y de hecho, los residuos con el mayor valor para k fueron diferentes en todas las redes, con excepción del residuo R371 que se repite en dos modelos de VacAp55Col en dos aislamientos colombianos; esto es un reflejo de la plasticidad topológica y evolutiva de las proteínas estudiadas. El mismo comportamiento fue determinado para los parámetros C_i y C_{li} ($p < 0,05$), sugiriendo que tanto los clústeres de residuos, como qué tan cerca se ubica un residuo del otro en la red, también difieren entre controles y aislamientos.

Con el fin de confirmar si la conectividad de los residuos estaba implicada en la diferenciación de los aislamientos colombianos, se estudiaron otros parámetros topológicos relacionados. Dos de estos parámetros la conectividad y el coeficiente topológico, mostraron diferencias significativas entre las redes ($p < 0,05$). Estos resultados sugieren diferencias importantes entre las redes, indicando, primero que los nodos más conectados tienden a conformar grupos que se correlacionan positivamente con sus valores para el parámetro k y, el segundo ratifica que algunos residuos tienden a formar grupos o clústeres.

Algo muy significativo y sorprendente fue identificar que los residuos R110, R181 y N126 propuestos aquí como ejes topológicos esenciales para las variantes de la proteína AlpA también hacen parte de la carga mutacional que es diferente a nivel de secuencia entre los

grupos de patología gástrica. La conectividad de las redes de esta proteína se asocia con la diferenciación entre los grupos gastritis/metaplasia intestinal y gastritis atrófica/adenocarcinoma gástrico. Una vez más, no se afirma que estos residuos se asocien a la patología, ya que solo reflejan que las variantes alélicas para los factores de patogenicidad podrían tener cargas mutacionales diferenciales entre los grupos de patología. Pero sí se menciona que esta aproximación metodológica no ha sido explorada previamente para *Helicobacter pylori* y, abre una posibilidad novedosa para el desarrollo de nuevos métodos computacionales orientados al descubrimiento de marcadores diagnósticos enfocados a población mestiza.

Un análisis más detallado de los residuos más significativos para cada parámetro topológico reveló, que muchos de ellos estaban ya sea bajo selección purificadora o en zonas predichas a ser reconocidas por la inmunidad del hospedero, o inclusive en ambas. Se podría sugerir que, a este nivel, la selección purificadora elimina aquellas variantes deletéreas que no cumplan los requisitos mínimos de conectividad y asortividad por nodo para estos factores de patogenicidad y que, a su vez, la conectividad (estabilidad de la red) de las redes es lo que diferencia los linajes independientes en la zona de estudio de los referentes poblacionales. El hecho de haber encontrado que muchos de estos residuos topológicamente significativos, se encontraban en zonas predichas de reconocimiento inmune, muestra el que el proceso coevolutivo se está dando entre la bacteria y su hospedero mestizo en la zona de estudio.

Los resultados encontrados a lo largo de este estudio, pueden ser soportados en el contexto de la teoría de la reina roja (Van Valen, 1974). Esta teoría plantea que la competencia entre organismos es lo que provoca el cambio microevolutivo, ya que induce la adaptación y

contra-adaptación en un proceso coevolutivo entre las especies involucradas (Van Valen, 1973; Stenseth y Smith, 1984). Es por este motivo que se logran identificar grandes ritmos de cambio evolutivo en genes asociados con procesos infecciosos (Clark et al, 2007; Hedrick, 1994; Obbard et al., 2006; Blan et al., 2005; Mu et al., 2006; Barret et al., 2009), o también en genes involucrados en resistencia a la infección (Paterson et al., 2010) o como los detectados en el presente estudio.

Latinoamérica es considerada como un laboratorio evolutivo para *Helicobacter pylori*. La presencia de un haplotipo mestizo que presenta unos porcentajes de ancestría muy diversos, tanto interespecifica como intraespecificamente, sumado a las características genéticas y adaptativas de la bacteria, significa que es posible encontrar linajes divergentes e independientes en diferentes zonas del continente con población mestiza, lo cual tiene implicaciones biomédicas y en salud pública importantes. Este trabajo es un ejemplo de cómo la medicina y la teoría evolutiva se pueden integrar de tal manera, que no solo se obtenga una descripción histórica de los eventos evolutivos sino también, cómo su metodología permite identificar variantes alélicas que podrían ser aplicadas para posibles desarrollos clínicos.

9. Conclusiones

- Los aislamientos estudiados en la presente tesis fueron obtenidos en la ciudad de Bogotá, Tunja y pueblos aledaños, la cual, como zona montañosa es considerada una región de alto riesgo de cáncer gástrico en Colombia. En esta región se identificó un proceso microevolutivo que dio origen a linajes independientes de *Helicobacter pylori*. Se argumenta que los inductores de este proceso fueron la reciente aparición del haplotipo mestizo en América Latina como un hospedero emergente y un bucle selectivo entre la selección inmune y la generación de variantes alélicas para factores de patogenicidad de *Helicobacter pylori*, que evolucionan bajo selección positiva y purificadora
- Los residuos identificados como nodos topológicos esenciales para las redes de interacción de residuos, también son objeto de la acción de la selección natural. Estos residuos regulan la conectividad de los factores de patogenicidad estudiados y pueden ser reconocidos por la inmunidad del hospedero, lo cual abre un nuevo panorama de análisis para la identificación de variantes genéticas de importancia clínica o diagnóstica para esta infección desde una perspectiva topológica.
- Mediante la aplicación de métodos evolutivos se lograron obtener resultados de posible relevancia clínica y diagnóstica. Por primera vez se reporta que el residuo Y321 del dominio VacAP55Col, que evoluciona bajo selección positiva en los aislamientos colombianos y que los residuos R110, R181 y N126 considerados como ejes topológicos esenciales para las variantes de la proteína AlpA, hacen parte de una carga mutacional que es diferente a nivel de secuencia proteica entre grupos conformados por diferentes

patologías gástricas. Esto, abre un área de investigación que correlaciona la selección natural, la conectividad de las redes de interacción de residuos de factores de patogenicidad de *Helicobacter pylori*, con la diferenciación entre patologías gástricas humanas. Sin embargo, una vez más, como se hizo a lo largo del estudio, no se afirma que estos residuos se asocien a la patología, solo tienen la capacidad de diferenciarlas.

- Los métodos basados en K-meros libres de alineamiento producen mejor resolución filogenómica que los métodos basados en alineamiento de secuencias para realizar análisis evolutivos con secuencias de *Helicobacter pylori*.

10. Referencias

Achtman M, Azuma T, Berg DE, Ito Y, Morelli G, Pan ZJ, Suerbaum S, Thompson SA, van der Ende A, van Doorn LJ. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Molecular Microbiology* 32(3):459-70, 1999. doi: 10.1046/j.1365-2958.1999.01382.x

Adamsson J, Lundin SB, Hansson LE, Sjövall H, Svennerholm AM. Immune responses against *Helicobacter pylori* in gastric cancer patients and in risk groups for gastric cancer. *Helicobacter* 18(1):73-82, 2013. doi: 10.1111/j.1523-5378.2012.00991.x.

Adhikari K, Chacón-Duque JC, Mendoza-Revilla J, Fuentes-Guajardo M, Ruiz-Linares A. The Genetic Diversity of the Americas. *Annual Review of Genomics and Human Genetics* 18:277-296, 2017. doi: 10.1146/annurev-genom-083115-022331

Agren J, Sundstrom A, Hafstrom T, Segerman B. Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. *PLoS One* 7(6):e39107, 2012. <https://doi.org/10.1371/journal.pone.0039107>

Aguileta G, Refregier G, Yockteng R, Fournier E, Giraud T. Rapidly evolving genes in pathogens: methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. *Infection Genetics and Evolution* 9:656–70, 2009. doi: 10.1016/j.meegid.2009.03.010

Ali A, Naz A, C Soares S, Bakhtiar M, Tiwari S, S Hassan S, Hanan F, Ramos R, Pereira U, Barh D, César Pereira Figueiredo H, W. Ussery D, Miyoshi A, Silva A, and Azevedo V. Pan-Genome Analysis of Human Gastric Pathogen *H. pylori*: Comparative Genomics and Pathogenomics Approaches to Identify Regions Associated with Pathogenicity and Prediction of Potential Core Therapeutic Targets. *Biomed Res International*. 2015 (2015): 139580. doi: [10.1155/2015/139580]

Allen LA, Schlesinger LS, Kang B. Virulent strains of *Helicobacter pylori* demonstrate delayed phagocytosis and stimulate homotypic phagosome fusion in macrophages. *Journal and Experimental Medicine* 191: 115-128, 2000. doi: 10.1084/jem.191.1.115].

Alm R, Bina J, Andrews B, Doig P, Hancock RE, Trust TJ. Comparative genomics of *Helicobacter pylori*: analysis of the outer membrane protein families. *Infection and Immunity* 68(7):4155–4168, 2000

Amieva MR, El-Omar EM. Host-bacterial interactions in *Helicobacter pylori* infection. *Gastroenterology* 134(1):306-23, 2008. doi: 10.1053/j.gastro.2007.11.009.

Amilon KR, Letley DP, Winter JA, Robinson K, Atherton JC. Expression of the *Helicobacter pylori* virulence factor vacuolating cytotoxin A (*vacA*) is influenced by a potential stem-loop

structure in the 5' untranslated region of the transcript. *Molecular Microbiology* 98(5):831–846, 2015. doi:10.1111/mmi.13160.

Argent RH, Thomas RJ, Letley DP, Rittig MG, Hardie KR, Atherton JC. Functional association between the *Helicobacter pylori* virulence factors VacA and CagA. *Journal of Medical Microbiology* 57(Pt 2):145–150, 2008. doi:10.1099/jmm.0.47465-0.

Asahi M, Tanaka Y, Izumi T, Ito Y, Naiki H, Kersulyte D, Tsujikawa K, Saito M, Sada K, Yanagi S, Fujikawa A, Noda M, Itokawa Y. *Helicobacter pylori* CagA containing ITAM-like sequences localized to lipid rafts negatively regulates VacA-induced signaling in vivo. *Helicobacter* 8(1):1–14, 2003.

Atherton JC. The pathogenesis of *Helicobacter pylori*-induced gastro-duodenal diseases. *Annual Review Pathology* 1:63–96, 2006. doi:10.1146/annurev.pathol.1.110304.100125.

Atherton JC, Blaser MJ. Coadaptation of *Helicobacter pylori* and humans: ancient history, modern implications. *J Clin Invest* 119(9):2475–2487, 2009. doi:10.1172/jci38605.

Atherton JC, Cao P, Peek RM Jr, Tummuru MK, Blaser MJ, Cover TL. Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. Association of specific vacA types with cytotoxin production and peptic ulceration. *Journal of Biological Chemistry* 270(30):17771–17777, 1995.

Backert S, Yamaoka Y. *Helicobacter pylori* Research: From Bench to Bedside. 1st ed. Springer. Japan, p 606, 2016. doi: 10.1007/978-4-431-55936-8

Baily SL, Míguez EJ. MassMigration to Modern Latin America. Wilmington, Delaware, A Scholarly Resources Inc. Jaguar books on Latin America N° 24, pp. 293. 2003.

Barrett LG1, Thrall PH, Dodds PN, van der Merwe M, Linde CC, Lawrence GJ, Burdon JJ. Diversity and evolution of effector loci in natural populations of the plant pathogen *Melampsora lini*. *Molecular Biology and Evolution* 26, 2499–2513, 2009. doi: 10.1093/molbev/msp166.

Berg, D. E., P. S. Hoffman, B. J. Appelmelk, J. G. Kusters. The *Helicobacter pylori* genome sequence: genetic factors for long life in the gastric mucosa. *Trends in Microbiology* 5:468–474, 1997.

Bergman MP, Engering A, Smits HH, van Vliet SJ, van Bodegraven AA, Wirth HP, Kapsenberg ML, Vandenbroucke- Grauls CM, van Kooyk Y, Appelmelk BJ. *Helicobacter pylori* modulates the T helper cell 1/T helper cell 2 balance through phase-variable interaction between lipopolysaccharide and DC-SIGN. *Journal of Experimental Medicine* 200: 979-990, 2004. [PMID: 15492123 DOI: 10.1084/jem.20041061].

Bianchine PJ, Russo TA. The role of epidemic infectious diseases in the discovery of America. *Allergy and Asthma Proceedings* 13(5):225-232, 1992. doi: https://doi.org/10.2500/108854192778817040

Bina J, Bains M, Hancock RE. Functional expression in *Escherichia coli* and membrane topology of porin HopE, a member of a large family of conserved proteins in *Helicobacter pylori*. *Journal of Bacteriology* 182(9):2370-2375, 2000.

Blanc G, Ngwamidiba M, Ogata H, Fournier PE, Claverie JM, Raoult D. Molecular evolution of *Rickettsia* surface antigens: evidence of positive selection. *Molecular Biology and Evolution* 22, 2073–2083, 2005. doi: 10.1093/molbev/msi199

Bloom JD, Drummond DA, Arnold FH, Wilke CO. Structural determinants of the rate of protein evolution in yeast. *Molecular Biology and Evolution* 23, 1751–1761, 2006.

Bode C, Kovacs IA, Szalay MS, Palotai R, Korcsmaros T, Csermely P. Network analysis of protein dynamics. *FEBS Letters* 581, 2776–2782, 2007.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27(4):578-9, 2011. doi: 10.1093/bioinformatics/btq683.

Boncrisiano M, Paccani SR, Barone S, Ulivieri C, Patrussi L, Ilver D, Amedei A, D’Elios MM, Telford JL, Baldari CT. The *Helicobacter pylori* vacuolating toxin inhibits T cell activation by two independent mechanisms. *Journal of Experimental Medicine* 198: 1887-1897, 2003. [PMID: 14676300 DOI: 10.1084/jem.20030621]

Boyd-Bowman P. Patterns of Spanish emigration to the Indies until 1600. *The Hispanic American Historical Review* 66:580–604, 1976. doi: 10.2307/2514194

Bravo LE, Cortés A, Carrascal E, Jaramillo R, García LS, Bravo PE, Badel A, Bravo PA. *Helicobacter pylori*: patología y prevalencia en biopsias gástricas en Colombia. *Colombia Médica* 34 (3): 1-9, 2003.

Bryc K, Velez C, Karafet T, Moreno-Estrada A, Reynolds A, Auton A, Hammer M, Bustamante CD, Ostrer H. Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proceedings of the National Academy of Sciences of the United States of America* 107 Suppl 2:8954–61, 2010. doi:10.1073/pnas.0914618107.

Burkholder MA, Johnson LL. Colonial Latin America. Oxford, UK: Oxford University Press. 2003.

Bustamante CD, Townsend JP, Hartl DL. Solvent accessibility and purifying selection within proteins of *Escherichia coli* and *Salmonella enterica*. *Molecular Biology and Evolution* 17, 301–308, 2000.

Camorlinga-Ponce M, Perez-Perez G, Gonzalez-Valencia G, Mendoza I, Penaloza-Espinosa R, Ramos I, Kersulyte D, Reyes-Leon D, Romo C, Granados J, Muñoz L, Berg DE, Torres J. *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, 2011. <https://doi.org/10.1371/journal.pone.0027212>

Cao DM, Lu QF, Li SB, Wang JP, Chen YL, Huang YQ, Bi HK. Comparative Genomics of *H. pylori* and Non-Pylori Helicobacter Species to Identify New Regions Associated with Its Pathogenicity and Adaptability. *BioMed Research International* 2016:6106029, 2016. doi: 10.1155/2016/6106029

Cao P, Lee KJ, Blaser MJ, Cover TL. Analysis of hopQ alleles in East Asian and Western strains of *Helicobacter pylori*. *FEMS Microbiology Letters* 251(1): 37-43, 2005. doi: 10.1016/j.femsle.2005.07.023.

Cao Q, Didelot X, Wu Z, Li Z, He L, Li Y, Ni M, You Y, Lin X, Li Z, Gong Y, Zheng M, Zhang M, Liu J, Wang W, Bo X, Falush D, Wang S, Zhang J. Progressive genomic convergence of two *Helicobacter pylori* strains during mixed infection of a patient with chronic gastritis. *Gut* 64(4):554-61, 2015. doi: 10.1136/gutjnl-2014-307345

Charlebois RL, Doolittle WF. Computing prokaryotic gene ubiquity: rescuing the core from extinction. *Genome Research* 14:2469–2477, 2004.

Cavalli-Sforza LL, Menozzi P, Piazza A. The History and Geography of Human Genes. Princeton NJ: Princeton University Press.1994.

Celli JP, Turner BS, Afdhal NH, Keates S, Ghiran I, Kelly CP, Ewoldt RH, McKinley GH, So P, Erramilli S, Bansil R. *Helicobacter pylori* moves through mucus by reducing mucin viscoelasticity. *Proceedings of the National Academy of Sciences of the United States of America* 106(34):14321–14326, 2009. doi:10.1073/pnas.0903438106.

Claeys D, Faller G, Appelmelk BJ, Negrini R, Kirchner T. The gastric H⁺, K⁺-ATPase is a major autoantigen in chronic *Helicobacter pylori* gastritis with body mucosa atrophy. *Gastroenterology*; 115: 340-347, 1998.

Clark AG et al. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450, 203–218, 2007. doi: 10.1038/nature06341

Cohan, F. M. The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. *The American Naturalist*. 143:965-986, 1994a. <https://doi.org/10.1086/285644>

Cohan, F. M. Genetic exchange y evolutionary divergence in prokaryotes. *Trends in Ecology & Evolution* 9:175-180, 1994b.

Coil D, Jospin G, Darling AE. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics*15; 31(4):587-9, 2015.

Colquhoun A, Arnold M, Ferlay J, Goodman KJ, Forman D, Soerjomataram I. Global patterns of cardia and non-cardia gastric cancer incidence in 2012. *Gut* 64(12):1881–1888, 2015. doi:10.1136/gutjnl-2014-308915.

Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Applied and Environmental Microbiology* 7 (24):7696-701, 2013.

Cooper AM. IL-17 and anti-bacterial immunity: protection versus tissue damage. *European Journal of Immunology*; 39: 649-652, 2009. Doi: 10.1002/eji.200839090

Correa P, Cuello C, Duque E. Carcinoma and intestinal metaplasia of the stomach in Colombian migrants. *Journal of the National Cancer Institute*; 44: 297–306, 1970.

Correa P, Haenszel W, Cuello C, Tannenbaum S, Archer M. A model for gastric cancer epidemiology. *The Lancet* 2: 58–60, 1975.

Correa P, Piazuelo MB. The gastric precancerous cascade. *Journal of Digestive Diseases* 13(1):2-9, 2012. doi: 10.1111/j.1751-2980.2011.00550.x.

Cover TL, Blaser MJ. Helicobacter pylori in health and disease. *Gastroenterology* 136(6):1863-73, 2009.

Cua DJ, Tato CM. Innate IL-17-producing cells: the sentinels of the immune system. *Nature Reviews Immunology*; 10: 479-489, 2010. [PMID: 20559326 DOI: 10.1038/nri2800].

De Sablet T, Piazuelo MB, Shaffer CL, Schneider BG, Asim M, Chaturvedi R, Bravo LE, Sicinski LA, Delgado AG, Mera RM, Israel DA, Romero-Gallo J, Peek RM Jr, Cover TL, Correa P, Wilson KT. Phylogeographic origin of Helicobacter pylori is a determinant of gastric cancer risk. *Gut* 60(9):1189-95, 2011. doi: 10.1136/gut.2010.234468

D'Elis MM, Manghetti M, De Carli M, Costa F, Baldari CT, Burroni D, Telford JL, Romagnani S, Del Prete G. T helper1 effector cells specific for Helicobacter pylori in the gastric antrum of patients with peptic ulcer disease. *The Journal of Immunology* 158: 962-967, 1997.

Delpont W, Cunningham M, Olivier B, Preisig O, van der Merwe SW. A population genetics pedigree perspective on the transmission of Helicobacter pylori. *Genetics* 174:2107–18, 2006. doi: 10.1534/genetics.106.057703

Denevan WM. The Native Population of the Americas in 1492. Madison: *University of Wisconsin Press*. 1992.

Devi SM, Ahmed I, Francalacci P, Hussain MA, Akhter Y, Alvi A, Sechi LA, Mégraud F, Ahmed N. Ancestral European roots of Helicobacter pylori in India. *BMC Genomics* 8:184, 2007. doi: 10.1186/1471-2164-8-184

Sengupta D, Kundu S. Do topological parameters of amino acids within protein contact networks depend on their physico-chemical properties? *Physica A: Statistical Mechanics and its Applications* 391 (17) 4266-4278, 2012. <https://doi.org/10.1016/j.physa.2012.03.034>

Dillehay TD. Probing deeper into first American studies. *Proceedings of the National Academy of Sciences of the United States of America* 106:971–78, 2009.

Dixon MF, Genta RM, Yardley JH, Correa P. Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. *The American Journal of Surgical Pathology* 20 (10) 1161-1181, 1996. [PMID: 8827022 DOI: 10.1097/00000478-199610000-00001]

Doig P1, de Jonge BL, Alm RA, Brown ED, Uria-Nickelsen M, Noonan B, Mills SD, Tummino P, Carmel G, Guild BC, Moir DT, Vovis GF, Trust TJ. Helicobacter pylori physiology predicted from genomic comparison of two strains. *Microbiology and Molecular Biology Reviews* 63:675–707, 1999.

Domínguez-Bello MG, Perez ME, Bortolini MC, Salzano FM, Pericchi LR, Zambrano-Guzmán O, Linz B. Amerindian Helicobacter pylori strains go extinct, as European strains expand their host range. *PLoS One* 3(10):e3307, 2008. <https://doi.org/10.1371/journal.pone.0003307>

Dorer MS, Cohen IE, Sessler TH, Fero J, Salama NR. Natural competence promotes Helicobacter pylori chronic infection. *Infection and Immunity* 81(1): 209-15, 2013. doi: 10.1128/IAI.01042-12

Du MQ, Atherton JC. Molecular subtyping of gastric MALT lymphomas: implications for prognosis and management. *Gut* 55(6):886–893, 2006.

Dubois A, Berg DE, Incecik ET, Fiala N, Heman-Ackah LM, Del Valle J, Yang M, Wirth HP, Perez-Perez GI, Blaser MJ. Host specificity of Helicobacter pylori strains and host responses in experimentally challenged non-human primates. *Gastroenterology* 116(1): 90–6, 1999.

Dykhuisen DE, Polin DS, Dunn JJ, Wilske B, Preac-Mursic V, Dattwyler RJ, Luft BJ. Borrelia burgdorferi is clonal: implications for taxonomy y vaccine development. *Proceedings of the National Academy of Sciences of the United States of America* 90 (21): 10162-10167, 1993.

Echave J, Spielman SJ, Wilke CO. Causes of evolutionary rate variation among protein sites. *Nature Review Genetics* 17(2):109-21, 2016. doi: 10.1038/nrg.2015.18.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5):1792-7, 2004.

Edwards NJ1, Monteiro MA, Faller G, Walsh EJ, Moran AP, Roberts IS, High NJ. Lewis X structures in the O antigen side-chain promote adhesion of Helicobacter pylori to the gastric epithelium. *Molecular Microbiology* 35(6):1530-1539, 2000.

England, J.L., Shakhnovich, E.I., Structural determinant of protein designability. *Physical Review Letters* 90: 218101, 2003.

Falush D1, 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. Traces of human migrations in *Helicobacter pylori* populations. *Science* 299 (5612):1582-5, 2003. doi: 10.1126/science.1080857

Falush D, Kraft C, Taylor NS, Correa P, Fox JG, Achtman M, Suerbaum S. Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. *Proceedings of the National Academy of Sciences of the United States of America* 98, 15056–15061, 2001. doi: 10.1073/pnas.251396098

Feil EJ, Spratt BG. Recombination and the population structures of bacterial pathogens. *Annual Review of Microbiology* 55:561-590, 2001.

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer* 136(5):E359–E386, 2015. doi:10.1002/ijc.29210.

Fernández - Armesto F. The Americas: The History of a Hemisphere. London: Weidenfeld & Nicolson. 2003.

Filipe MI1, Muñoz N, Matko I, Kato I, Pompe-Kirn V, Jutersek A, Teuchmann S, Benz M, Prijon T. Intestinal metaplasia types and the risk of gastric cancer: a cohort study in Slovenia. *International Journal of Cancer* 57: 324–9, 1994.

Foegeding NJ, Caston RR, McClain MS, Ohi MD, Cover TL. An overview of *Helicobacter pylori* VacA toxin biology. *Toxins* 8(6), 2016. doi:10.3390/toxins8060173.

Ford AC, Gurusamy KS, Delaney B, Forman D, Moayyedi P. Eradication therapy for peptic ulcer disease in *Helicobacter pylori*-positive people. *Cochrane Database of Systematic Reviews* 4: CD003840, 2016. doi:10.1002/14651858.CD003840.pub5.

Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. Evolutionary rate in the protein interaction network. *Science* 296,750–752, 2002.

Fu YX, Li WH. Statistical tests of neutrality mutations. *Genetics* 133:693-709, 1993.

Fujikawa A, Shirasaka D, Yamamoto S, Ota H, Yahiro K, Fukada M, Shintani T, Wada A, Aoyama N, Hirayama T, Fukamachi H, Noda M. Mice deficient in protein tyrosine phosphatase receptor type Z are resistant to gastric ulcer induction by VacA of *Helicobacter pylori*. *Nature Genetics* 33(3):375–381, 2003. doi:10.1038/ng1112.

Fujimoto Y, Furusyo N, Toyoda K, Takeoka H, Sawayama Y, Hayashi J. Intrafamilial transmission of *Helicobacter pylori* among the population of endemic areas in Japan. *Helicobacter* 12(2):170-6, 2007.

Furuta Y, Konno M, Osaki T, Yonezawa H, Ishige T, Imai M, Shiwa Y, Shibata-Hatta M, Kanesaki Y, Yoshikawa H, Kamiya S, Kobayashi I. Microevolution of Virulence-Related Genes in *Helicobacter pylori* Familial Infection. *PLoS One* 10(5):e0127197, 2015. doi: 10.1371/journal.pone.0127197.

Futuyma, D. *Evolutionary biology*. 2nd ed. Sinauer Associates Inc. Sunderland, Massachusetts. 1986.

Galibert F, Finan TM, Long SR, Puhler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, Bothe G, Boutry M, Bowser L, Buhrmester J, Cadieu E, Capela D, Chain P, Cowie A, Davis RW, Dreano S, Federspiel NA, Fisher RF, Gloux S, Godrie T, Goffeau A, Golding B, Gouzy J, Gurjal M, HeARNandez-Lucas I, Hong A, Huizar L, Hyman RW, Jones T, Kahn D, Kahn ML, Kalman S, Keating DH, Kiss E, Komp C, Lelaure V, Masuy D, Palm C, Peck MC, Pohl TM, Portetelle D, Purnelle B, Ramsperger U, Surzycki R, Thebault P, Vandenbol M, Vorholter FJ, Weidner S, Wells DH, Wong K, Yeh KC, Batut J. The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* 293:668-672, 2001.

Gangwer KA, Shaffer CL, Suerbaum S, Lacy DB, Cover TL, Bordenstein SR. Molecular evolution of the *Helicobacter pylori* vacuolating toxin gene *vacA*. *Journal of Bacteriology* 192(23):6126-35, 2010.

Gardner SN, Slezak T, Hall BG. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* 31:2877–8, 2015.

Gebert B, Fischer W, Weiss E, Hoffmann R, Haas R. *Helicobacter pylori* vacuolating cytotoxin inhibits T lymphocyte activation. *Science* 301(5636):1099–1102, 2003. doi:10.1126/science.1086871.

Genta RM, Hamner HW, Graham DY. Gastric lymphoid follicles in *Helicobacter pylori* infection: frequency, distribution, and response to triple therapy. *Human Pathology* 24(6):577–583. 1993.

Gewirtz AT, Yu Y, Krishna US, Israel DA, Lyons SL, Peek RM. *Helicobacter pylori* flagellin evades toll-like receptor 5-mediated innate immunity. *Journal of Infectious Diseases* 189: 1914-1920, 2004. [PMID: 15122529 DOI: 10.1086/386289].

Ginisty H, Sicard H, Roger B, Bouvet P. Structure and functions of nucleolin. *Journal of Cell Science* 112 (Pt 6) 761–772, 1999.

Goebel T, Waters MR, O'Rourke DH. The late Pleistocene dispersal of modern humans in the Americas. *Science* 319:1497–502, 2008.

Gogarten JP, Doolittle WF, Lawrence JG. Prokaryotic evolution in light of gene transfer. *Molecular Biology and Evolution* 19:2226-2238. 2002.

González CA, Pardo ML, Liso JM, Alonso P, Bonet C, Garcia RM, Sala N, Capella G, Sanz-Anquela JM. Gastric cancer occurrence in preneoplastic lesions: a long-term follow-up in a high-risk area in Spain. *International Journal of Cancer* 127: 2654–60, 2010. doi: 10.1002/ijc.25273

Gordon DM, Wexler M, Reardon TB, Murphy PJ. The genetic structure of *Rhizobium* populations. *Soil Biology and Biochemistry* 27:491-499, 1995.

Graham DY, Opekun AR, Osato MS, El-Zimaity HM, Lee CK, Yamaoka Y, Qureshi WA, Cadoz M, Monath TP. Challenge model for *Helicobacter pylori* infection in human volunteers. *Gut* 53: 1235-1243, 2004. [PMID: 15306577 DOI: 10.1136/gut.2003.037499]

Greene LH, Higman VA. Uncovering network systems within protein structures. *Journal of Molecular Biology* 334: 781–791, 2003.

Greene LH. Protein structure networks. *Briefings in Functional Genomics* 11: 469–478, 2012.

Grossman SR1, Andersen KG, Shlyakhter I, Tabrizi S, Winnicki S, Yen A, Park DJ, Griesemer D, Karlsson EK, Wong SH, Cabili M, Adegbola RA, Bamezai RN, Hill AV, Vannberg FO, Rinn JL; 1000 Genomes Project, Lander ES, Schaffner SF, Sabeti PC. Identifying recent adaptations in large-scale genomic data. *Cell* 152(4): 703–13, 2013. doi:10.1016/j.cell.2013.01.035.

Gu X, Zou Y, Su Z, Huang W, Zhou Z, Arendsee Z, Zeng Y. An update of DIVERGE software for functional divergence analysis of protein family. *Molecular Biology and Evolution* 30(7): 1713–9, 2013. doi: 10.1093/molbev/mst069.

Gu X. Statistical methods for testing functional divergence after gene duplication. *Molecular Biology and Evolution* 16(12): 1664–74, 1999.

Gu X. Maximum-likelihood approach for gene family evolution under functional divergence. *Molecular Biology and Evolution* 18: 453–464, 2001.

Gupta VR, Patel HK, Kostolansky SS, Ballivian RA, Eichberg J, Blanke SR. Sphingomyelin functions as a novel receptor for *Helicobacter pylori* VacA. *PLOS Pathogens* 4, e1000073, 2008.

Gutiérrez-Escobar AJ, Trujillo E, Acevedo O, Bravo MM. Phylogenomics of Colombian *Helicobacter pylori* isolates. *Gut Pathogens* 9:52, 2017. doi: 10.1186/s13099-017- 0201-1.

Gutiérrez-Escobar, AJ. Molecular evolutionary analysis of adherence: evidence of positive selection operating on AlpAB locus and horB adhesins of *Helicobacter pylori*. *Revista udca actualidad y divulgación científica* 16 (1) 3-15, 2013.

Guevara AA, Criollo A, Suarez JJ, Bohórquez ME. A Systematic Review of Genetic Coevolution of *Homo Sapiens* and *Helicobacter Pylori*: Implications for Development of

Gastric Cancer. *Revista Colombiana de Gastroenterologia* 31(4):376-390, 2016. ISSN 0120-9957

Hannelien V, Karel G, Jo VD, Sofie S. The role of CXC chemokines in the transition of chronic inflammation to esophageal and gastric cancer. *Biochimica et Biophysica Acta* 1825:117–129, 2012.

Harris PR, Wright SW, Serrano C, Riera F, Duarte I, Torres J, Peña A, Rollán A, Viviani P, Guiraldes E, Schmitz JM, Lorenz RG, Novak L, Smythies LE, Smith PD. Helicobacter pylori gastritis in children is associated with a regulatory T-cell response. *Gastroenterology* 134:491-499, 2008. doi: 10.1053/j.gastro.2007.11.006.

Hedrick PW. Evolutionary genetics of the major histocompatibility complex. *The American Naturalist* 143, 945–964, 1994.

Helicobacter and Cancer Collaborative Group. Gastric cancer and Helicobacter pylori: A combined analysis of 12 case control studies nested within prospective cohorts. *Gut* 49:347-53, 2001. <http://dx.doi.org/10.1136/gut.49.3.347>

Hennig EE, Allen JM, Cover TL. Multiple chromosomal loci for the babA gene in Helicobacter pylori. *Infection and Immunity* 74(5): 3046–3051, 2006. doi: 10.1128/IAI.74.5.3046-3051.2006.

Herrera V, Parsonnet J. Helicobacter pylori and gastric adenocarcinoma. *Clinical Microbiology and Infection* 15:971–76, 2009.

Herrero R, Park JY, Forman D The fight against gastric cancer—the IARC Working Group report. *Best Practice & Research: Clinical Gastroenterology* 28(6):1107–1114, 2014. doi:10.1016/j.bpg.2014.10.003.

Hirano K, Miki Y, Hirai Y, Sato R, Itoh T, Hayashi A, Yamanaka M, Eda S, Beppu M. A multifunctional shuttling protein nucleolin is a macrophage receptor for apoptotic cells. *Journal of Biological Chemistry* 280: 39284–39293, 2005.

Holmes EC, Urwin R, Maiden MC. The influence of recombination on the population structure and evolution of the human pathogen Neisseria meningitidis. *Molecular Biology and Evolution* 16:741-749, 1999.

Huang JY, Sweeney EG, Sigal M, Zhang HC, Remington SJ, Cantrell MA, Kuo CJ, Guillemin K, Amieva MR. Chemodetection and destruction of host urea allows Helicobacter pylori to locate the epithelium. *Cell Host & Microbe* 18(2):147–156, 2015. doi:10.1016/j.chom.2015.07.002.

Huson DH. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 14:68–73, 1998.

International Agency for Research on Cancer-IARC. Infection with *Helicobacter Pylori*. *Monographs on the Evaluation of Carcinogenic Risks to Humans* 61 177–220, 1994.

International Agency for Research on Cancer-IARC. Schistosomes, liver flukes and *Helicobacter pylori*. *Monographs on the evaluation of carcinogenic risks to humans* 61, 1994.

International Agency for Research on Cancer. Globocan, 2012. http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx.

Ilver D, Arnqvist A, Ogren J, Frick IM, Kersulyte D, Incecik ET, Berg DE, Covacci A, Engstrand L, Boren T. *Helicobacter pylori* adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science* 279 (5349):373-377, 1998.

Ingvarsson PK. Gene expression and protein length influence codon usage and rates of sequence evolution in *Populus tremula*. *Molecular Biology and Evolution* 24, 836–844, 2007.

Israel DA, Lou AS, Blaser MJ. Characteristics of *Helicobacter pylori* natural transformation. *FEMS Microbiology Letters* 186, 275–280, 2000.

Italiani P, Boraschi D. From monocytes to M1/M2 macrophages: phenotypical vs. functional differentiation. *Frontiers in Immunology* 5:514, 2014. doi:10.3389/fimmu.2014.00514.

Ito Y, Azuma T, Ito S, Suto H, Miyaji H, Yamazaki Y, Kohli Y, Kuriyama M. Full-length sequence analysis of the *vacA* gene from cytotoxic and noncytotoxic *Helicobacter pylori*. *Journal of Infectious Diseases* 178: 1391–1398, 1998.

Jang JY, Yoon HJ, Yoon JY, Kim HS, Lee SJ, Kim KH, Kim DJ, Jang S, Han BG, Lee BI, Suh SW. Crystal structure of the TNF- α -Inducing protein (Tipalpha) from *Helicobacter pylori*: Insights into Its DNA-binding activity. *Journal of Molecular Biology* 392(1):191-7, 2009.

Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Research* 2: 2, 2006. doi: 10.1186/1745-7580-2-2

de Jonge R, Durrani Z, Rijpkema SG, Kuipers EJ, van Vliet AH, Kusters JG. Role of the *Helicobacter pylori* outer-membrane proteins AlpA and AlpB in colonization of the guinea pig stomach. *Journal of Medical Microbiology* 53:375–379, 2004.

Jordan IK. The Columbian Exchange as a source of adaptive introgression in human populations. *Biology Direct* 11(1):17, 2016. doi: 10.1186/s13062-016-0121-x.

Kaas RS, Friis C, Ussery DW, Aarestrup FM. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. *BMC Genomics* 13:577, 2012. doi: 10.1186/1471-2164-13-577.

Kamen H. Spain's Road to Empire: The Making of a World Power 1492–1763. London: Allen Lane. pp 656. 2002.

Keilberg D, Ottemann KM. How *Helicobacter pylori* senses, targets and interacts with the gastric epithelium. *Environmental Microbiology* 18(3):791–806, 2016. doi:10.1111/1462-2920.13222.

Gangwer KA1, Mushrush DJ, Stauff DL, Spiller B, McClain MS, Cover TL, Lacy DB. Cover, and D. Borden Lacy. Crystal structure of the *Helicobacter pylori* vacuolating toxin p55 domain. *Proceedings of the National Academy of Sciences of the United States of America* 104(41): 16293–16298, 2007.

Kenneth L, McColl. *Helicobacter pylori* Infection. *The new England Journal of Medicine* 362(17):1597-1604, 2010.

Kersulyte D, Kalia A, Gilman RH, Mendez M, Herrera P, Cabrera L, Velapatiño B, Balqui J, Paredes Puente de la Vega F, Rodriguez Ulloa CA, Cok J, Hooper CC, Dailide G, Tamma S, Berg DE. *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, 2010. doi: 10.1371/journal.pone.0015076.

Khalifa MM, Sharaf RR, Aziz RK. *Helicobacter pylori*: a poor man's gut pathogen? *Gut Pathogens* 2(1):2, 2010. doi: 10.1186/1757-4749-2-2.

Kim DJ, Park JH, Franchi L, Backert S, Núñez G. The Cag pathogenicity island and interaction between TLR2/NOD2 and NLRP3 regulate IL-1b production in *Helicobacter pylori* infected dendritic cells. *European Journal of Immunology* 43(10):2650–2658, 2013. doi:10.1002/eji.201243281.

Kim IJ, Blanke SR. Remodeling the host environment: modulation of the gastric epithelium by the *Helicobacter pylori* vacuolating toxin (VacA). *Frontiers in Cellular and Infection Microbiology* 2:37, 2012. doi:10.3389/fcimb.2012.00037.

Kim N, Weeks D, Moo J, Scott D, Young M, Sachs. Proteins released by *Helicobacter pylori* in vitro. *Journal of Bacteriology* 184(22):6155-6162, 2002.

Kim SS1, Ruiz VE, Carroll JD, Moss SF. *Helicobacter pylori* in the pathogenesis of gastric cancer and gastric lymphoma. *Cancer Letters* 305 228–238, 2011. doi:10.1016/j.canlet.2010.07.014.

Kimura, M. Evolutionary rate at the molecular level. *Nature* 217(129):624-626, 1968.

Kimura, M. The neutral theory of molecular evolution. Cambridge University Press. 1983.

Kodaman N, Pazos A, Schneider BG, Piazuolo MB, Mera R, Sobota RS, Sicinschi LA, Shaffer CL, Romero-Gallo J, de Sablet T, Harder RH, Bravo LE, Peek RM Jr, Wilson KT, Cover TL, Williams SM, Correa P. Human and *Helicobacter pylori* coevolution shapes the

risk of gastric disease. *Proceedings of the National Academy of Sciences of the United States of America* 111(4):1455-60, 2014. doi: 10.1073/pnas.1318093111.

Koonin EV, Wolf YI. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Research* 36:6688–6719, 2008.

Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution* 22(5):1208-22, 2005. doi: 10.1093/molbev/msi105

Kraft C, Suerbaum S. Mutation and recombination in *Helicobacter pylori*: mechanisms and role in generating strain diversity. *International Journal of Medical Microbiology* 295(5):299-305, 2005.

Krishnan A1, Zbilut JP, Tomita M, Giuliani A. Proteins as networks: Usefulness of graph theory in protein science. *Current Protein & Peptide Science* 9, 28–38, 2008.

Kristensen DM, Kannan L, Coleman MK, Wolf YI, Sorokin A, Koonin EV, Mushegian A. A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. *Bioinformatics* 26:1481–1487, 2010.

Kumar N, Albert MJ, Al Abkal H, Siddique I, Ahmed N. What constitutes an Arabian *Helicobacter pylori*? Lessons from comparative genomics. *Helicobacter* 22(1), 2017. doi: 10.1111/hel.12323.

Kumar N, Mariappan V, Baddam R, Lankapalli AK, Shaik S, Goh KL, Loke MF, Perkins T, Benghezal M, Hasnain SE, Vadivelu J, Marshall BJ, Ahmed N. Comparative genomic analysis of *Helicobacter pylori* from Malaysia identifies three distinct lineages suggestive of differential evolution. *Nucleic Acids Research* 43(1):324-35, 2015. doi: 10.1093/nar/gku1271.

Kumar N, Mukhopadhyay AK, Patra R, De R, Baddam R, Shaik S, Alam J, Tiruvayipati S, Ahmed N. Next-generation sequencing and de novo assembly, genome organization, and comparative genomic analyses of the genomes of two *Helicobacter pylori* isolates from duodenal ulcer patients in India. *Journal of Bacteriology* 194(21):5963-4, 2012. doi: 10.1128/JB.01371-12

Kumar S, Bhan Ch, Kumar A, Kumar A, Nath G. Diagnosis of *Helicobacter pylori*: What should be the gold standard?. *World Journal of Gastroenterology* 20(36): 12847–12859, 2014. doi: 10.3748/wjg.v20.i36.12847

Kusters JG, Vliet AM, Kuipers EJ. Pathogenesis of *Helicobacter pylori* Infection. *Clinical Microbiology Reviews* 19(3):449–490, 2006. doi: 10.1128/CMR.00054-05

Larson B. *Trials of Nation Making: Liberalism, Race, and Ethnicity in the Andes, 1810–1910*. Cambridge, UK: Cambridge University Press. 2004.

Lavrin A. Sexuality and Marriage in Colonial Latin America. Lincoln: University of Nebraska Press. 1989.

Lawrence JG. Gene transfer, speciation, y the evolution of bacterial genomes. *Current Opinion in Microbiology* 2:519-523, 1999.

Lawrence JG. Clustering of antibiotic resistance genes: beyond the selfish operon. *ASM News* 66:281-286, 2000.

Lawrence JG. Gene transfer in bacteria: speciation without species? *Theoretical Population Biology* 61:449-460, 2002.

Lenski RE, May RM. The evolution of virulence in parasites and pathogens: reconciliation between two competing hypotheses. *Journal of Theoretical Biology* 169:253–65, 1994.

Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* 44(W1):W242-5, 2016. doi: 10.1093/nar/gkw290

Levin BR, Bergstrom CT. Bacteria are different: observations, interpretations, speculations, y opinions about the mechanisms of adaptive evolution in prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America* 97:6981-6985, 2000. <https://doi.org/10.1073/pnas.97.13.6981>

Li WH. Molecular Evolution. Sinauer Associates Incorporated, Sunderland, USA. pp. 487. 1997. ISBN: 0878934634

Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11):1451-2, 2009.

Lindholm C, Quiding-Järbrink M, Lönröth H, Hamlet A, Svennerholm AM. Local cytokine response in Helicobacter pylori-infected subjects. *Infection and Immunity*; 66 (12): 5964-5971, 1998.

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. An African origin for the intimate association between humans and Helicobacter pylori. *Nature* 445(7130):915-8, 2007. doi:10.1038/nature05562

Linz B, Windsor HM, Gajewski JP, Hake CM, Drautz DI, Schuster SC, Marshall BJ. Helicobacter pylori genomic microevolution during naturally occurring transmission between adults. *PLoS One* 10: 8 (12):e82187, 2013. doi: 10.1371/journal.pone.0082187.

Linz B, Vololonantenainab CR, Seck A, Carod JF, Dia D, Garin B, Ramanampamonjy RM, Thiberge JM, Raymond J, Breurec S. Population genetic structure and isolation by distance of Helicobacter pylori in Senegal and Madagascar. *PLoS One* 9(1):e87355, 2014.

Linz B, Windsor HM, McGraw JJ, Hansen LM, Gajewski JP, Tomsho LP, Hake CM, Solnick JV, Schuster SC, Marshall BJ. A mutation burst during the acute phase of *Helicobacter pylori* infection in humans and rhesus macaques. *Nature Communications* 5:4165, 2014b. doi: 10.1038/ncomms5165.

Loveman M. Whiteness in Latin America: measurement and meaning in national censuses. *Journal de la société des américanistes* 95:207–34, 2009.

Loveman M. National Colors: Racial Classification and the State in Latin America. New York: Oxford University Press. 2014.

Lu H, Wu JY, Beswick EJ, Ohno T, Odenbreit S, Haas R, Reyes VE, Kita M, Graham DY, Yamaoka Y. Functional and intracellular signaling differences associated with the *Helicobacter pylori* AlpAB adhesin from western and East Asian strains. *Journal of Biological Chemistry* 282(9):6242–6254, 2007.

Lundgren A, Suri-Payer E, Enarsson K, Svennerholm AM, Lundin BS. *Helicobacter pylori*-specific CD4⁺ CD25^{high} regulatory T cells suppress memory T-cell responses to *H. pylori* in infected individuals. *Infection and Immunity* 71(4):1755-1762, 2003. doi: 10.1128/IAI.71.4.1755-1762.2003

Luzza F, Parrello T, Monteleone G, Sebkova L, Romano M, Zarrilli R, Imeneo M, Pallone F. Up-regulation of IL-17 is associated with bioactive IL-8 expression in *Helicobacter pylori*-infected human gastric mucosa. *The Journal of Immunology* 165: 5332-5337, 2000.

Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science* 290(5494):1151–5, 2000.

Maeda S, Ogura K, Yoshida H, Kanai F, Ikenoue T, Kato N, Shiratori Y, Omata M. Major virulence factors, VacA and CagA, are commonly positive in *Helicobacter pylori* isolates in Japan. *Gut* 42(3): 338-43, 1998.

Magalhães A, Reis CA. *Helicobacter pylori* adhesion to gastric epithelial cells is mediated by glycan receptors. *Brazilian Journal of Medical and Biological Research* 43:611-618, 2010.

Maggi Solcà N1, Bernasconi MV, Valsangiacomo C, Van Doorn LJ, Piffaretti JC. Population genetics of *Helicobacter pylori* in the southern part of Switzerland analysed by sequencing of four housekeeping genes (atpD, glnA, scoB and recA), and by vacA, cagA, iceA and IS605 genotyping. *Microbiology* 147:1693-1707, 2001.

Mahdavi J, Boreín T, Vandenbroucke C, Appelmelk B. Limited role of lipopolysaccharide Lewis antigens in adherence of *Helicobacter pylori* to the human gastric epithelium. *Infection and Immunity* 71(5):2876–2880, 2003.

Mahdavi J, Sonden B, Hurtig M, Olfat FO, Forsberg L, Roche N, Angstrom J, Larsson T, Teneberg S, Karlsson KA, Altraja S, Wadstrom T, Kersulyte D, Berg DE, Dubois A,

Petersson C, Magnusson KE, Norberg T, Lindh F, Lundskog BB, Arnqvist A, Hammarstrom L, Boren T. Helicobacter pylori SabA adhesin in persistent infection and chronic inflammation. *Science* 297(5581):573-578, 2002. doi: 10.1126/science.1069076

Malfertheiner P. The intriguing relationship of Helicobacter pylori infection and acid secretion in peptic ulcer disease and gastric cancer. *Digestive Diseases* 29(5):459–464, 2011. doi:10.1159/000332213.

Malfertheiner P, Fry LC, Monkemuller K. Can gastric cancer be prevented by Helicobacter pylori eradication? *Best Practice & Research Clinical Gastroenterology* 20(4):709–719, 2006. doi:10.1016/j.bpg.2006. 04.005.

Marais A, Mendz GL, Hazell SL, Mégraud F. Metabolism and genetics of Helicobacter pylori: the genome era. *Microbiology and Molecular Biology Reviews* 63:642–674, 1999.

Marsh L, Griffiths CS. Protein structural influences in rhodopsin evolution. *Molecular Biology and Evolution* 22, 894–904, 2005. Doi: 10.1093/molbev/msi081

Marshall BJ, Armstrong JA, McGeachie DB, Glancy RJ. Attempt to fulfil Koch's postulates for pyloric Campylobacter. *The Medical Journal of Australia* 142:436–439, 1985.

Matsukura N, Suzuki K, Kawachi T, Aoyagi M, Sugimura T, Kitaoka H, Numajiri H, Shirota A, Itabashi M, Hirota T. Distribution of marker enzymes and mucin in intestinal metaplasia in human stomach and relation to complete and incomplete types of intestinal metaplasia to minute gastric carcinomas. *Journal of the National Cancer Institute* 65: 231–40, 1980.

Mattsson A, Quiding-Järbrink M, Lönroth H, Hamlet A, Ahlstedt I, Svennerholm A. Antibody-secreting cells in the stomachs of symptomatic and asymptomatic Helicobacter pylori-infected subjects. *Infection and Immunity* 66: 2705-2712, 1998.

McAlister LN. Spain and Portugal in the New World, 1492–1700. University of Minnesota Press. 1st ed. Minneapolis, Minnesota. 1984.

Meltzer DJ. First Peoples in a New World: Colonizing Ice Age America. University of California Press. Berkeley. 2009.

Montenegro RA, Stephens C. Indigenous health in Latin America and the Caribbean. *The Lancet* 367:1859–69, 2006.

Montinaro F, Busby GB, Pascali VL, Myers S, Hellenthal G, Capelli C. Unravelling the hidden ancestry of American admixed populations. *Nature Communications* 6:6596, 2015. doi:10.1038/ncomms7596.

Moodley Y, Linz B, Yamaoka Y, Windsor HM, Breurec S, Wu JY, Maady A, Bernhöft S, Thiberge JM, Phuanukoonnon S, Jobb G, Siba P, Graham DY, Marshall BJ, Achtman M. The peopling of the Pacific from a bacterial perspective. *Science* 323(5913):527 -30, 2009. doi: 10.1126/science.1166083.

Moreno-Estrada A, Gravel S, Zakharia F, McCauley JL, Byrnes JK, Gignoux CR, Ortiz-Tello PA, Martínez RJ, Hedges DJ, Morris RW, Eng C, Sandoval K, Acevedo-Acevedo S, Norman PJ, Layrisse Z, Parham P, Martínez-Cruzado JC, González Burchard E, Cuccaro ML, Martin ER, Bustamante CD. Reconstructing the population genetic history of the Caribbean. *PLOS Genetics* 9:e1003925, 2013. doi: 10.1371/journal.pgen.1003925

Morner M. Race Mixture in the History of Latin America. Little, Brown and Company. Boston. 1st ed. pp 178. 1967.

Morris AJ, Ali MR, Nicholson GI, Perez-Perez GI, Blaser MJ. Long term follow-up of voluntary ingestion of *Helicobacter pylori*. *Annals of Internal Medicine* 114:662-663, 1991.

Morris A, Nicholson G. Ingestion of *Campylobacter pyloridis* causes gastritis and raised fasting gastric pH. *The American Journal of Gastroenterology* 82(3):192-199, 1987.

Moyat M, Velin D. Immune responses to *Helicobacter pylori* infection. *World Journal of Gastroenterology* 20(19): 5583–5593, 2014. doi: 10.3748/wjg.v20.i19.5583.

Mu J, Awadalla P, Duan J, McGee KM, Keebler J, Seydel K, McVean GA, Su XZ. Genome-wide variation and identification of vaccine targets in the *Plasmodium falciparum* genome. *Nature Genetics* 39, 126–130, 2006. doi: 10.1038/ng1924

Doncheva NT, Assenov Y, Domingues FS, Albrecht M. Topological analysis and interactive visualization of biological networks and protein structures. *Nature Protocols* 7, 670–685, 2012.

Nedenskov, P. Nutritional requirements for growth of *Helicobacter pylori*. *Applied and Environmental Microbiology* 60:3450–3453. 1994.

Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics* 39:121-52, 2005.

Nei M, Kumar S. Molecular evolution and phylogenetics. Oxford University Press, New York. 2000.

Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America* 76(10):5269-5273, 1979.

Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, White O, Salzberg SL, Smith HO, Venter JC, Fraser CM. Evidence for lateral gene transfer between Archaea y bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399(6734):323-329, 1999. doi:10.1038/20601

Nurgalieva ZZ, Conner ME, Opekun AR, Zheng CQ, Elliott SN, Ernst PB, Osato M, Estes MK, Graham DY. B-cell and T-cell immune responses to experimental *Helicobacter pylori* infection in humans. *Infection and Immunity* 73: 2999-3006, 2005. doi: 10.1128/IAI.73.5.2999-3006.2005

Obbard DJ, Jiggins FM, Halligan DL, Little TJ. Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Current Biology* 16(6): 580–585, 2006.

Odenbreit S, Faller G, Haas R. Role of the AlpAB proteins and lipopolysaccharide in adhesion of *Helicobacter pylori* to human gastric tissue. *International Journal of Medical Microbiology* 2002; 292(3):247-256.

Odenbreit S, Swoboda K, Barwig I, Ruhl S, Borén T, Koletzko S, Haas R. Outer membrane protein expression profile in *Helicobacter pylori* clinical isolates. *Infection and Immunity* 77(9): 3782-3790, 2009.

Odenbreit S, Till M, Hofreuter D, Faller G, Haas R. Genetic and functional characterization of the alpAB gene locus essential for the adhesion of *Helicobacter pylori* to human gastric tissue. *Molecular Microbiology* 31(5):1537-1548, 1999.

Odenbreit S. Adherence properties of *Helicobacter pylori*: impact on pathogenesis and adaptation to the host. *International Journal of Medical Microbiology* 295:317–324, 2005.

Oertli M, Noben M, Engler DB, Semper RP, Reuter S, Maxeiner J, Gerhard M, Taube C, Muller A. *Helicobacter pylori* gamma-glutamyl transpeptidase and vacuolating cytotoxin promote gastric persistence and immune tolerance. *Proceedings of the National Academy of Sciences of the United States of America* 110(8):3047–3052, 2013. doi:10.1073/pnas.1211248110.

Ohno S. Evolution by gene duplication. Springer Science+Business Media. New York, USA. 1970. Doi: 10.1007/978-3-642-86659-3

Ohta T. The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics* 23:263-286, 1992.

Oleastro M, Cordeiro R, Ménard A, Yamaoka Y, Queiroz D, Mégraud F, Monteiro L. Allelic diversity and phylogeny of homB, a novel co-virulence marker of *Helicobacter pylori*. *BMC Microbiology* 9:248, 2009.

Oleastro M, Cordeiro R, Yamaoka Y, Queiroz D, Mégraud F, Monteiro L, Ménard A. Disease association with two *Helicobacter pylori* duplicate outer membrane protein genes, homB and homA. *Gut Pathogens* 1(1): 12, 2009b. doi: 10.1186/1757-4749-1-12.

O'Rourke DH. Human migrations: the two roads taken. *Current Biology* 19(5):R203-R205, 2009.

Ossa H, Aquino J, Pereira R, Ibarra A, Ossa RH, Pérez LA, Granda JD, Lattig MC, Groot H, Fagundes de Carvalho E, Gusmão L. Outlining the Ancestry Landscape of Colombian Admixed Populations. *PLoS One*. 11(10):e0164414, 2016. doi: 10.1371/journal.pone.0164414.

Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research* 42(Database issue):D206-D214, 2014. doi:10.1093/nar/gkt1226

Pagán I, Holmes EC, Simon-Loriere E. Level of gene expression is a major determinant of protein evolution in the viral order mononegavirales. *Journal of Virology* 86(9):5253-63, 2012. doi: 10.1128/JVI.06050-11.

Pál C, Papp B, Lercher MJ. An integrated view of protein evolution. *Nature Reviews Genetics* 7, 337–348, 2006. doi: 10.1038/nrg1838

Pardo C, Cendales R. Incidencia estimada y mortalidad por cáncer en Colombia, 2002-2006. Instituto Nacional de Cancerología, Bogotá, Colombia. 2010.

Parrish CR, Holmes EC, Morens DM, Park EC, Burke DS, Calisher CH, Laughlin CA, Saif LJ, Daszak P. Cross species virus transmission and the emergence of new epidemic diseases. *Microbiology and Molecular Biology Reviews* 72(3):457-70, 2008. doi: 10.1128/MMBR.00004-08

Paterson S, Vogwill T, Buckling A, Benmayor R, Spiers AJ, Thomson NR, Quail M, Smith F, Walker D, Libberton B, Fenton A, Hall N, Brockhurst MA. Antagonistic coevolution accelerates molecular evolution. *Nature* 464(7286):275-8, 2010. doi: 10.1038/nature08798.

Peck B, Ortkamp M, Diehl KD, Hundt E, Knapp B. Conservation, localization and expression of HopZ, a protein involved in adhesion of *Helicobacter pylori*. *Nucleic Acids Research* 27(16):3325- 3333, 1999.

Peng Y, Leung HC, Yiu SM, Chin FY. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28(11):1420-8, 2012. doi: 10.1093/bioinformatics/bts174.

Perez-Perez GI, Dworkin BM, Chodos JE, Blaser MJ. *Campylobacter pylori* antibodies in humans. *Annals of Internal Medicine* 109: 11-17, 1988. doi: 10.7326/0003-4819-109-1-11

Perez-Perez GI, Rothenbacher D, Brenner H. Epidemiology of *Helicobacter pylori* infection. *Helicobacter* 9 Suppl 1:1-6, 2004. doi: 10.1111/j.1083-4389.2004.00248.x

Perna NT, Plunkett G 3rd, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA, Pósfai G, Hackett J, Klink S, Boutin A, Shao Y, Miller L, Grotbeck EJ, Davis NW, Lim A, Dimalanta ET, Potamosis KD, Apodaca J, Anantharaman

TS, Lin J, Yen G, Schwartz DC, Welch RA, Blattner FR. Genome sequence of entero haemorrhagic *Escherichia coli* O157:H7. *Nature* 409:529-533, 2001. doi: 10.1038/35054089

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 25(13):1605-12, 2004.

Pflock M, Finsterer N, Joseph B, Mollenkopf H, Meyer TF, Beier D. Characterization of the ArsRS regulon of *Helicobacter pylori*, involved in acid adaptation. *Journal of Bacteriology* 188(10):3449– 3462, 2006. doi:10.1128/JB.188.10.3449-3462.2006.

Pflock M, Kennard S, Delany I, Scarlato V, Beier D. Acid-induced activation of the urease promoters is mediated directly by the ArsRS two-component system of *Helicobacter pylori*. *Infection and Immunity* 73(10):6437–6445, 2005. doi:10.1128/IAI.73.10.6437-6445.2005.

Pickett BE, Liu M, Sadat EL, Squires RB, Noronha JM, He S, Jen W, Zaremba S, Gu Z, Zhou L, Larsen CN, Bosch I, Gehrke L, McGee M, Klem EB, Scheuermann RH. Metadata-driven comparative analysis tool for sequences (meta-CATS): an automated process for identifying significant sequence variations that correlate with virus attributes. *Virology* 447(1-2):45-51, 2013.

Pinto AV, Mathieu A, Marsin S, Veaute X, Ielpi L, Labigne A, Radicella JP. Suppression of homologous and homeologous recombination by the bacterial MutS2 protein. *Molecular Cell* 17, 113–120, 2005.

Plummer M, Franceschi S, Vignat J, Forman D, de Martel C. Global burden of gastric cancer attributable to *Helicobacter pylori*. *International Journal of Cancer* 136(2):487–490, 2015. doi:10.1002/ijc.28999.

Polk DB, Peek RM Jr., *Helicobacter pylori*: gastric cancer and beyond. *Nature Reviews Cancer* 10: 403–414, 2010. doi: 10.1038/nrc2857.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013. <http://www.R-project.org/>.

Rad R, Brenner L, Bauer S, Schwendy S, Layland L, da Costa CP, Reindl W, Dossumbekova A, Friedrich M, Saur D, Wagner H, Schmid RM, Prinz C. CD25+/Foxp3+ T cells regulate gastric inflammation and *Helicobacter pylori* colonization in vivo. *Gastroenterology* 131: 525-537, 2006. doi: 10.1053/j.gastro.2006.05.001

Raghavan S, Fredriksson M, Svennerholm AM, Holmgren J, Suri-Payer E. Absence of CD4+CD25+ regulatory T cells is associated with a loss of regulation leading to increased pathology in *Helicobacter pylori*-infected mice. *Clinical & Experimental Immunology* 132: 393-400, 2003. doi: 10.1046/j.1365-2249.2003.02177.x

Ramarao N, Gray-Owen SD, Backert S, Meyer TF. Helicobacter pylori inhibits phagocytosis by professional phagocytes involving type IV secretion components. *Molecular Microbiology* 37: 1389-1404, 2000. doi: 10.1046/j.1365-2958.2000.02089.x

Ramarao N, Gray-Owen SD, Meyer TF. Helicobacter pylori induces but survives the extracellular release of oxygen radicals from professional phagocytes using its catalase activity. *Molecular Microbiology* 38: 103-113, 2000. doi:10.1046/j.1365-2958.2000.02114.x

Ramírez A, Recavarren S, Arias J, Gilman H, León R, Garrido G, Biber M, Rodríguez C, Chinga E. Helicobacter pylori, Gastritis crónica, úlcera gástrica y úlcera duodenal: estudio de 1638 pacientes. *Revista de Gastroenterología del Peru* 19:196–201, 1999.

Ramsay H, Rieseberg LH, Ritland K. The correlation of evolutionary rate with pathway position in plant terpenoid biosynthesis. *Molecular Biology and Evolution* 26, 1045–1053, 2009. doi: 10.1093/molbev/msp021.

Rhead JL, Letley DP, Mohammadi M, Hussein N, Mohagheghi MA, Eshagh Hosseini M, Atherton JC. A new Helicobacter pylori vacuolating cytotoxin determinant, the intermediate region, is associated with gastric cancer. *Gastroenterology* 133(3):926–936, 2007.

Ribeiro D. O Povo Brasileiro - A Formação o Sentido do Brasil. Companhia das Letras. São Paulo. Segunda edição. 1995.

Robinson K, Letley DP, Kaneko K. The Human Stomach in Health and Disease: Infection Strategies by Helicobacter pylori. In: Tegtmeyer N., Backert S. (eds) Molecular Pathogenesis and Signal Transduction by Helicobacter pylori. *Current Topics in Microbiology and Immunology* 400. Springer, Cham. 2017.

Roche N, Ilver D, Angström J, Barone S, Telford JL, Teneberg S. Human gastric glycosphingolipids recognized by Helicobacter pylori vacuolating cytotoxin VacA. *Microbes and Infection* 9, 605–614, 2007. doi: 10.1016/j.micinf.2007.01.023

Rokkas T, Filipe MI, Sladen GE. Detection of an increased incidence of early gastric cancer in patients with intestinal metaplasia type III who are closely followed up. *Gut* 32: 1110–3, 1991.

Rorick MM, Wagner GP. Protein structural modularity and robustness are associated with evolvability. *Genome Biology and Evolution* 3: 456–475, 2011.

Rossez Y, Gosset P, Boneca IG, Magalhaes A, Ecobichon C, Reis CA, Cieniewski-Bernard C, Joncquel Chevalier Curt M, Leonard R, Maes E, Sperandio B, Slomianny C, Sansonetti PJ, Michalski JC, Robbe-Masselot C. The lacdiNAc-specific adhesin LabA mediates adhesion of Helicobacter pylori to human gastric mucosa. *J Infect Dis* 210(8):1286–1295, 2014. doi:10.1093/infdis/jiu239.

Ruiz-Linares A, Adhikari K, Acuña-Alonzo V, Quinto-Sanchez M, Jaramillo C, Arias W, Fuentes M, Pizarro M, Everardo P, de Avila F, Gómez-Valdés J, León-Mimila P, Hunemeier

T, Ramallo V, Silva de Cerqueira CC, Burley MW, Konca E, de Oliveira MZ, Veronez MR, Rubio-Codina M, Attanasio O, Gibbon S, Ray N, Gallo C, Poletti G, Rosique J, Schuler-Faccini L, Salzano FM, Bortolini MC, Canizales-Quinteros S, Rothhammer F, Bedoya G, Balding D, Gonzalez-José R. Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genetics* 10(9):e1004572, 2014. doi:10.1371/journal.pgen.1004572.

Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. Positive natural selection in the human lineage. *Science* 312(5780): 1614–20, 2006. doi:10.1126/science.1124309.

Sachs G, David RS, Wen Y. Gastric Infection by *Helicobacter pylori*. *Current Gastroenterology Reports* 13:540–546, 2011. doi: 10.1007/s11894-011-0226-4

Sagaert X, Van Cutsem E, De Hertogh G, Geboes K, Tousseyn T. Gastric MALT lymphoma: a model of chronic inflammatory-induced tumor development. *Nature Reviews Gastroenterology & Hepatology* 7:336–46, 2010. doi: 10.1038/nrgastro.2010.58.

Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4(4):406-25, 1987.

Sakar SF, Guttman DS. Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. *Applied and Environmental Microbiology* 70:1999-2012, 2004. doi: 10.1128/AEM.70.4.1999-2012.2004

Santoyo G, Romero D. Gene conversion and concerted evolution in bacterial genomes. *FEMS Microbiology Letters* 29(2): 169-83, 2005. doi: 10.1016/j.femsre.2004.10.004.

Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. *Nature Reviews Genetics* 12: 683–691, 2011.

Sawyer S. Statistical tests for detecting gene conversion. *Molecular Biology and Evolution* 6(5):526–38, 1989.

Schmees C, Prinz C, Treptau T, Rad R, Hengst L, Voland P, Bauer S, Brenner L, Schmid RM, Gerhard M. Inhibition of T-cell proliferation by *Helicobacter pylori* gamma-glutamyl transpeptidase. *Gastroenterology* 132: 1820-1833, 2007. doi: 10.1053/j.gastro.2007.02.031

Schuler MM, Nastke MD, Stevanovic S. SYFPEITHI: database for searching and T cell epitope prediction. *Methods in Molecular Biology* 409:75-93, 2007.

Schwarz S, Morelli G, Kusecek B, Manica A, Balloux F, Owen RJ, Graham DY, van der Merwe S, Achtman M, Suerbaum S. Horizontal versus familial transmission of *Helicobacter pylori*. *PLoS Pathogens* 4:e1000180, 2008. doi: 10.1371/journal.ppat.1000180

Senkovich OA, Yin J, Ekshyyan V, Conant C, Traylor J, Adegboyega P, McGee DJ, Rhoads RE, Slepnev S, Testerman TL. *Helicobacter pylori* AlpA and AlpB bind host laminin and

influence gastric inflammation in gerbils. *Infection and Immunity* 79(8) 3106–3116, 2011. doi: 10.1128/IAI.01275-10.

Senthilkumar C, Niranjali S, Jayanthi V, Ramesh T, Devaraj H. Molecular and histological evaluation of tumor necrosis factor-alpha expression in *Helicobacter pylori* mediated gastric carcinogenesis. *Journal of Cancer Research and Clinical Oncology* 137:577–583, 2011.

Serrano C, Wright SW, Bimczok D, Shaffer CL, Cover TL, Venegas A, Salazar MG, Smythies LE, Harris PR, Smith PD. Down-regulated Th17 Responses Are Associated with Reduced Gastritis in *Helicobacter pylori*-infected Children. *Mucosal Immunology* 6: 950-959, 2013. doi: 10.1038/mi.2012.133

Seto K, Hayashi-Kuwabara Y, Yoneta T, Suda H, Tamaki H. Vacuolation induced by cytotoxin from *Helicobacter pylori* is mediated by the EGF receptor in HeLa cells. *FEBS Letters* 431: 347–350, 1998.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13:2498-04, 2003. doi: 10.1101/gr.1239303

Shao SH, Wang H, Chai SG, Liu L. M. Research progress on *Helicobacter pylori* outer membrane protein. *World Journal of Gastroenterology* 11(20):3011-3013, 2005.

Shimizu N, Inada K, Nakanishi H, Tsukamoto T, Ikehara Y, Kaminishi M, Kuramoto S, Sugiyama A, Katsuyama T, Tatematsu M. *Helicobacter pylori* infection enhances glandular stomach carcinogenesis in Mongolian gerbils treated with chemical carcinogens. *Carcinogenesis* 20: 669–676, 1999.

Shiota S, Suzuki R, Matsuo Y, Miftahussurur M, Tran TT, Binh TT, Yamaoka Y. *Helicobacter pylori* from gastric cancer and duodenal ulcer show same phylogeographic origin in the Andean region in Colombia. *PLoS One* 9(8):e105392, 2014. doi: 10.1371/journal.pone.0105392

Sicinschi LA, Correa P, Peek RM, Camargo MC, Piazuelo MB, Romero-Gallo J, Hobbs SS, Krishna U, Delgado A, Mera R, Bravo LE, Schneider BG. CagA C-terminal variations in *Helicobacter pylori* strains from Colombian patients with gastric precancerous lesions. *Clinical Microbiology and Infection* 16(4):369-78, 2010. doi: 10.1111/j.1469-0691.2009.02811.x.

Sierra MS, Cueva P, Bravo LE, Forman D. Stomach cancer burden in Central and South America. *Cancer Epidemiology* 44 Suppl 1:S62-S73, 2016. doi: 10.1016/j.canep.2016.03.008.

Silva C, Eguiarte LE, Souza V. Reticulated y epidemic population genetic structure of *Rhizobium etli* biovar phaseoli in a traditionally managed locality in Mexico. *Molecular Ecology* 8:277-287, 1999.

Simmons KL, Churchill GA, Aquadro CF. Properties of statistical tests of neutrality for ADN polymorphism data. *Genetics* 141:413-429, 1995.

Simoons-Smit IM, Appelmeik BJ, Verboom T, Negrini R, Penner JL, Aspinall GO, Moran AP, Fei SF, Shi BS, Rudnica W, Savio A, de Graaff J. Typing of *Helicobacter pylori* with monoclonal antibodies against Lewis antigens in lipopolysaccharide. *Journal of Clinical Microbiology* 34: 2196-2200, 1996.

Simpson JT, Durbin R. Efficient de novo assembly of large genomes using compressed data structures. *Genome Research* 22(3):549-56, 2012.

Sinnett CG, Letley DP, Narayanan GL, Patel SR, Hussein NR, Zaitoun AM, Robinson K, Atherton JC. *Helicobacter pylori* vacA transcription is genetically-determined and stratifies the level of human gastric inflammation and atrophy. *Journal of Clinical Pathology* 69(11):968-973, 2016. doi:10.1136/jclinpath-2016-203641.

Smith, JM. Do bacteria have population genetics? Pages 1-12 in: Population genetics of bacteria. E. A. Boumberg (ed.) Cambridge University Press. 1995.

Smith JM, Smith NH, O'Rourke M, Spratt BG. How clonal are bacteria? *Proceedings of the National Academy of Sciences of the United States of America* 90:4384-4388, 1993.

Snelling WJ, Moran AP, Ryan KA, Scully P, McGourty K, Cooney JC, Annuk H, O'Toole PW. *HP0127* is a gastric epithelial cell adhesin. *Helicobacter* 12(3):200-209, 2007. doi: 10.1111/j.1523-5378.2007.00499.x

Snipen L, Almoy T, Ussery DW. Microbial comparative pan-genomics using binomial mixture models. *BMC Genomics* 10:385, 2009. doi: 10.1186/1471-2164-10-385

Snyder L, Champness W. Molecular genetics of bacteria. ASM Press, Washington, USA. 1997.

Souza V, Nguyen TT, Hudson RR, Piñero D, Lenski RE. Hierarchical analysis of linkage disequilibrium in *Rhizobium* populations: evidence for sex? *Proceedings of the National Academy of Sciences of the United States of America* 89:8389-8393, 1992.

Spratt G, Maiden MC. Bacterial population genetics, evolution y epidemiology. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 354:701-710, 1999.

Stenseth NC, Smith JM. Coevolution in ecosystems: Red Queen evolution or stasis? *Evolution* 38, 870-880, 1984.

Stepan N. The Hour of Eugenics: Race, Gender, and Nation in Latin America. Cornell University Press. Ithaca, Nueva York, Estados Unidos. 1991.

Storck SM, Shukla S, Dimitrov S, Bouvet P. Functions of the histone chaperone nucleolin in diseases. *Subcellular Biochemistry* 41: 125–144, 2007.

Suda Y, Kim YM, Ogawa T, Yasui N, Hasegawa Y, Kashihara W, Shimoyama T, Aoyama K, Nagata K, Tamura T, Kusumoto S. Chemical structure and biological activity of a lipid A component from *Helicobacter pylori* strain 206. *Journal of endotoxin research* 7: 95-104, 2001.

Suganuma M, Yamaguchi K, Ono Y, Matsumoto H, Hayashi T, Ogawa T, Imai K, Kuzuhara T, Nishizono A, Fujiki H. TNF- α -inducing protein, a carcinogenic factor secreted from *H. pylori*, enters gastric cancer cells. *International Journal of Cancer* 123:117–122, 2008. doi: 10.1002/ijc.23484.

Suganuma M, Kurusu M, Suzuki K, Nishizono A, Murakami K, Fujioka T, Fujiki H. New tumor necrosis factor- α -inducing protein released from *Helicobacter pylori* for gastric cancer progression. *Journal of Cancer Research and Clinical Oncology* 131 305–313, 2005. doi: 10.1007/s00432-004-0652-x

Suganuma M, Okabe S, Kurusu M, Iida N, Ohshima S, Saeki Y, Kishimoto T, Fujiki H. Discrete roles of cytokines, TNF- α , IL-1, IL-6 in tumor promotion and cell transformation. *International Journal of Oncology* 20:131–136, 2002.

Suganuma M, Okabe S, Marino MW, Sakai A, Sueoka E, Fujiki H. Essential role of tumor necrosis factor α (TNF- α) in tumor promotion as revealed by TNF- α -deficient mice. *Cancer Research* 59: 4516–4518, 1999.

Suganuma M, Kuzuhara T, Yamaguchi K, Fujiki H. Carcinogenic role of tumor necrosis factor- α inducing protein of *Helicobacter pylori* in human stomach. *Journal of biochemistry and molecular biology* 39(1):1–8, 2006.

Sugiyama A, F. Maruta, T. Ikeno, K. Ishida, S. Kawasaki, T. Katsuyama, N. Shimizu, M. Tatematsu, *Helicobacter pylori* infection enhances N-methyl-Nnitrosourea- induced stomach carcinogenesis in the Mongolian gerbil, *Cancer Res.* 58 (1998) 2067–2069.

Sullivan JT, Ronson CW. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proceedings of the National Academy of Sciences of the United States of America* 95:5145-5149, 1998.

Sun W, He J. From isotropic to anisotropic side chain representations: comparison of three models for residue contact estimation. *PLoS One* 6:e19238, 2011.

Suzuki R, Shiota S, Yamaoka Y. Molecular epidemiology, population genetics, and pathogenic role of *Helicobacter pylori*. *Infection, Genetics and Evolution* 12(2):203- 13, 2012.

Taghvaei T, Abadi A, Ghasemzadeh A, Naderi BK, Mobarez AM. Prevalence of horB gene among the Helicobacter pylori strains isolated from dyspeptic patients: first report from Iran. *Internal and Emergency Medicine* 7(6):505-508, 2012.

Tajima, F. Evolutionary relationship of ADN sequences in finite populations. *Genetics* 105:437-460, 1983.

Tajima, F. Statistical method for testing the neutral mutation hypothesis by AND polymorphism. *Genetics* 123:585-595, 1989.

Talarico S, Whitefield SE, Fero J, Haas R, Salama NR. Regulation of Helicobacter pylori adherence by gene conversion. *Molecular Microbiology* 84(6): 1050-61, 2012. doi: 10.1111/j.1365-2958.2012.08073.x

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 30(12):2725-9, 2013.

Tava F, Luinetti O, Ghigna MR, Alvisi C, Perego M, Trespi E, Klersy C, Fratti C, Fiocca R, Solcia E. Type or extension of intestinal metaplasia and immature/atypical 'indefinite-for-dysplasia' lesions as predictors of gastric neoplasia. *Human Pathology* 37: 1489–97, 2006.

Taylor NR. Small world network strategies for studying protein structures and binding. *Computational and Structural Biotechnology Journal* J.5, e201302006, 2013.

Terebiznik MR, Raju D, Vazquez CL, Torbricki K, Kulkarni R, Blanke SR, Yoshimori T, Colombo MI, Jones NL. Effect of Helicobacter pylori's vacuolating cytotoxin on the autophagy pathway in gastric epithelial cells. *Autophagy* 5(3):370–379, 2009.

Terrés AM, Pajares JM. An increased number of follicles containing activated CD69+ helper T cells and proliferating CD71+ B cells are found in H. pylori-infected gastric mucosa. *American Journal of Gastroenterology* 93: 579-583, 1998. doi: 10.1111/j.1572-0241.1998.168_b.x

Tettelin H, Massignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJ, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome" *Proceedings of the National Academy of Sciences of the United States of America* 102:13950–13955, 2005.

Thomas H. The Slave Trade: The Story of the Atlantic Slave Trade: 1440 – 1870. Simon & Schuster. New York, USA. 1997.

Thorell K, Hosseini S, Palacios Gonzales RV, Chaotham C, Graham DY, Paszat L, Rabeneck L, Lundin SB, Nookaew I, Sjöling Å. Identification of a Latin American-specific BabA adhesin variant through whole genome sequencing of *Helicobacter pylori* patient isolates from Nicaragua. *BMC Evolutionary Biology* 16:53, 2016. doi: 10.1186/s12862-016-0619-y.

Thorell K, Yahara K, Berthenet E, Lawson DJ, Mikhail J, Kato I, Mendez A, Rizzato C, Bravo MM, Suzuki R, Yamaoka Y, Torres J, Sheppard SK, Falush D. Rapid evolution of distinct *Helicobacter pylori* subpopulations in the Americas. *PLoS Genetics* 13(2):e1006546, 2017. doi: 10.1371/journal.pgen.1006546.

Thornton R. American Indian Holocaust and Survival: A Population History Since 1492. University of Oklahoma Press. Norman, Oklahoma. 1987.

Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA, Nelson K, Quackenbush J, Zhou L, Kirkness EF, Peterson S, Loftus B, Richardson D, Dodson R, Khalak HG, Glodek A, McKenney K, Fitzgerald LM, Lee N, Adams MD, Hickey EK, Berg DE, Gocayne JD, Utterback TR, Peterson JD, Kelley JM, Cotton MD, Weidman JM, Fujii C, Bowman C, Watthey L, Wallin E, Hayes WS, Borodovsky M, Karp PD, Smith HO, Fraser CM, Venter JC. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388: 539-547, 1997.

Torres J, Pelayo Correa, Catterina Ferreccio, Gustavo Hernandez-Suarez, Rolando Herrero, Maria Cavazza-Porro, Ricardo Dominguez, Douglas Morgan. Gastric cancer incidence and mortality is associated with altitude in the mountainous regions of Pacific Latin America. *Cancer Causes & Control* 24:249–256, 2013. doi: 10.1007/s10552-012-0114-8.

Tosi P, Filipe MI, Luzzi P, Miracco C, Santopietro R, Lio R, Sforza V, Barbini P. Gastric intestinal metaplasia type III cases are classified as low-grade dysplasia on the basis of morphometry. *Journal of Pathology* 169: 73–8, 1993. doi: 10.1002/path.1711690112

Tu S, Bhagat G, Cui G, Takaishi S, Kurt-Jones EA, Rickman B, Betz KS, Penz-Oesterreicher M, Bjorkdahl O, Fox JG, Wang TC. Overexpression of Interleukin-1b induces gastric inflammation and cancer and mobilizes myeloid-derived suppressor cells in mice. *Cancer Cell* 14:408–419, 2008. doi: 10.1016/j.ccr.2008.10.011.

Tummuru MK, Cover TL, Blaser MJ. Mutation of the cytotoxin-associated cagA gene does not affect the vacuolating cytotoxin activity of *Helicobacter pylori*. *Infection and Immunity* 62(6):2609–2613, 1994.

Ubukata H, Nagata H, Tabuchi T, Konishi S, Kasuga T, Tabuchi T. Why is the coexistence of gastric cancer and duodenal ulcer rare? Examination of factors related to both gastric cancer and duodenal ulcer. *Gastric Cancer* 14(1):4–12, 2011. doi:10.1007/s10120-011-0005-9.

Utt M, Danielsson B, Wadström T. *Helicobacter pylori* vacuolating cytotoxin binding to a putative cell surface receptor, heparan sulfate, studied by surface plasmon resonance. *FEMS Immunology and Medical Microbiology* 30, 109–113, 2001.

Vale FF, Vitor JM. Transmission pathway of *Helicobacter pylori*: does food play a role in rural and urban areas? *International Journal of Food Microbiology* 138:1–12, 2010.

Van Doorn LJ, Figueiredo C, Mégraud F, Pena S, Midolo P, Queiroz DM, Carneiro F, Vanderborght B, Pegado MD, Sanna R, De Boer W, Schneeberger PM, Correa P, Ng EK, Atherton J, Blaser MJ, Quint WG. Geographic distribution of *vacA* allelic types of *Helicobacter pylori*. *Gastroenterology* 116(4):823-30, 1999.

Van Valen L. A New Evolutionary Law. *Evolutionary Theory* 1, 1–30, 1973.

Van Valen, L. Molecular evolution as predicted by natural selection. *Journal of Molecular Evolution* 3(2), 89–101, 1974.

van Vliet AH, Kusters JG. Use of alignment-free phylogenetics for rapid genome sequence-based typing of *Helicobacter pylori* virulence markers and antibiotic susceptibility. *Journal of Clinical Microbiology* 53:2877–88, 2015. Doi: 10.1128/JCM.01357-15

Vanaja SK, Rathinam VA, Fitzgerald KA. Mechanisms of inflammasome activation: recent advances and novel insights. *Trends in Cell Biology* 25(5):308–315, 2015. doi:10.1016/j.tcb.2014.12.009.

Viala J, Chaput C, Boneca IG, Cardona A, Girardin SE, Moran AP, Athman R, Memet S, Huerre MR, Coyle AJ, DiStefano PS, Sansonetti PJ, Labigne A, Bertin J, Philpott DJ, Ferrero RL. Nod1 responds to peptidoglycan delivered by the *Helicobacter pylori* *cag* pathogenicity island. *Nature Immunology* 5(11):1166–1174, 2004. doi:10.1038/ni1131.

Vinuesa P, Silva C, Werner D, Martínez-Romero E. Population genetics y phylogenetic inference in bacterial molecular systematics: the roles of migration y recombination in *Bradyrhizobium* species cohesion y delineation. *Molecular Phylogenetics and Evolution* 34:29-54, 2005c. doi:10.1016/j.ympev.2004.08.020

Wall DP, Hirsh AE, Fraser HB, Kumm J, Giaever G, Eisen MB, Feldman MW. Functional genomic analysis of the rates of protein evolution. *Proceedings of the National Academy of Sciences of the United States of America* 102, 5483–5488, 2005. doi:10.1073/pnas.0501761102

Wang S, Ray N, Rojas W, Parra MV, Bedoya G, Gallo C, Poletti G, Mazzotti G, Hill K, Hurtado AM, Camrena B, Nicolini H, Klitz W, Barrantes R, Molina JA, Freimer NB, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Dipierri JE, Alfaro EL, Bailliet G, Bianchi NO, Llop E, Rothhammer F, Excoffier L, Ruiz-Linares A. Geographic patterns of genome admixture in Latin American Mestizos. *PLoS Genetics* 4(3):e1000037, 2008. doi:10.1371/journal.pgen.1000037.

Ward DM. A natural species concept for prokaryotes. *Current Opinion in Microbiology* 1:271-277, 1998.

Warren JR, Marshall BJ. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. *The Lancet* i:1273–1275, 1983.

Watanabe T, Takahashi A, Suzuki K, Kurusu-Kanno M, Yamaguchi K, Fujiki H, Suganuma M. Epithelial-mesenchymal transition in human gastric cancer cell lines induced by TNF- α -inducing protein of *Helicobacter pylori*. *International Journal of Cancer* 134(10):2373-82, 2014. doi: 10.1002/ijc.28582.

Watterson GA. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* 7:256-276, 1975.

Weeks AR, Turelli M, Harcombe WR, Reynolds KT, Hoffmann AA. Rapid evolution of *Wolbachia* in natural populations of *Drosophila*. *PLoS Biology* 5:e114, 2007.

Weeks DL, Eskandari S, Scott DR, Sachs G. A H⁺-gated urea channel: the link between *Helicobacter pylori* urease and gastric colonization. *Science* 287(5452):482–485, 2000. doi: 10.1126/science.287.5452.482

Welch RA1, Burland V, Plunkett G 3rd, Redford P, Roesch P, Rasko D, Buckles EL, Liou SR, Boutin A, Hackett J, Stroud D, Mayhew GF, Rose DJ, Zhou S, Schwartz DC, Perna NT, Mobley HL, Donnenberg MS, Blattner FR. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 99:17020-17024, 2002. doi: 10.1073/pnas.252529799

Wen Y, Feng J, Scott DR, Marcus EA, Sachs G. The HP0165-HP0166 two-component system (ArsRS) regulates acid-induced expression of HP1186 alpha-carbonic anhydrase in *Helicobacter pylori* by activating the pH-dependent promoter. *Journal of Bacteriology* 189(6):2426–2434, 2007. doi:10.1128/JB.01492-06.

Wernegreen J, Riley MA. Comparison of the evolutionary dynamics of symbiotic y housekeeping loci: A case for the genetic coherence of rhizobial lineages. *Molecular Biology and Evolution* 16:98-113, 1999.

Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. RELAX: detecting relaxed selection in a phylogenetic framework. *Molecular Biology and Evolution* 32(3): 820–32, 2015. doi: 10.1093/molbev/msu400.

Wood DW, Setubal JC, Kaul R, Monks DE, Kitajima JP, Okura VK, Zhou Y, Chen L, Wood GE, Almeida NF Jr, Woo L, Chen Y, Paulsen IT, Eisen JA, Karp PD, Bovee D Sr, Chapman P, Clendenning J, Deatherage G, Gillet W, Grant C, Kutuyavin T, Levy R, Li MJ, McClelland E, Palmieri A, Raymond C, Rouse G, Saenphimmachak C, Wu Z, Romero P, Gordon D, Zhang S, Yoo H, Tao Y, Biddle P, Jung M, Krespan W, Perry M, Gordon-Kamm B, Liao L, Kim S, Hendrick C, Zhao ZY, Dolan M, Chumley F, Tingey SV, Tomb JF, Gordon MP, Olson MV, Nester EW. 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science* 294:2317-2323. doi: 10.1126/science.1066804

Xue J, Bai Y, Chen Y, Wang JD, Zhang ZS, Zhang YL, Zhou DY. Expression of *Helicobacter pylori* AlpA protein and its immunogenicity. *World Journal of Gastroenterology* 11(15):2260-2263, 2005. doi: 10.3748/wjg.v11.i15.2260

Yahara K, Furuta Y, Oshima K, Yoshida M, Azuma T, Hattori M, Uchiyama I, Kobayashi I. Chromosome painting in silico in a bacterial species reveals fine population structure. *Molecular Biology and Evolution* 30:1454–1464, 2013.

Yahara K, Kawai M, Furuta Y, Takahashi N, Handa N, Tsuru T, Oshima K, Yoshida M, Azuma T, Hattori M, Uchiyama I, Kobayashi I. Genome-wide survey of mutual homologous recombination in a highly sexual bacterial species. *Genome Biology and Evolution* 4:628–640, 2012.

Yahiro K, Satoh M, Nakano M, Hisatsune J, Isomoto H, Sap J, Suzuki H, Nomura F, Noda M, Moss J, Hirayama T. Low-density lipoprotein receptor-related protein-1 (LRP1) mediates autophagy and apoptosis caused by *Helicobacter pylori* VacA. *Journal of Biological Chemistry* 287, 31104–31115, 2012. doi:10.1074/jbc.M112.387498

Yamaoka Y, Kikuchi S, el-Zimaity HM, Gutierrez O, Osato MS, Graham DY. Importance of *Helicobacter pylori* oipA in clinical presentation, gastric inflammation, and mucosal interleukin 8 production. *Gastroenterology* 123(2): 414-24, 2002. Doi: 10.1053/gast.2002.34781

Yamaoka Y, Orito E, Mizokami M, Gutierrez O, Saitou N, Kodama T, Osato MS, Kim JG, Ramirez FC, Mahachai V, Graham DY. *Helicobacter pylori* in North and South America before Columbus. *FEBS Letters* 517 (1-3):180-4, 2002.

Yamaoka Y. *Helicobacter pylori* typing as a tool for tracking human migration. *Clinical Microbiology and Infection* 15(9):829-34, 2009. doi: 10.1111/j.1469-0691.2009.02967.x.

Yamaoka Y, Kwon DH, Graham DY. A M(r) 34,000 proinflammatory outer membrane protein (oipA) of *Helicobacter pylori*. *Proceedings of the National Academy of Sciences of the United States of America* 97(13):7533-7538, 2000.

Yang I, Woltemate S, Piazuolo MB, Bravo LE, Yopez MC, Romero-Gallo J, Delgado AG, Wilson KT, Peek RM, Correa P, Josenhans C, Fox JG, Suerbaum S. Different gastric microbiota compositions in two human populations with high and low gastric cancer risk in Colombia. *Scientific Reports* 6:18594, 2016. doi: 10.1038/srep18594.

Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: Protein structure and function prediction. *Nature Methods* 12: 7-8, 2015. doi: 10.1038/nmeth.3213

Yun CH, Lundgren A, Azem J, Sjolting A, Holmgren J, Svennerholm AM, Lundin BS. Natural killer cells and *Helicobacter pylori* infection: bacterial antigens and interleukin-12 act synergistically to induce gamma interferon production. *Infection and Immunity* 73(3):1482–1490, 2005.

Zabaleta J, McGee DJ, Zea AH, Hernández CP, Rodríguez PC, Sierra RA, Correa P, Ochoa AC. Helicobacter pylori arginase inhibits T cell proliferation and reduces the expression of the TCR zeta-chain (CD3zeta). *Journal of Immunology* 173: 586-593, 2004.

Zhang J. Evolution by gene duplication: an update. *Trends in Ecology & Evolution* 18(6):292–8, 2003.

Zhang YX, Perry K, Vinci VA, Powell K, Stemmer WP, del Cardayré SB. Genome shuffling leads to rapid phenotypic improvement in bacteria. *Nature* 415 (6872):644-646, 2002. doi: 10.1038/415644a

Zhou T, Drummond DA, Wilke CO. Contact density affects protein evolutionary rate from bacteria to animals. *Journal of Molecular Evolution* 66:395–404, 2008. doi: 10.1007/s00239-008-9094-4.

11. Anexos

11.1. Anexo 1: Estadística de los genomas de *Helicobacter pylori*

11.2. Anexo 2: Análisis Inmuno-Informáticos

11.3. Anexo 3: Artículos científicos

11.3.1. Artículo publicado en Gut Pathogens

11.3.2. Artículo publicado en PeerJ

11.3.3. Artículo sometido a PeerJ

11.4. Productos obtenidos

11.4.1. Participación en congresos

- VII International Symposium of *Helicobacter pylori* infection and Gastric Cancer, 2016, Cartagena Colombia.

Ponencia Oral: Doctora María Mercedes Bravo, Phylogenomics of Colombian isolates of *Helicobacter pylori*.

- VIII International Symposium of *Helicobacter pylori* infection and Gastric Cancer, 2018.
Belo Horizonte, Brasil.

Ponencia Oral: Andrés Julián Gutiérrez Escobar, Emergence of a new phylogenetic group of *H. pylori* in Latin American mestizos.

Póster: Molecular evolution by functional divergence of the p55 binding domain of VacA in Colombian isolates of *Helicobacter pylori*.

Póster: Rapid evolution of AlpA adhesin of *Helicobacter pylori* in Colombia.

11.4.2. Pasantía de investigación

National Institute of Basic Biology. Okazaki, Japan. Octubre - Noviembre, 2017.

Genome Informatics Laboratory

Supervisor: Ikuo Uchiyama, Ichizo Kobayashi.

Friederich-Alexander University of Erlangen, Alemania, Erlangen. Noviembre, 2018.

Bacterial infection Laboratory

Supervisor: Steffen Backert.

11.4.3. Publicaciones

- Gutiérrez-Escobar AJ, Trujillo E, Acevedo O, Bravo MM. Phylogenomics of Colombian *Helicobacter pylori* isolates. Gut Pathogens 9:52, 2017. doi:10.1186/s13099-017-0201-1.

- Gutiérrez-Escobar AJ, Mendez-Casallas G, Acevedo O, Bravo MM. Rapid evolution of the AlpA adhesin of *Helicobacter pylori* from a high gastric cancer gastric risk region from Colombia. PeerJ. 2018 May 25;6: e4846. doi: 10.7717/peerj.4846.

- Gutiérrez-Escobar AJ, Bravo MM, Acevedo O, Backert S. Molecular evolution of the VacA p55 binding domain of *Helicobacter pylori* in mestizos from a high gastric cancer region of Colombia. PeerJ (sometido).