



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
Maestría y Doctorado en Ciencias Bioquímicas

Secuenciación de genomas de *Helicobacter pylori* asociado a
patología gástrica

TESIS

Que para optar por el grado de
Maestro en Ciencias

P R E S E N T A :

José Eduardo Mucito Varela

TUTOR PRINCIPAL:

Dra. Yolanda López Vidal
Facultad de Medicina, UNAM

COMITÉ TUTOR:

Dra. Maricarmen Quirasco Baruch
Facultad de Química, UNAM
Dr. Luis Servín González
Instituto de Investigaciones Biomédicas, UNAM

México, D.F., Febrero, 2014



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

- El presente trabajo fue desarrollado en el Programa de Inmunología Molecular Microbiana, de la Facultad de Medicina de la Universidad Nacional Autónoma de México bajo la Tutoría de la Dra. Yolanda López Vidal y la Co-tutoría del Dr. Gonzalo Castillo Rojas. El proyecto fue financiado por el CONACyT con los proyectos CB2007-01-78787M, SALUD2009-C01-112588 e ICYTDF/150/2012.
- El sustentante fue becario del CONACyT en el programa Becas Nacionales durante el periodo 31 de enero del 2011 al 30 de enero del 2013, y en el programa de Apoyo para Ayudantes de investigador SNI III.
- A la Dra. Yolanda López Vidal por permitirme desarrollarme en su grupo de trabajo, por su confianza, motivación y su asesoría en el diseño experimental.
- Al Dr. Gonzalo Castillo Rojas por sus enseñanzas, su apoyo y asesoría en el diseño experimental.
- A la Dra. Patricia Orduña Estrada por su amistad, por la asesoría y apoyo brindada en el trabajo de experimental.
- A la Dra. Antonia Castillo Rodal por su confianza y motivación.
- Se agradece el apoyo del jurado de exámen por las aportaciones que hicieron a este trabajo con sus comentarios y sugerencias. El jurado estuvo conformado por:
 - Dr. Juan Enrique Morett Sánchez
 - Dr. Ricardo Oropeza Navarro

- Dr. Julio César Carrero Sánchez
 - Dr. Víctor Manuel González Zúñiga
 - Dr. Miguel Angel Cevallos Gaos
-
- Al M. en C. Victor Missael Flores López por el apoyo brindado durante el proceso de ensamble y anotación de los genomas.
 - Al Dr. Gabriel Guarneros Peña por facilitar el uso de su servidor en el Lab. de Genética y Biología Molecular del Centro de Investigación y Estudios Avanzados, del Instituto Politécnico Nacional.
 - Al Doctor Miguel Ángel Cevallos Gaos por su asesoría y apoyo en el mejoramiento del ensamble.
 - Al Doctor Luis F. Lozano Aguirre por su asesoría y apoyo en el mejoramiento del ensamble.
 - A la Dra. Claudia Rangel Escareño por facilitar el uso de su servidor en el Laboratorio de Genómica Computacional del Instituto Nacional de Medicina Genómica.
 - Al Mat. Iván Imaz Rosshandler por su apoyo y asesoría en el uso de herramientas informáticas.
 - A la QFB. Rosa Isabel Amieva Fernández por su apoyo en el cultivo microbiológico de las cepas de *H. pylori* empleadas en este trabajo.
 - A mi amiga Lucy González Villaroel, estudiante de Medicina, por su apoyo incondicional en cada meta lograda y su gran esfuerzo y dedicación en el análisis de datos.
 - Al estudiante de Medicina Joshua Martínez Domínguez por su apoyo en la elaboración de gráficos, figuras y tablas que se presentan en esta tesis.
 - A todos mis compañeros y amigos del laboratorio por hacer más ameno el trabajo diario.

Dedicatoria

A mi familia y amigos...

Índice general

Agradecimientos	II
Dedicatoria	IV
Índice general	VI
Índice de figuras	VII
Índice de tablas	VIII
Resumen	IX
1. Introducción	1
1.1. Epidemiología	1
1.1.1. Adquisición y transmisión de la infección	2
1.2. Patogénesis	3
1.2.1. Factores de Virulencia	3
1.3. Características del genoma de <i>H. pylori</i>	6
1.3.1. Poblaciones de <i>H. pylori</i>	7
1.3.2. Relación entre el origen filogeográfico y patología gástrica	8
2. Justificación	9
3. Hipótesis	10
4. Objetivo general	11
4.1. Objetivos específicos	11

5. Material y métodos	12
5.1. Cepas de <i>H. pylori</i> utilizadas y condiciones de crecimiento	12
5.2. Extracción de DNA genómico	13
5.3. Secuenciación de los genomas	13
5.4. Ensamble de los genomas	14
5.5. Anotación de los genomas “draft”	15
5.6. Análisis de los principales factores de virulencia: <i>cag</i> -PAI, CagA y VacA	16
5.6.1. Análisis de los polimorfismos de VacA	17
5.7. Comparación de los genomas de <i>H. pylori</i> 29CaP y 7C con otros ge- nomas de <i>H. pylori</i>	19
5.7.1. Análisis del genoma “core” de <i>H. pylori</i>	21
5.8. Determinación del origen filogeográfico de las cepas de <i>H. pylori</i> . . .	21
6. Resultados	23
6.1. Secuenciación de genomas de <i>H. pylori</i>	23
6.2. Ensamble de los genomas	23
6.3. Anotación de los genomas “draft”	25
6.4. Análisis de los principales factores de virulencia: <i>cag</i> -PAI, CagA y VacA	26
6.5. Comparación de los genomas de <i>H. pylori</i> 29CaP y 7C con otros ge- nomas de <i>H. pylori</i>	28
6.5.1. Análisis del genoma “core” de <i>H. pylori</i>	29
6.6. Determinación del origen filogeográfico	29
7. Discusión	41
8. Conclusiones	46
A. Bacteriófagos y profagos de <i>H. pylori</i>	47
B. Publicaciones	50
Bibliografía	72

Índice de figuras

6.1.	Anotación del profago de <i>H. pylori</i> 29CaP	30
6.2.	Alineamiento BLAST de los plásmidos de <i>H. pylori</i> contra el plásmido de <i>H. pylori</i> 7C	31
6.3.	Alineamiento BLAST de la isla de patogenicidad <i>cag</i> de <i>H. pylori</i> 26695 contra contigs de <i>H. pylori</i> 29CaP	32
6.4.	Alineamiento BLAST de la isla de patogenicidad <i>cag</i> de <i>H. pylori</i> 26695 contra contigs de <i>H. pylori</i> 7C	33
6.5.	Tipificación de los polimorfismos de la región señal de VacA de <i>H. pylori</i> 29CaP y 7C	34
6.6.	Tipificación de los polimorfismos de la región intermedia de VacA de <i>H. pylori</i> 29CaP y 7C	35
6.7.	Tipificación de los polimorfismos de la región media de VacA de <i>H. pylori</i> 29CaP y 7C	36
6.8.	Alineamiento en MAUVE de los contigs de <i>H. pylori</i> 29CaP y 7C con genomas de <i>H. pylori</i> disponibles en la base de datos de la NCBI.	37
6.9.	Regiones homólogas de los contigs de <i>H. pylori</i> 29CaP y 7C con genomas de <i>H. pylori</i> disponibles en la base de datos de la NCBI	38
6.10.	Representación del genoma "core" de <i>H. pylori</i>	38
6.11.	MLST de <i>H. pylori</i>	39
6.12.	Filogenia de <i>H. pylori</i> generada con las regiones homólogas del alineamiento de genomas completos en Mauve	40
A.1.	Alineamiento de los bacteriófagos y profagos de <i>H. pylori</i> contra la cepa 29CaP	49

Índice de tablas

5.1. Cepas de <i>H. pylori</i> utilizadas en este estudio	12
5.2. Secuencias de VacA de <i>H. pylori</i> utilizadas para determinar los polimorfismos de las cepas de estudio	17
5.3. Secuencias de referencia de las regiones señal e intermedia de VacA	18
5.4. Secuencias genómicas y de plásmidos de <i>H. pylori</i> utilizados en el análisis comparativo	19
6.1. Secuenciación de los genomas de <i>H. pylori</i> 29CaP y 7C	23
6.2. Contigs generados con el ensamble con MIRA	24
6.3. Secuencia “draft” de dos genomas de <i>H. pylori</i> aislado en México	24
6.4. Alineamiento BLAST de contigs concatenados de <i>H. pylori</i> 29CaP con genomas de <i>H. pylori</i>	25
6.5. Alineamiento BLAST de contigs concatenados de <i>H. pylori</i> 7C con genomas de <i>H. pylori</i>	25
6.6. Genomas de <i>H. pylori</i> empleados para anotar los contigs de <i>H. pylori</i> 29CaP	26
6.7. Genomas de <i>H. pylori</i> empleados para anotar los contigs de <i>H. pylori</i> 7C	27
6.8. Número de elementos transferidos con el programa RATT a los contigs de <i>H. pylori</i> 29CaP y 7C	28
A.1. Genomas de bacteriófagos de <i>H. pylori</i> secuenciados	47
A.2. Profagos de <i>H. pylori</i> predichos con PHAST en genomas secuenciados	47

Resumen

H. pylori infecta a más de la mitad de la población mundial y se asocia al desarrollo de gastritis crónica, úlcera péptica y cáncer gástrico. Los factores de virulencia de *H. pylori* más estudiados son la citotoxina vacuolizante (VacA) y la proteína asociada a la citotoxina (CagA), los cuales presentan polimorfismos y determinadas variantes presentan mayor actividad citopática. Sin embargo, la presencia de estos polimorfismos en los factores de virulencia no determina el desarrollo de enfermedad lo que sugiere que otros factores de virulencia desempeñan una función en la patogénesis. Los genomas de *H. pylori* de distintas partes del mundo presentan diferencias relacionadas con su origen filogeográfico que se asocian al desarrollo de mayor daño tisular. En las bases de datos públicas no se cuenta con genomas de cepas de *H. pylori* aisladas en México por lo que en este proyecto se secuenció el genoma de dos cepas de *H. pylori* aisladas en México, una de ellas de un paciente con cáncer gástrico y la otra de un paciente con gastritis crónica. Los genomas se secuenciaron por pirosecuenciación, se ensamblaron *de novo* y se anotaron por transferencia empleando varios genomas de *H. pylori* como referencia. Aunque no se logró cerrar el genoma, se estimó que los “contigs” de los genomas cubren el 92% del genoma. Interesantemente, ninguna de las dos cepas presentaron la isla de patogenicidad *cag*. El análisis reveló la presencia de un profago en la cepa aislada de cáncer gástrico, que es similar a los pocos bacteriófagos de *H. pylori* que se han reportado en la literatura. La cepa aislada de gastritis crónica contiene un plásmido que codifica una copia del elemento de inserción IS605. Ambas cepas poseen ancestría europea, lo que correlaciona con las características genéticas de la población Mexicana actual. No se evidenció la presencia de alguna característica genómica específica de aislamientos de cáncer gástrico empleando 48 genomas de *H. pylori* por lo que se requieren estudios

en donde se analice un mayor número de cepas de diferentes padecimientos gástricos.

Capítulo 1

Introducción

Helicobacter pylori es la bacteria dominante del microbioma gástrico humano y causa la segunda infección bacteriana más común en el mundo [17, 19]. La gastritis inducida por este microorganismo es el principal factor de riesgo para desarrollar cáncer de estómago [17]. La bacteria tiene forma de bacilo curvo o de "S", cuyas dimensiones oscilan entre 2.5-5 x 0.5 μm , también puede presentar formas cocoides en cultivos viejos que se sugiere son formas viables no cultivables y cuya función en el ciclo vital de la bacteria aún no se esclarece. Este microorganismo es móvil debido a la presencia de 4-8 flagelos unipolares o bipolares envainados. Su crecimiento óptimo es a 37°C, en ambiente microaerófilico. Todas las cepas de *H. pylori* reportadas son ureasa, catalasa y oxidasa positivas [71].

El principal reservorio de *H. pylori* es el humano, aunque también es común la infección en primates y fortuitamente se puede aislar de otros mamíferos, tales como felinos silvestres [71].

1.1. Epidemiología

H. pylori infecta a más de la mitad de la población global. La prevalencia de la infección difiere de un país a otro e incluso entre grupos étnicos, sociales o de edad dentro del mismo país. Esto se relaciona principalmente a diferencias en la tasa de adquisición de la bacteria en la infancia y al nivel de vida de las poblaciones. En general, la infección ocurre en la infancia siendo la prevalencia mayor en países en

vías de desarrollo que en países desarrollados [39]. En países en vías de desarrollo, como nuestro país, se estima que el 70-90 % de la población está infectada con *H. pylori*, mientras que en países desarrollados la prevalencia oscila entre 25-50 % [8].

De los individuos infectados, algunos de ellos son capaces de superar la colonización inicial o eliminar la infección establecida mediante mecanismos que aún se desconocen [43]. El 10-20 % de los infectados desarrollan úlcera péptica, del 1-2 % adenocarcinoma gástrico y menos de 1 % linfoma asociado a tejido gástrico linfoide (MALT)[21].

La infección con *H. pylori* aumenta hasta 6 veces el riesgo de desarrollar cáncer gástrico [75]. Este riesgo involucra factores específicos de cada cepa de *H. pylori*, así como características genéticas del hospedero y ambientales [17].

1.1.1. Adquisición y transmisión de la infección

La adquisición de la infección ocurre principalmente en la infancia en países en vías de desarrollo como México [53]. La dosis infectiva para el humano se estima de aproximadamente 10^5 unidades formadoras de colonia (UFC) [31]. En México, Torres y cols. [78], realizaron un estudio en el que demostraron que la seroprevalencia de la infección aumenta conforme avanza la edad de la población. En población menor de un año la seroprevalencia fue del 20 %, mientras que en niños de 10 años fue del 50 % y a partir de los 25 años fue igual a 80 %. El incremento de la incidencia anual medida a través de la seropositividad fue mayor a 6 % en niños menores de 5 años y disminuyó a menos de 3 % en niños de 10-14 años y menos de 0.5 % en individuos de 30-69 años de edad [78]. Estos hallazgos demuestran que la infección por *H. pylori* sucede en los primeros cinco años de vida.

El principal mecanismo de transmisión de la infección por *H. pylori* se considera que ocurre de forma directa (de persona a persona), debido a que *H. pylori* parece tener una amplitud de hospedero muy estrecha, aunque existen evidencias de que el agua, los alimentos o los animales fungen como fuentes de transmisión para adquirir la infección [51, 80].

La ruta de transmisión de persona a persona no está bien establecida, pero se considera que ocurre de dos formas: familiar y/o horizontal. La primera vía de transmisión es la diseminación de la infección entre padres e hijos o entre hermanos y

la segunda vía entre individuos sin relación parental. La vía de transmisión que se considera la predominante es la familiar, con base en estudios que documentaron la diseminación clonal de la infección entre familias [69]. Las rutas más probables de transmisión de la infección por *H. pylori* son: gastro-oral, oral-oral y fecal-oral, con base en las evidencias de detección del microorganismo por cultivo microbiológico o reacción en cadena de la polimerasa (PCR) en el jugo gástrico, el vómito, la saliva y las heces [27] de los individuos infectados.

1.2. Patogénesis

Una vez que el *H. pylori* ingresa al tracto digestivo, se desplaza a la proximidad de la mucosa gástrica y permanece libremente en una delgada capa de moco, donde el pH neutro le es favorable. Esto lo logra a través de su movilidad flagelar y su capacidad quimiotáctica a gradientes de pH. Además, *H. pylori* es capaz de producir hasta el 6 % de su peso húmedo de ureasa, que al adherirse en su superficie mantiene un pH neutro que garantiza su sobrevivencia. Este proceso permite que una parte de la población microbiana se adhiera a la superficie de las células epiteliales de la mucosa mediante proteínas de membrana externa denominadas adhesinas. La adherencia de la bacteria incluye tropismo específico por uniones intercelulares o algunas ocasiones se describe su capacidad de internalización en las células epiteliales [4].

Las interacciones de las diferentes adhesinas y de los factores de virulencia de *H. pylori*, contribuyen al proceso inflamatorio de la mucosa gástrica (gastritis), en la que participan de forma muy importante los neutrófilos e IL-8. La gastritis persiste durante la infección que con gran frecuencia permanece de por vida, a través de alcanzar el equilibrio con el hospedero [17].

1.2.1. Factores de Virulencia

Además de las adhesinas que favorecen la colonización, otros factores de virulencia participan en el desarrollo de la enfermedad. Muchos de estos factores están relacionados con la promoción de respuesta inflamatoria, quizás la respuesta más importante en la infección por *H. pylori* [10]. Dos factores de virulencia de *H. pylori* que se han estudiado ampliamente son: la proteína asociada a la citotoxina (CagA),

codificada por el gen asociado a la citotoxina (*cagA*), presente en isla de patogenicidad *cag* (*cag*-PAI) [74]; y la citotoxina vacuolizante (*VacA*), codificada por el gen *vacA*.

Isla de patogenicidad *cag* (*cag*-PAI) y proteína asociada a la citotoxina (*CagA*)

La isla de patogenicidad *cag* es un segmento de DNA de aproximadamente 40 kilobases (kb) que contiene 31 genes que codifican para los componentes de un sistema de secreción tipo IV (T4SS) y la proteína *CagA* [82]. La isla puede estar dividida en dos regiones por el elemento de inserción IS605, mismo que puede originar la pérdida de una o ambas regiones en algunos aislamientos de *H. pylori* [13].

La proporción de cepas de *H. pylori* que presentan la isla completa varía dependiendo de la región geográfica desde el 60 a 100 % [41, 60]. Se sugiere que la infección con *H. pylori cag*-PAI positivas incrementa el riesgo de desarrollar gastritis crónica, úlcera duodenal y cáncer gástrico [7], aunque esta asociación con la enfermedad es aún controversial ya que la frecuencia de positividad de la isla es similar en pacientes sanos que con patología gástrica [5, 60].

El T4SS codificado en la isla funciona como jeringa molecular para incorporar productos bacterianos a la célula epitelial, que incluyen la proteína *CagA* y el peptidoglucano, además algunos de sus componentes inducen la producción de IL-8 (estimulan inflamación) y de esta manera desempeña un papel importante en el proceso de patogénesis [55]. El genoma de *H. pylori* contiene por lo menos otro T4SS que participa en la competencia natural de la bacteria [14], lo que permite adquirir DNA externo y originar diversidad genética en la misma población de *H. pylori*. La diversificación genética permitiría que *H. pylori* se adapte a un nuevo hospedero después de la transmisión, a diferentes micro-nichos dentro de un mismo hospedero y a condiciones cambiantes en el hospedero durante la infección para lograr su supervivencia en el estómago [21].

CagA es una proteína inmunodominante de tamaño molecular variable entre 121-145 kDa. El gen *cagA* está codificado en el extremo 5' de la isla de patogenicidad. La traslocación de la proteína *CagA* a las células epiteliales gástricas altera las uniones estrechas y adherentes celulares, lo que lleva a una pérdida de la polaridad celular, a la

inducción de la respuesta pro-inflamatoria y efecto mitogénico, que activa múltiples vías de señalización celular que desregulan la respuesta celular oncogénica, lo que aumenta el riesgo de transformación celular y el desarrollo de cáncer, aunado a prolongados períodos de colonización por *H. pylori* [21]. La actividad biológica de CagA se debe principalmente a su capacidad de fosforilación en los motivos EPIYA (Glu-Pro-Ile-Tyr-Ala) presentes en el extremo C-terminal de la proteína, que a su vez se presentan en un número variable dentro de la misma proteína [17].

Cuatro distintos patrones de aminoácidos que contienen los motivos EPIYA dispuestos en el gen *cagA* se denominan A, B, C y D. Los motivos EPIYA-A y EPIYA-B están presentes en todas las cepas de *H. pylori* del mundo. En contraste, EPIYA-C se encuentra predominantemente en las cepas de países occidentales, y EPIYA-D se encuentra en las cepas de países de Asia oriental. Las cepas de *H. pylori* que poseen tres o más motivos EPIYA-C o el motivo EPIYA-D, se asocian con mayor frecuencia al desarrollo de atrofia gástrica, metaplasia intestinal y cáncer gástrico [17].

Citotoxina vacuolizante (VacA)

La citotoxina vacuolizante (VacA) induce la formación de vacuolas citoplasmáticas en cultivos de células epiteliales y se observan variaciones en su actividad vacuolizante, asociadas a la diversidad de su secuencia nucleotídica. Además, VacA estimula la apoptosis en las células epiteliales gástricas, e inhibe la expansión de las células T que se activan por la presentación de los antígenos bacterianos, lo que permite a *H. pylori* evadir la respuesta inmune adaptativa al inducir un tipo de respuesta tolerogénica[17].

VacA está codificada por el gen *vacA* como protoxina (140 kDa), que incluye un péptido señal amino-terminal y un dominio autotransportador carboxilo-terminal. La proteína VacA que se secreta sufre una proteólisis que genera dos fragmentos proteicos; p33 y p55, que son los dominios funcionales. El dominio p33 contiene una secuencia hidrofóbica que está involucrada en la formación de poros en las membranas de las células epiteliales gástricas, mientras que el fragmento p55 contiene dominios de unión a las células. Todas las cepas de *H. pylori* contienen el gen *vacA*, pero existe polimorfismo en sus secuencias nucleotídicas, con las regiones de mayor diversidad localizadas en; la región señal 5' (tipos de alelo s1a, s1b, s1c y s2), localizada en el

péptido señal que dirige el transporte de la proteína a través de la membrana interna de la bacteria; la región media (tipos de alelos m1 y m2), que se localiza en el dominio p55 y que media el tropismo celular; y la región intermedia (tipos de alelo i1, i2 e i3), que se localiza en el dominio p33 y se asocia a la actividad vacuolizante de la proteína [40]. Cada gen *vacA* contiene un solo alelo de región señal, región media y región intermedia, siendo los alelos s1, m1 e i1 los que se asocian con una mayor actividad citotóxica.

1.3. Características del genoma de *H. pylori*

H. pylori fue la primera bacteria de la que se reportaron dos genomas secuenciados de aislamientos independientes [28]. Las cepas fueron *H. pylori* 26695 y J99 [2, 77]. Hoy en día, se han secuenciado 268 genomas de *H. pylori*, de los cuales 50 se encuentran cerrados y 218 como genomas tipo “draft”.

El estudio de los genomas de *H. pylori* reveló una marcada macro diversidad de su secuencia, es decir la presencia de una variación de hasta el 10 % en la secuencia nucleotídica entre genomas de diferentes cepas de *H. pylori* [20], así como la presencia de distinta composición y de arreglo génico. Además, existe mosaicismo genético, con una diversidad alélica que se encuentra entre las más altas observadas en cualquier otra especie bacteriana, debida a una alta tasa de recombinación [28]. El 24 % de los genes de *H. pylori* no tienen similitud de secuencia con los genes disponibles en las bases de datos públicas [2].

El tamaño del genoma de *H. pylori* varía de 1.54-1.67 megabases, con un promedio de 39 % de G+C. Cada cepa de *H. pylori* codifica en su genoma un diferente número de genes, que varía de 1485-1749 marcos de lectura abiertos (por sus siglas en inglés ORFs), de los cuales aproximadamente el 91 % son genes del “core” y el resto (9-10 %) cepa específicos o compartidos entre al menos dos cepas, de acuerdo a la comparación de 20 genomas realizada por Kawai y cols., en 2011 [38]. Los genes cepa específicos probablemente permiten la adaptación de la bacteria a diferentes nichos ambientales, los cuales comprenden hospederos genéticamente distintos [21]. La mayoría de los genes cepa específicos (aprox. 45 %) se concentran en una región cromosomal denominada zona de plasticidad [29]. Esta región contiene un considerable

número de genes de enzimas de restricción-modificación, transposasas, topoisomerasas y proteínas de membrana externa [67].

Algunas cepas de *H. pylori* contienen plásmidos que oscilan entre 1.5 y 40 kilobases [73]. Hasta la fecha se desconoce el papel que desempeñan los plásmidos en el ciclo vital de la bacteria, pero se hipotetiza que éstos son los responsables de la transferencia horizontal de DNA y la recombinación entre cepas de *H. pylori*, ya que son portadores de diferentes regiones cromosomales [2, 34, 64]. El estudio de la replicación, el mantenimiento y la segregación de los plásmidos de *H. pylori* permitirá profundizar el conocimiento en la biología de la bacteria [56].

1.3.1. Poblaciones de *H. pylori*

La diversidad de secuencia nucleotídica en los genes core de *H. pylori* permite correlacionar las cepas de *H. pylori* con diferentes poblaciones humanas étnicas y geográficas, lo que muestra que la infección de *H. pylori* entre humanos cercanamente relacionados ha continuado de manera ininterrumpida por decenas de cientos de años [21]. En general, se distinguen siete poblaciones de *H. pylori* designadas de acuerdo a su origen filogeográfico con base en la comparación de grupos de genes conservados por análisis de secuencias multilocus (MLST). Estas subpoblaciones se denominan hpAfrica1, hpAfrica2, hpEastAsia, hpEurope, hpAsia2, hpNEAfrica y hpSahul [48, 54]. A su vez, hpEastAsia se subdivide en tres subpoblaciones; hspAmerind, hspEAsia y hspMaori. Mientras que hpAfrica1 en dos: hspWAfrica y hspSAfrica [22]. Recientemente, esta clasificación se confirmó al comparar 20 genomas completos de *H. pylori* [38]. En este mismo estudio, se identificó una marcada diferencia en la evolución del genoma entre cepas de *H. pylori* del grupo hspEAsia en comparación con las del grupo hpEurope, especialmente en genes involucrados en la interacción con el hospedero como son; los factores de virulencia (CagA y VacA), proteínas de membrana externa, adhesinas y enzimas de biosíntesis de lipopolisacárido; así como en genes relacionados con la transferencia de electrones, en las reacciones de oxido-reducción y en genes involucrados en la fidelidad de la traducción [38].

1.3.2. Relación entre el origen filogeográfico y patología gástrica

La distribución mundial de la incidencia del cáncer gástrico está cercanamente relacionada a las poblaciones de *H. pylori* definidas por MLST. Una alta incidencia de cáncer se presenta en las regiones donde prevalecen las cepas del grupo hpEastAsia (especialmente del subgrupo hspEAsia). Por otra parte, la incidencia de cáncer gástrico es muy baja en países donde la mayoría de las cepas son hpNEAfrica, hpAfrica1, hpAfrica2 o hpAsia2 [72].

En Latinoamérica, las diferencias geográficas en incidencia de cáncer se documentaron, a pesar de la similitud en la prevalencia de infección con *H. pylori* con genotipos de CagA y VacA asociados con mayor riesgo de cáncer. En Colombia, De Sablet y cols. [18] encontraron que las cepas de *H. pylori* aisladas en regiones de alto riesgo de cáncer gástrico se agrupan mayoritariamente en la población hpEurope y producen un mayor daño histológico, en contraste con las cepas de las regiones de menor prevalencia de cáncer donde la mayoría pertenece al grupo hpAfrica1 y en menor frecuencia al grupo hpEurope, lo que sugiere que el origen filogeográfico determinado por MLST sea un predictor de riesgo de cáncer gástrico.

Capítulo 2

Justificación

Un número creciente de genomas de *H. pylori* de diferentes latitudes y padecimientos clínicos se reportan en las bases de datos públicas, sin embargo no se incluyen aislamientos de *H. pylori* de individuos mexicanos. Por otro lado, los genomas de *H. pylori* han aportado información sobre la evolución de la bacteria ligada a la de las poblaciones humanas, por lo tanto es necesario secuenciar genomas de *H. pylori* aislados en México provenientes de distintas entidades clínicas para incluir las variaciones específicas de las cepas que circulan en el país y estudiar de manera generalizable cuáles son los genes comunes entre las cepas de *H. pylori* que se asocian a cáncer gástrico.

Capítulo 3

Hipótesis

Este trabajo es un proyecto genómico de exploración de cepas de *H. pylori* aisladas en México para tratar de identificar perfiles genómicos que se asocien a patología gástrica y las distinga geográficamente.

Capítulo 4

Objetivo general

Identificar algunos genes de *H. pylori* asociados a cáncer gástrico y describir las características genómicas específicas de las cepas de *H. pylori* aisladas en México.

4.1. Objetivos específicos

- Secuenciar el genoma de las cepas de *H. pylori* 29CaP y 7C
- Ensamblar *de novo* los genomas
- Anotar los genomas obtenidos
- Determinar el origen filogeográfico de las cepas estudiadas
- Comparar los genomas de las cepas secuenciadas entre sí y con los genomas disponibles de otras cepas de *H. pylori*

Capítulo 5

Material y métodos

5.1. Cepas de *H. pylori* utilizadas y condiciones de crecimiento

Dos cepas de *H. pylori* aisladas de biopsias gástricas de sujetos mexicanos se seleccionaron para su secuenciación. Ambas cepas se aislaron de estudios realizados previamente en el laboratorio de individuos cuyo criterio de inclusión fue que hubieran nacido en México y cuyos padres tuvieran al menos dos generaciones residiendo en el país. Las cepas se encontraban conservadas en caldo *Brucella*-glicerol 10 % a -70°C , se cultivaron en placas de agar *Haemophilus* Test Medium (HTM) suplementado con 8 % de suero de caballo y se incubaron a 37°C en atmósfera de 5 % de CO_2 . La pureza de las cepas se verificó por observación de la morfología microscópica en un frote teñido por Gram y la presencia de una sola morfología colonial en el crecimiento en placa, además producción de ureasa, catalasa y oxidasa. Las características de las cepas se resumen en la Tabla 5.1.

Para la obtención de biomasa una colonia de cada cepa se cultivó por estría masiva en tres placas de medio HTM y se incubaron en las condiciones antes mencionadas.

Tabla 5.1: Cepas de *H. pylori* utilizadas en este estudio

Cepas de <i>H. pylori</i>	29CaP	7C
Padecimiento de origen	Cáncer gástrico avanzado	Gastritis crónica
Susceptibilidad a antibióticos	Sensible	Multi-resistente

5.2. Extracción de DNA genómico

El DNA genómico de ambas cepas se obtuvo con el método de Bromuro de cetiltrimetilamonio (CTAB) [83], previamente estandarizado. En breve, la biomasa cultivada de tres placas para cada cepa se concentró y se resuspendió en 600 μL de una solución de TE [Tris-HCl 50 mM, pH8; EDTA 25 mM]. Las bacterias se lisaron por la adición de 30 μL de SDS 10 % (p/v), 3 μL de proteinasa k (20 mg/mL), 20 μL de RNasa (10 mg/mL) y se incubaron a 37°C durante una hora. Enseguida, los polisacáridos y las proteínas del lisado bacteriano se removieron mediante la adición de 106 μL de NaCl 5 M y 80 μL de solución CTAB/NaCl (CTAB 10 %, NaCl 0.7 M), a 65°C durante un minuto. A continuación, se removieron las proteínas con 800 μL de cloroformo-alcohol isoamílico (24:1). La fase acuosa se recuperó y se mezcló con un volumen igual de fenol-cloroformo-alcohol isoamílico (25:25:1). Posteriormente, la mezcla se centrifugó 5 minutos a 10,000 rpm. Después se recuperó el sobrenadante y se extrajo una vez más con cloroformo-alcohol isoamílico (24:1) para remover remanentes de fenol y/o CTAB. Finalmente, el DNA se precipitó con 700 μL de isopropanol frío y se concentró por centrifugación a 12,000 rpm durante 5 minutos. La pastilla se resuspendió con 1 mL de etanol al 70 % (v/v), se centrifugó a 8,000 rpm durante 5 minutos, se favoreció su evaporación a temperatura ambiente y se resuspendió en 110 μL de solución TE. La calidad del DNA se cuantificó y se estimó espectrofotométricamente en el equipo NanoDrop 1000 (Thermo Scientific). La integridad del DNA se verificó por electroforesis en un gel de agarosa al 0.8 %.

5.3. Secuenciación de los genomas

El DNA genómico de *H. pylori* se secuenció por pirosecuenciación en la unidad de secuenciación del LANGEBIO (Cinvestav, IPN) con el equipo 454 Titanium (Roche). Se utilizó una placa completa de secuenciación para obtener lecturas de un sólo extremo o “single reads” de cada genoma de *H. pylori*, debido a políticas de la unidad de secuenciación que limita el uso de fracciones de placa dependiendo de los trabajos en espera. LANGEBIO obsequió una corrida adicional de media placa en la plataforma XL+ para la cepa 29CaP. En breve, el DNA se fragmentó por nebulización y se construyó una librería con fragmentos ligados a adaptadores complementarios a

oligonucleótidos inmovilizados en microesferas. Una amplificación de los fragmentos de DNA se realizó mediante PCR en emulsión en las microesferas. Las microesferas con los fragmentos amplificados se inmovilizaron en placas donde se llevó a cabo la reacción de pirosecuenciación que consiste en la síntesis de los segmentos de DNA mediante la adición secuencial de desoxiribonucleótidos y la detección de la incorporación de un nucleótido mediante la hidrólisis del pirofosfato que genera una reacción luminosa catalizada por el complejo enzimático luciferasa-sulfurilasa presente en las placas de reacción.

5.4. Ensamble de los genomas

Las lecturas obtenidas de la pirosecuenciación de los genomas de las cepa *H. pylori* 29CaP y 7C se ensamblaron con el software MIRA (acrónimo en inglés de ensamble inteligente mimetizado de lecturas) versión 3.4.0 [15](Ver tablas 6.1 y 6.2). MIRA selecciona regiones de nucleótidos de alta calidad para cada lectura, luego compara cada lectura contra todas las lecturas (y su cadena complementaria reversa) con un algoritmo rápido y tolerante a sustituciones de bases para detectar empalmes potenciales, estos son aceptados como verdaderos si cumplen criterios de calidad establecidos en el algoritmo y los ensambla en “contigs” de manera iterativa adicionando lecturas hasta que no existan más posibilidades de empalme. Como paso final, los “contigs” se extienden gradualmente mediante el solapamiento con otros de ellos con un mínimo de error [15].

El ensamble se realizó con las opciones ajustadas para lecturas de 454 con los parámetros siguientes: método=denovo, tipo=genoma, calidad=accurate en el comando `mira -project=genome_assembly -job=denovo,genome,accurate,454`.

El ensamble de las lecturas obtenidas en la segunda secuenciación del genoma de la cepa 29CaP se realizó por mapeo de las lecturas empleando como molde los contigs obtenidos del ensamble las primeras lecturas, con los parámetros método=mapping y la opción -abnc:yes, para construir nuevos contigs.

Los contigs obtenidos en el ensamble con MIRA se sometieron a ensamble con el software Velvet [85] con el tamaño de “hash” ajustado a 191 para extender los contigs. El ajuste del “hash” se realizó con el programa `optimize_assembly` escrito en lengua-

je bash (cortesía del Lab. De Genética y Biología Molecular, Cinvestav-Zacatenco). Velvet se ejecutó con los comandos `velveth . 191 -fasta -long reads.fna` y `velvetg . -read_trkg yes -amos_file yes -exp_cov auto -cov_cutoff auto -long_mult_cutoff 0`. Velvet primero lee las lecturas y construye un diccionario de todas las palabras de tamaño k , donde k es el tamaño del “hash” definido por el usuario, y define alineamientos locales exactos entre las lecturas. Velvetg lee los alineamientos y construye una gráfica de de Bruijn, remueve errores y finalmente procede a simplificar la gráfica y resolver repetidos [84]. Los errores sin resolver automáticamente con los parámetros establecidos en el algoritmo del programa, se resolvieron mediante la edición manual en los archivos de texto de los “contigs”. La corrección de errores se realizó por consenso o mediante la comparación con los genomas secuenciados de *H. pylori* visualizando el ensamble con el programa Tablet versión 1.12.02.06 [52].

El ensamble se verificó mediante la comprobación de la cobertura continua y de profundidad mayor a 10x en un alineamiento de las lecturas originales de 454 con los contigs obtenidos en el ensamble con Velvet. Para estimar la cobertura del genoma de *H. pylori* cubierta por los contigs, los contigs de cada cepa se concatenaron como si se tratara de una sola secuencia y se realizó un alineamiento BLAST contra los genomas de referencia de *H. pylori* en la base de datos "genome" de la NCBI (<http://www.ncbi.nlm.nih.gov/genome/>).

5.5. Anotación de los genomas “draft”

Se realizó un alineamiento en BLAST de los “contigs” de las cepas *H. pylori* 29CaP y 7C contra los genomas de *H. pylori* disponibles en la base de datos “genome” de la NCBI. Se seleccionaron los genomas de *H. pylori* que tuvieron el mejor alineamiento para cada uno de los “contigs” de las cepas estudiadas y se emplearon para realizar la anotación con el programa RATT (acrónimo en inglés de herramienta de anotación rápida por transferencia) [58] el cual busca bloques sinténicos entre la referencia y la secuencia blanco mediante el algoritmo de NUCmer dentro del paquete Mummer3 [42] y ubica las posiciones de las anotaciones dentro de estos bloques, las transfiere corrigiendo los cambios en el marco de lectura y los codones de inicio y de paro en la nueva anotación.

Para la búsqueda de profagos, se utilizó el programa PHAST (acrónimo en inglés de herramienta de búsqueda de fagos) [87] (disponible en <http://phast.wisharlab.com>), las predicciones y las anotaciones generadas se compararon con las transferidas con RATT, si la anotación no se transfirió con el programa RATT, ésta se generó manualmente en el programa Artemis [68].

5.6. Análisis de los principales factores de virulencia: *cag*-PAI, CagA y VacA

Las secuencias de las proteínas de la isla de patogenicidad *cag* y de su proteína efectora CagA, así como la secuencia de VacA se buscaron en los “contigs” anotados de las cepas *H. pylori* 29CaP y 7C, empleando el programa Artemis [68] para visualizar las anotaciones de los contigs. Se identificaron únicamente las secuencias de la proteína VacA de ambas cepas, las cuales se guardaron en formato FASTA para determinar los tipos de polimorfismos que poseen.

Debido a que no se encontraron las secuencias de la isla de patogenicidad ni de la proteína CagA, se realizó la búsqueda de sus secuencias en los contigs de cada una de las cepas estudiadas mediante alineamientos en BLAST+ version 2.2.28 [3]. Como referencia se tomó la secuencia de la isla de patogenicidad *cag* del genoma de la cepa *H. pylori* 26695, en los extremos se incluyó la secuencia de los genes que flanquean la isla (HP0519 y glutamato racemasa). El alineamiento contra la secuencia nucleotídica de la isla completa se realizó con el algoritmo “blastn”, el alineamiento contra las secuencias de aminoácidos de los genes de la isla, incluyendo la proteína CagA y de los dos genes que la flanquean se realizó mediante el algoritmo “blastx”.

Adicionalmente, se alinearon todas las lecturas obtenidas en las corridas de secuenciación contra la región completa de la isla de patogenicidad, con el fin de evaluar si la ausencia de ésta pudiera ser un error en el ensamble. Para este análisis se utilizó el algoritmo SMALT [63], tomando como parámetros para el índice de la secuencia de referencia un tamaño de k-mero de 13, “skip” 1 y para las lecturas mapeadas una calidad mayor o igual a Q10 para incluir lecturas que probablemente no se hayan incluido en el ensamble.

5.6.1. Análisis de los polimorfismos de VacA

Se realizó un alineamiento múltiple con el algoritmo de clustalW2 [45] de las secuencias de aminoácidos de la proteína VacA de ocho cepas de *H. pylori* disponibles en la base datos de la NCBI y cuyos polimorfismos se encuentran reportados en la literatura. Los números de acceso de las secuencias utilizadas se enlistan en la Tabla 5.2. Las secuencias de VacA de las cepas 29CaP y 7C se incluyeron en el alineamiento, además de las secuencias de las regiones señal e intermedia reportadas en la literatura [6, 65], las cuales se consideraron como secuencias de referencia (Tabla 5.3). Se consideró como región media la secuencia a partir del aminoácido 270, de acuerdo a las secuencias disponibles de cepas caracterizadas.

Se hizo un análisis filogenético con el paquete para inferencia de filogenias, PHYLIP versión 3.695 [23]. En resumen, se calculó una matriz de distancia con el programa *protdist* usando el parámetro de Kimura 2. Las relaciones filogenéticas se infirieron por el método de Neighbor-joining y se desplegaron árboles filogénicos con el programa SeaView versión 4 [30]. Los polimorfismos de VacA de las cepas 29CaP y 7C se establecieron por su similitud con las secuencias de referencia en los alineamientos múltiples y por su agrupación con las secuencias de referencia y filogramas generados.

Tabla 5.2: Secuencias de VacA de *H. pylori* utilizadas para determinar los polimorfismos de las cepas de estudio

Cepa	Tipo de polimorfismo			No. de acceso
	Región señal	Región intermedia	Región media	
60190	s1a	*	m1	U05676
F57	s1a	i1	m1	BAJ59976
F32	s1a	i1	m1	BAJ58044
HPAG1	s1b	i1	m1	ABF84934
G27	s1b	i1	m1	ACI27595
Cuz20	s1a	i2	m2	ADO04098
Tx30a	s2	*	m2	U29401
B38	s2	i2	m2	CAX29398

*= No reportado

Tabla 5.3: Secuencias de referencia de las regiones señal e intermedia de VacA

Región	Polimorfismo	Secuencia
Señal	s1a	MEIQQTHRRKINRPLVSLALVGAIVSTTPQQSHAAF
	s1b	MEIQQTHRRKINRPLVSLVLA GALISAIPQESHAAF
	s2	MEIQQTHRRKINRPLSLALVGVLMGTELGANTPNDPIHSESRAF
Intermedia	i1	GGDLDVNMQKATLRLGQFNGNSFTSYkdsadrttrvqFNAKNVLIDNPLEINNIRV GSGAGRKASSTVLTLLqaeqfsskNAEISLYDGATLNLssisvldngnVWMGRLQYVGAYL
	i2	GGDLDVNMQKATLRLGQFNGNSFTSIRkgaurtrvqFNAKNVLIDNVEINNIRV GSGAGRKASSTVLTLLksselktsrseNAEISLYDGATLNLvssisqsvdlygkVWMGRLQYVGAYL
	i1/i2	GGDLDVNMQKATLRLGQFNGNSFTSIRkdgaurtrvqFNAKNVLIDNVEINNIRV GSGAGRKASSTVLTLLqsselktsrseNAEISLYDGATLNLvssshslsvdlygkVWMGRLQYVGAYL

5.7. Comparación de los genomas de *H. pylori* 29CaP y 7C con otros genomas de *H. pylori*

Se realizó un alineamiento múltiple de genomas con el programa Mauve [16], empleando los contigs de los genomas secuenciados con 46 genomas completos de *H. pylori* y sus plásmidos, disponibles en la base de datos “genome” de la NCBI (<http://www.ncbi.nlm.nih.gov/genome>), ver Tabla 5.4. La presencia de bloques sinténicos entre todos los genomas de *H. pylori* se analizó y se identificaron los contigs de las cepas de estudio con similitud a secuencias de los plásmidos de *H. pylori*.

Las secuencias plasmídicas de *H. pylori* se alinearon mediante el algoritmo BLAST contra el contig de *H. pylori* 7C correspondiente al plásmido de esta cepa y se generó un gráfico comparativo con el programa generador de imágenes de BLAST en anillo (BRIG) versión 0.95 [1].

Tabla 5.4: Secuencias genómicas y de plásmidos de *H. pylori* utilizados en el análisis comparativo

Cepa de <i>H. pylori</i>	No de acceso GeneBank	
	Genoma	Plásmido
B8	NC_014256.1	NC_014257.1
35A	NC_017360.1	
PeCan4	NC_014555.1	NC_014556.1
SJM180	NC_014560.1	
26695	NC_000915.1	
J99	NC_000921.1	
HPAG1	NC_008086.1	NC_008087.1
Shi470	NC_010698.2	
G27	NC_011333.1	NC_011334.1
P12	NC_011498.1	NC_011499.1
B38	NC_012973.1	
ELS37	NC_017063.1	NC_017064.1

Continúa en la siguiente página

Continuación de la Tabla 5.4

Sat464	NC_017359.1	NC_017356.1
Lithuania75	NC_017362.1	NC_017363.1
Gambia94/24	NC_017371.1	NC_017364.1
Puno120	NC_017378.1	NC_017377.1
SNT49	NC_017376.1	NC_017380.1
v225d	NC_017355.1	NC_017383.1
52	NC_017354.1	
908	NC_017357.1	
Cuz20	NC_017358.1	
SouthAfrica7	NC_017361.1	NC_017373.1
F30	NC_017365.1	NC_017369.1
F32	NC_017366.1	NC_017370.1
F57	NC_017367.1	
F16	NC_017368.1	
India7	NC_017372.1	
83	NC_017375.1	
Puno135	NC_017379.1	
51	NC_017382.1	
Shi417	NC_017739.1	
Shi112	NC_017741.1	
Shi169	NC_017740.1	
PeCan18	NC_017742.1	
HUP-B14	NC_017733.1	NC_017734.1
XZ274	NC_017926.1	NC_017919.1
Rif1	NC_018937.1	
Rif2	NC_018938.1	
Aklavik117	NC_019560.1	NC_019562.1,NC_019561.1
Aklavik86	NC_019563.1	NC_019565.1,NC_019564.1
OK113	NC_020508.1	
OK310	NC_020509.1	NC_020556.1

Continúa en la siguiente página

Continuación de la Tabla 5.4	
UM032	NC_021215.1
UM299	NC_021216.1
UM037	NC_021217.1
UM066	NC_021218.1
HPM186	AF077006

5.7.1. Análisis del genoma “core” de *H. pylori*

Se determinaron grupos de ortólogos entre las proteínas de los 46 genomas completos de *H. pylori* y de las cepas secuenciadas, con el programa OrthoMCL [47] como se describe en la literatura [25, 44]. Un alineamiento con “blastp” de todas las secuencias contra todas se realizó con valor de corte de E de 10^{-5} y éste se guardó en formato de tabla. Luego, los resultados del alineamiento se cargaron a una base de datos de MySQL y se buscaron pares de proteínas con similitud significativa con el software OrthoMCL, los cuales se agruparon mediante el algoritmo de agrupamiento de Markov (MCL, disponible en: <http://www.micans.org/mcl>). La lista de grupos de ortólogos obtenidos con OrthoMCL se analizó con programas escritos en “perl” para localizar grupos comunes a todas las cepas “genoma core” y grupos compuestos por al menos dos genomas. Las secuencias de proteínas que no se asignaron a ningún grupo de ortólogos fueron consideradas cepa específicas.

Para tratar de identificar genes asociados a cáncer gástrico se realizó la búsqueda de grupos de ortólogos que incluyeran sólo los aislamientos de cáncer gástrico.

5.8. Determinación del origen filogeográfico de las cepas de *H. pylori*

La secuencia de los genes *atpA*, *efp*, *mutY*, *ppa*, *trpC*, *ureI* y *yphC* se obtuvo a partir de los contigs anotados de las cepas de *H. pylori* 29CaP y 7C. Estas secuencias se concatenaron y se combinaron con las secuencias de los mismos genes concatenadas

de 1,551 cepas disponibles en la base de datos de *Helicobacter* PubMLST (disponible en: <http://pubmlst.org/helicobacter/>).

Un alineamiento múltiple con el algoritmo de clustalW2 [45] se realizó con las secuencias concatenadas de los genes anteriormente mencionados. Este alineamiento se utilizó para inferir una filogenia con el programa Seaview versión 4 [30], se utilizó como método para medir la distancia el parámetro de Kimura2 y se agrupó mediante el método de Neighbor-joining. Se utilizó un valor de "bootstrap" de 100 para verificar la topología del árbol. El gráfico se generó en lenguaje de programación R, mediante programas escritos en el laboratorio que utilizan los paquetes APE [59] y GEIGER [33].

Adicionalmente, se comparó la filogenia realizada por MLST con el árbol generado con el alineamiento de 48 genomas de *H. pylori* en Mauve.

Capítulo 6

Resultados

6.1. Secuenciación de genomas de *H. pylori*

Más de un millón de lecturas se obtuvieron de cada uno de los genomas de *H. pylori* estudiados. La longitud promedio de las lecturas fue mayor de 300 bases, siendo las de mayor longitud las obtenidas para la cepa 29CaP con la plataforma XL+ (Tabla 6.1).

Tabla 6.1: Secuenciación de los genomas de *H. pylori* 29CaP y 7C

Cepa	29CaP		7C
Plataforma	Titanium	XL+	Titanium
No. de bases	458 194 053	274 175 257	617 214 966
No. de Lecturas	1 293 574	583 810	1 433 388
Longitud promedio(pb)	354.63	469.63	430.60
Cobertura teórica(veces)	458		386

pb = Pares de bases

6.2. Ensamble de los genomas

Como se observa en la Tabla 6.2, en promedio el 98 % de las lecturas obtenidas de la secuenciación de cada genoma se ensamblaron en “contigs” por MIRA. La profundidad que se obtuvo fue mayor a 100 veces pero menor a la estimada teóricamente (Tabla 6.1) .

Tabla 6.2: Contigs generados con el ensamble con MIRA

Cepa	29CaP	7C
Lecturas ensambladas	1 265 412 *	1 428 173
(%)	97.8	99.6
Cobertura promedio	115.74	178.34
Número de contigs	1623	572
Longitud maxima	53 090	68 964
N50	10 493	22 042

* Correspondiente a la primera corrida de secuenciación

Los contigs obtenidos con MIRA fueron ensamblados con Velvet para extender la secuencia de los mismos, la secuencia de los contig extendidos fue verificada alineando las lecturas originales filtradas con calidad mínima Q25, se observó que la cobertura de los contigs fuera continua y que la secuencia correspondiera al alineamiento consenso. El número de contigs de cada genoma se redujo considerablemente resultando en contigs de mayor de longitud, la combinación de estos contigs corresponden a la longitud promedio de un genoma de *H. pylori* (Tabla 6.3).

El alineamiento BLAST de los contigs concatenados de las cepas de *H. pylori* 29CaP y 7C con cinco genomas de *H. pylori* disponibles en base de datos de la NCBI contra los que se presentó el mejor alineamiento, permite estimar que la cobertura de los contigs cubre del 89 al 92% del genoma como se muestra en las Tablas 6.4 y 6.5.

Tabla 6.3: Secuencia “draft” de dos genomas de *H. pylori* aislado en México

Cepa	29CaP	7C
Número de contigs	17	39
Longitud combinada	1 650 830	1 624 053
N50	163 794	80 417
Longitud mínima	2 708	68
Longitud máxima	588 708	265 122

Tabla 6.4: Alineamiento BLAST de contigs concatenados de *H. pylori* 29CaP con genomas de *H. pylori*

Cepa	Padecimiento/Origen	Max Score	Total Score	% Cobertura	E-val	% Identidad
B38	Linfoma/Francia	1.248e5	2.525e6	91	0.0	96
ELS37	Cáncer gástrico/El Salvador	1.047e5	2.453e6	91	0.0	96
HUP-B14	No especificado/España	1.046e5	2.401e6	89	0.0	96
G27	No especificado/Italia	76718	2.480e6	92	0.0	95
B8	Gastritis severa en Gerbillo / Cepa parental úlcera gástrica de humano	73233	2.491e6	92	0.0	95

Tabla 6.5: Alineamiento BLAST de contigs concatenados de *H. pylori* 7C con genomas de *H. pylori*

Cepa	Padecimiento/Origen	Max Score	Total Score	% Cobertura	E-val	% Identidad
B38	Linfoma/Francia	1.227e5	2.523e6	94	0.0	96
B8	Gastritis severa en Gerbillo/Cepa parental úlcera gástrica de humano	1.214e5	2.479e6	94	0.0	96
Lituania75	No especificado/Europa	1.063e5	2.439e6	93	0.0	96
P12	Úlcera duodenal/Alemania	70181	2.497e6	94	0.0	96
26695	Gastritis/Reino Unido	64585	2.498e6	93	0.0	95

6.3. Anotación de los genomas “draft”

Todos los “contigs” de las cepas de *H. pylori* 29CaP y 7C tuvieron similitud con los genomas de *H. pylori*. Para la anotación por transferencia se emplearon los genomas de *H. pylori* que se presentan en las Tablas 6.6 y 6.7. Estos genomas se eligieron ya que el alineamiento en BLAST de los “contigs” presentó el máximo grado de similitud.

El número de genes transferidos y anotados con RATT para cada conjunto de contigs se resume en la Tabla 6.8. La anotación con el programa PFAST reveló la presencia de una región de 31.7 kb en un contig de *H. pylori* 29CaP, predicha como la secuencia completa de un profago, de acuerdo al puntaje asignado por el algoritmo. Se anotaron 45 secuencias codificantes de las cuales 38 (95%) tienen similitud con proteínas de bacteriófagos de *H. pylori* u otros géneros bacterianos y 7 tienen similitud con proteínas codificadas en algunos genomas de *H. pylori* o el profago de *Helicobacter acynonichis*. Adicionalmente, se predijeron cuatro probables sitios de inserción (att) dos attL y dos attR. En la figura 6.1 se representa la región del profago con las anotaciones relevantes.

Por otra parte, un contig de *H. pylori* 7C se anotó como un plásmido circular de 6,835 bases que codifica a 11 proteínas, incluidas las trasposasas del elemento de inserción IS605, una replicasa, los genes de movilización *mobA-D* y cuatro proteínas hipotéticas. El plásmido presentó similitud con los plásmidos de *H. pylori* reportados

en la base de datos “genome” de la NCBI (Figura 6.2).

Tabla 6.6: Genomas de *H. pylori* empleados para anotar los contigs de *H. pylori* 29CaP

Contig	Longitud (pb)*	Genoma	Score	% Identidad	E-val
29CaP_contig_1	204 216	HUP-B14	5.490×10^4	95	0.0
29CaP_contig_2	163 794	Lithuania75	8.907×10^4	95	0.0
29CaP_contig_3	588 708	B38	1.248×10^5	96	0.0
29CaP_contig_4	27 829	India7	4881	86	0.0
29CaP_contig_5	118 617	G27	3.879×10^4	95	0.0
29CaP_contig_6	104 476	F32	4.709×10^4	95	0.0
29CaP_contig_7	138 677	UM037	5.270×10^4	95	0.0
29CaP_contig_8	48 200	HPAG1	5.127×10^4	96	0.0
29CaP_contig_9	2708	B38	3275	93	0.0
29CaP_contig_10	100 497	B38	4.059×10^4	96	0.0
29CaP_contig_11	6999	HPAG1	4748	93	0.0
29CaP_contig_12	32 211	SNT49	4.599×10^4	95	0.0
29CaP_contig_13	48 031	26695	6.940×10^4	96	0.0
29CaP_contig_14	20 309	India7	2.895×10^4	94	0.0
29CaP_contig_15	10 018	PeCan18	1.744×10^4	99	0.0
29CaP_contig_16	27 550	ELS37	2.500×10^4	99	0.0
29CaP_contig_17	7990	Gambia94/24	8067	99	0.0

* pb = pares de bases

6.4. Análisis de los principales factores de virulencia: *cag*-PAI, CagA y VacA

La isla de patogenicidad *cag* y el gen que codifica la proteína efectora CagA, no se encontraron presentes en ninguno de los “contigs” de los genomas de *H. pylori* 29CaP o 7C. El alineamiento BLAST de la secuencia de la isla completa de la cepa 26695 mostró que únicamente se encuentran los genes que flanquean la isla (Figuras 6.3 y 6.4). El alineamiento de las CDS o de los genes de la isla contra los “contigs” tampoco mostró alineamientos significativos que revelaran la presencia de la *cag*-PAI.

Para corroborar que la ausencia de la *cag*-PAI no se tratara de un mal ensamble, se alinearon las lecturas de las corridas de secuenciación contra la secuencia de la *cag*-PAI de *H. pylori* 26695. El alineamiento mostró que hay cobertura completa de dos genes que flanquean la isla y la presencia de lecturas con alineamientos menores

Tabla 6.7: Genomas de *H. pylori* empleados para anotar los contigs de *H. pylori* 7C

Contig ¹	Longitud (pb) ²	Genoma	Score	% Identidad	E-val
7C_contig_1	59 084	UM299	7.219×10^4	94	0.0
7C_contig_2	119 977	Rif2	4.200×10^4	95	0.0
7C_contig_3	265 122	Rif2	6.459×10^4	95	0.0
7C_contig_4	80 281	B38	7.545×10^4	96	0.0
7C_contig_5	111 726	SJM180	8.159×10^4	96	0.0
7C_contig_6	60 019	Lithuania75	6.524×10^4	95	0.0
7C_contig_7	49 363	ELS37	3.295×10^4	95	0.0
7C_contig_8	33 247	B38	4.111×10^4	96	0.0
7C_contig_9	91 805	ELS37	9.393×10^4	96	0.0
7C_contig_10	56 377	HPAG1	3.686×10^4	96	0.0
7C_contig_11	98 368	Rif2	3.994×10^4	95	0.0
7C_contig_12	23 129	ELS37	3.751×10^4	96	0.0
7C_contig_13	31 450	P12	2.999×10^4	95	0.0
7C_contig_14	23 628	P12	3.061×10^4	95	0.0
7C_contig_15	58 902	ELS37	4.878×10^4	95	0.0
7C_contig_16	81 398	B38	1.227×10^5	96	0.0
7C_contig_17	80 417	B38	5.966×10^4	96	0.0
7C_contig_18	56 225	HPAG1	7.163×10^4	96	0.0
7C_contig_19	20 649	Aklavik117	1.165×10^4	95	0.0
7C_contig_20	53 808	UM037	4.960×10^4	95	0.0
7C_contig_21	6959	B8	5308	94	0.0
7C_contig_22	27 376	G27	4.236×10^4	96	0.0
7C_contig_23	35 294	Shi470	3.264×10^4	95	0.0
7C_contig_24	18 919	G27	2.196×10^4	96	0.0
7C_contig_25	26 759	Puno120	1.855×10^4	92	0.0
7C_contig_26	12 397	Shi417	1.600×10^4	98	0.0
7C_contig_27	4711	B38	7103	95	0.0
7C_contig_28	12 921	B38	1.771×10^4	96	0.0
7C_contig_29	7583	plasmid Aklavik86	3182	89	0.0
7C_contig_30	2063	P12	2998	95	0.0
7C_contig_31	2501	B38	3121	92	0.0
7C_contig_32	2679	Aklavik117	3240	94	0.0
7C_contig_33	3014	B38	5035	97	0.0
7C_contig_34	2230	HPAG1	4063	99	0.0
7C_contig_35	1111	SJM180	1989	99	0.0
7C_contig_36	1620	P12	2693	97	0.0
7C_contig_37	741	Aklavik117	1306	98	0.0

¹ Los contigs 7C_contig_38 (132 pb) y 39 (68 pb) no presentaron similitud con ningún genoma de referencia o alguna otra secuencia

² pb = pares de bases

Tabla 6.8: Número de elementos transferidos con el programa RATT a los contigs de *H. pylori* 29CaP y 7C

Cepa	29CaP	7C
Genes	1518	1512
CDS	1415	1446
tRNA	34	34
rRNA	6	4
23s	2	1
5s	3	2
16s	1	1

a 50 bases que alinean a lo largo de la isla pero que no cubren completamente algún gen.

Por otra parte, los polimorfismos presentes en las regiones señal, intermedia y media de la proteína VacA de *H. pylori* 29CaP y 7C resultaron ser s2, i2, m2, de acuerdo a los alineamientos múltiples y agrupación en los árboles filogenéticos realizados. En las figuras 6.5, 6.6 y 6.7 se muestran los alineamientos y filogramas de cada una de las regiones de VacA analizadas.

6.5. Comparación de los genomas de *H. pylori* 29CaP y 7C con otros genomas de *H. pylori*

Todos los “contigs” de ambas cepas de *H. pylori* secuenciadas presentaron algún grado de similitud con genomas de referencia o de plásmidos de *H. pylori* disponibles en las bases de datos públicas que se observó inicialmente por el alineamiento BLAST de cada uno de los “contigs” (Tablas 6.6 y 6.7). El alineamiento en Mauve con 46 genomas completos de *H. pylori* permitió identificar gráficamente regiones homólogas entre las secuencias analizadas (figura 6.9). En total 2,794 bloques colineares de la cepa 29CaP y 4,595 en la cepa 7C se encontraron. Los “contigs” correspondientes al plásmido de *H. pylori* 7C y al profago de *H. pylori* 29CaP presentaron uno o dos bloques colineares menores a 1,000 bases con los genomas incluidos en el análisis. Adicionalmente, la visualización de las líneas de conexión entre bloques de diferentes genomas mostró la variación en el arreglo de los genomas de *H. pylori* (Figura 6.8).

6.5.1. Análisis del genoma “core” de *H. pylori*

Al analizar 48 genomas de *H. pylori* se estimaron un total de 2,402 grupos de proteínas ortólogas. De estos, 931 grupos se comparten entre todas las genomas de las cepas incluidas, lo que correspondería al genoma “core”. El resto de las proteínas que se comparten entre 2 a 49 genomas, se les denominó genoma variable o “core extendido”. Adicionalmente, se obtuvo un número de secuencias únicas en cada cepa las cuales varían entre 1 a 187 y que corresponden a los denominados genes “cepa específicos”(ver figura 6.10).

No se encontró ningún grupo de ortólogos específico de los aislamientos obtenidos de cáncer gástrico.

6.6. Determinación del origen filogeográfico

La filogenia realizada con la secuencia concatenada de siete genes “housekeeping” de *H. pylori*, ubicó a las cepas 29CaP y 7C dentro del grupo hpEurope. La cepa 29CaP se encuentra más cerca de un clado formado con cepas de origen africano que se sitúa dentro de las cepas europeas (Figura 6.11).

La determinación del origen filogeográfico se realizó por MLST debido a que se cuenta con una base datos extensa, de más de 1,500 cepas en la que cada grupo filogeográfico de *H. pylori* que se ha descrito se encuentra representado por al menos 10 cepas, lo que da soporte a la filogenia. Se ha descrito que la filogenia realizada con grupos de genes ortólogos, obtenidos por la comparación de genomas completos, resulta en mejor topología de los árboles generados [72], pero las secuencias disponibles de genomas de *H. pylori* aún no representan a todas las poblaciones de *H. pylori* descritas. Aunque no se realizó una filogenia con los grupos de genes ortólogos, el alineamiento en Mauve generó un árbol basado en la similitud de los bloques homólogos entre todas las secuencias alineadas y que correlaciona con las poblaciones obtenidas por MLST como se observa en la Figura 6.12.

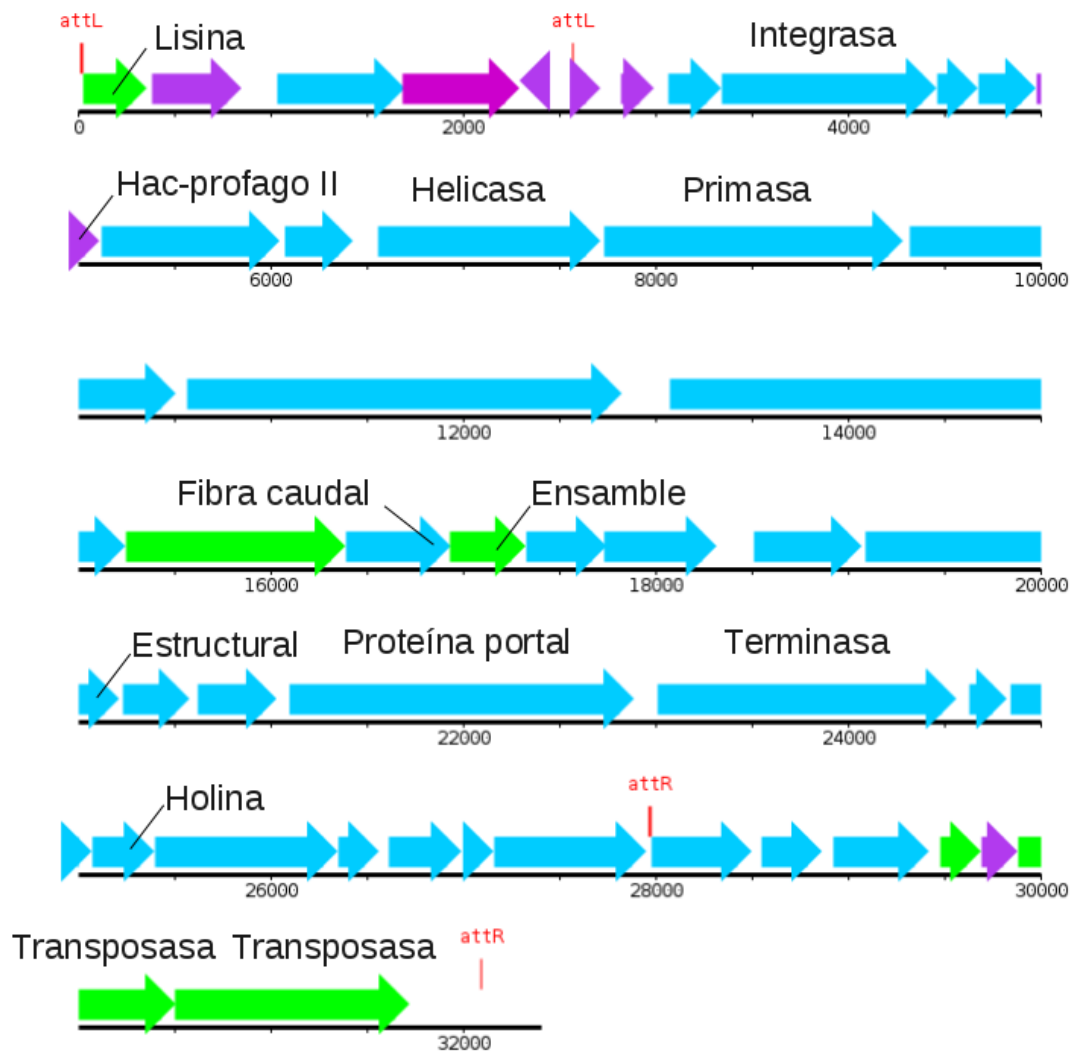


Figura 6.1: Anotación del profago de *H. pylori* 29CaP. Con líneas se señalan las funciones de algunas de las proteínas predichas. **Flechas azules**, genes con similitud a secuencias de genes de bacteriófagos de *H. pylori*. **Flechas verdes**, genes con similitud en bacteriófagos de otros géneros bacterianos. **Flechas moradas**, genes con similitud a genes presentes en genomas de *H. pylori* o *H. acynonichis*. *attL* y *attR*, sitios de unión similares a los sitios de unión de DNA del fago lambda. Anotaciones: Lisina, hidrolasa de mureína asociada a la fibra caudal (gi472437346); Integrasa, integrasa putativa (gi431810542); Hac-profago II, proteína homóloga a proteína del profago II de *H. acynonichis* (gi254780056); Helicasa, DNA helicasa (gi431810548); Primasa, DNA primasa (gi431810549); Fibra caudal, proteína de la fibra caudal (gi371671355); Ensamble, proteína que participa en el ensamble de la fibra caudal (gi388684690); Estructural, proteína de la cápside (gi431810554); Proteína portal, proteína que participa en el empaquetamiento del DNA (gi431810557); Terminasa, proteína implicada en el empaquetamiento del DNA (gi431810558); Holina, proteína que participa en la lisis de la bacteria (gi431810561); Transposasa, familia del elemento de inserción IS607 (gi422935176).

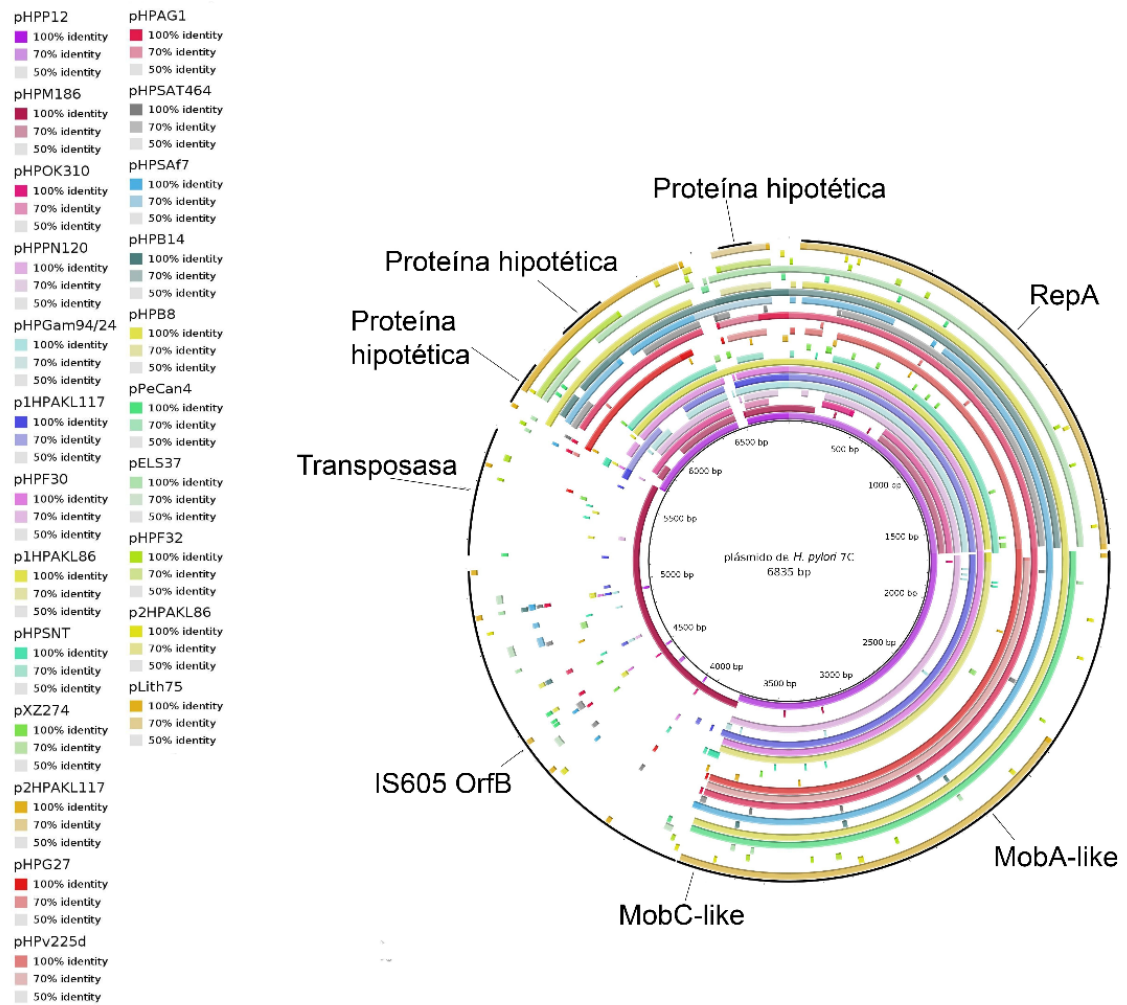


Figura 6.2: Alineamiento BLAST de los plásmidos de *H. pylori* contra el plásmido de *H. pylori* 7C. Imagen generada en BRIG

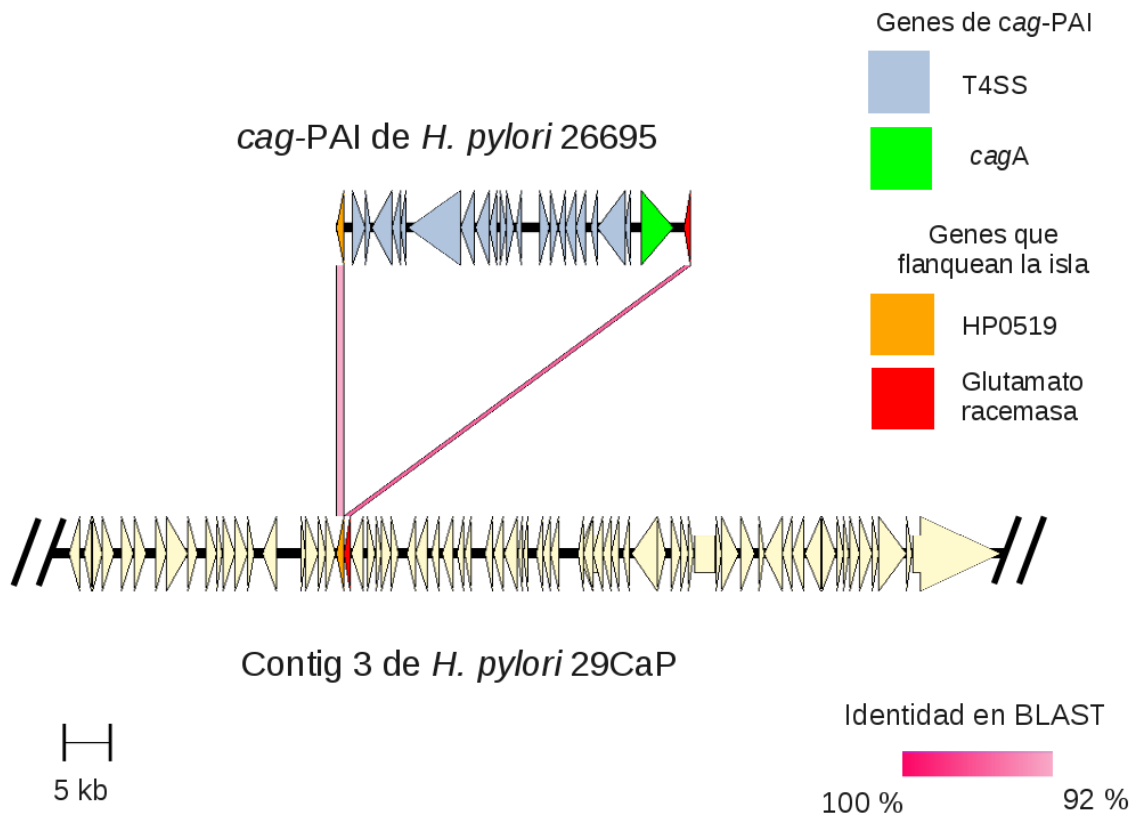


Figura 6.3: Alineamiento BLAST de la isla de patogenicidad *cag* de *H. pylori* 26695 contra contigs de *H. pylori* 29CaP. Las bandas rosas conectan los genes que tuvieron similitud, el porcentaje de similitud es indicado por el color de la barra que se muestra en la parte inferior derecha.

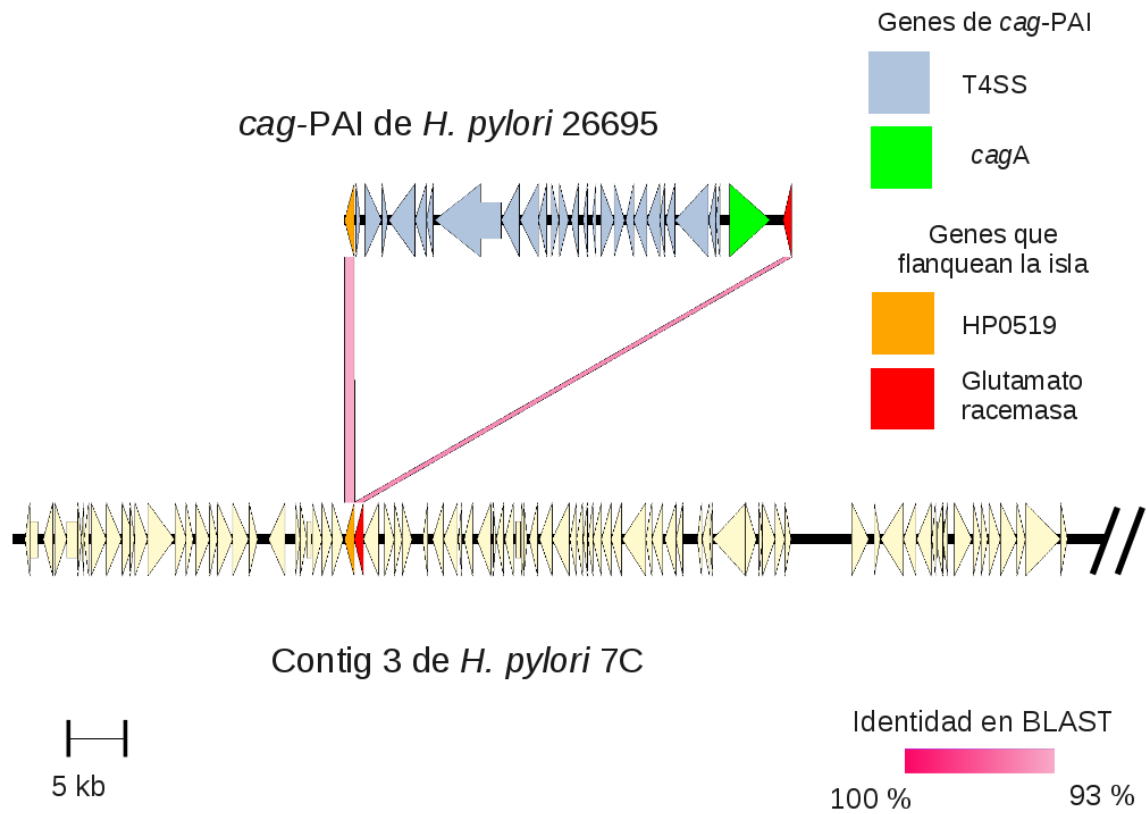


Figura 6.4: Alineamiento BLAST de la isla de patogenicidad *cag* de *H. pylori* 26695 contra contigs de *H. pylori* 7C. Las bandas rosas conectan los genes que tuvieron similitud, el porcentaje de similitud es indicado por el color de la barra que se muestra en la parte inferior derecha.

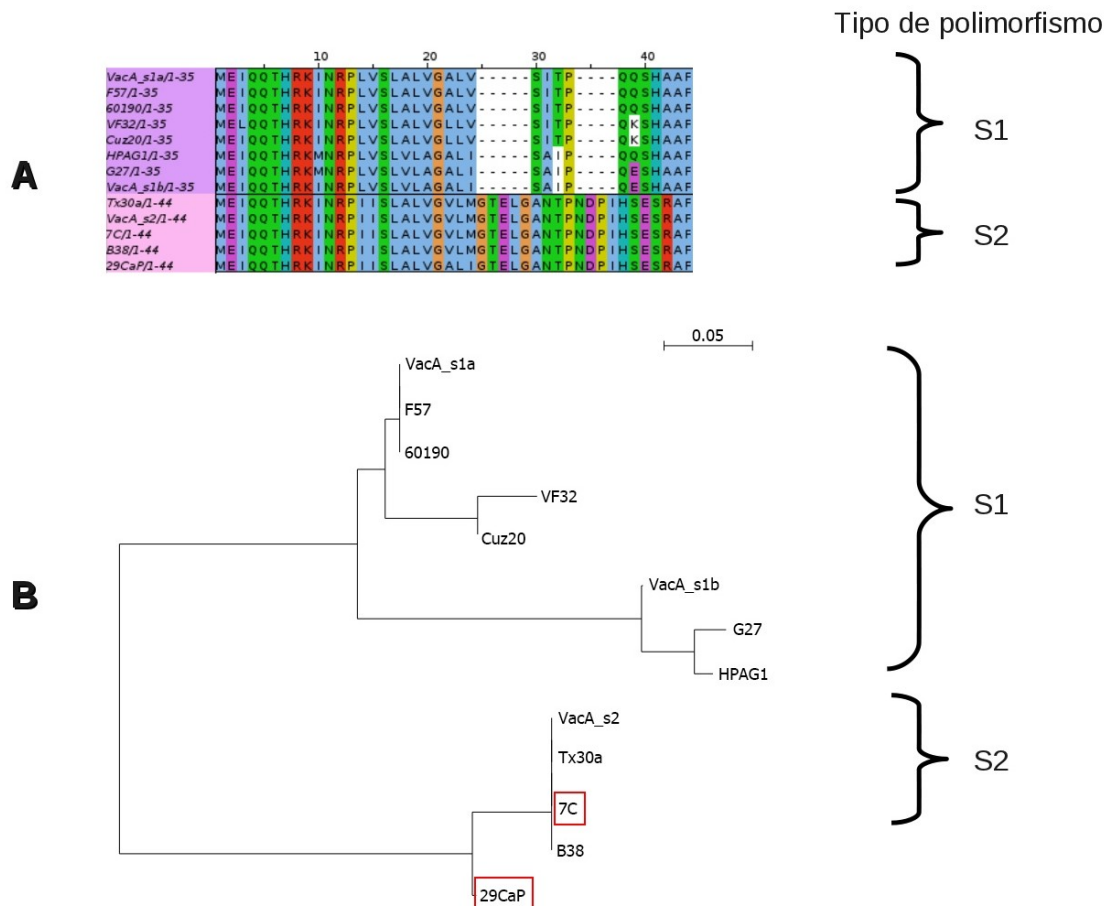


Figura 6.5: Tipificación de los polimorfismos de la región señal de VacA de *H. pylori* 29CaP y 7C. **A)** Alineamiento múltiple en ClustalW2 de la región señal de VacA de las cepas 29CaP y 7C con secuencias de referencia; **B)** Árbol filogenético realizado por el método Neighbor-Joining con la distancia del parámetro 2 de Kimura. En rectángulos rojos se muestran las cepas 29CaP y 7C.

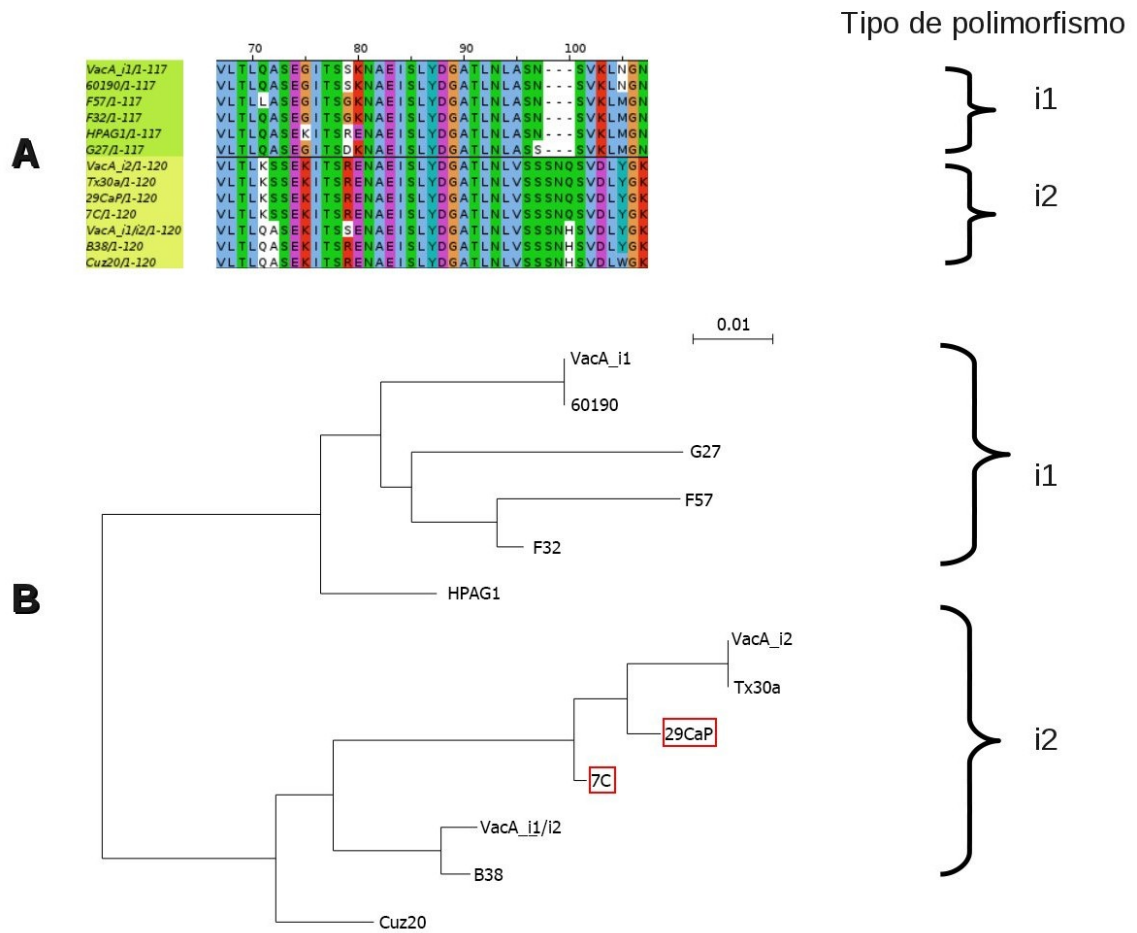


Figura 6.6: Tipificación de los polimorfismos de la región intermedia de VacA de *H. pylori* 29CaP y 7C. **A)** Alineamiento múltiple en ClustalW2 de la región intermedia de VacA de las cepas 29CaP y 7C con secuencias de referencia; **B)** Árbol filogenético realizado por el método Neighbor-Joining con la distancia del parámetro 2 de Kimura. En rectángulos rojos se muestran las cepas 29CaP y 7C.

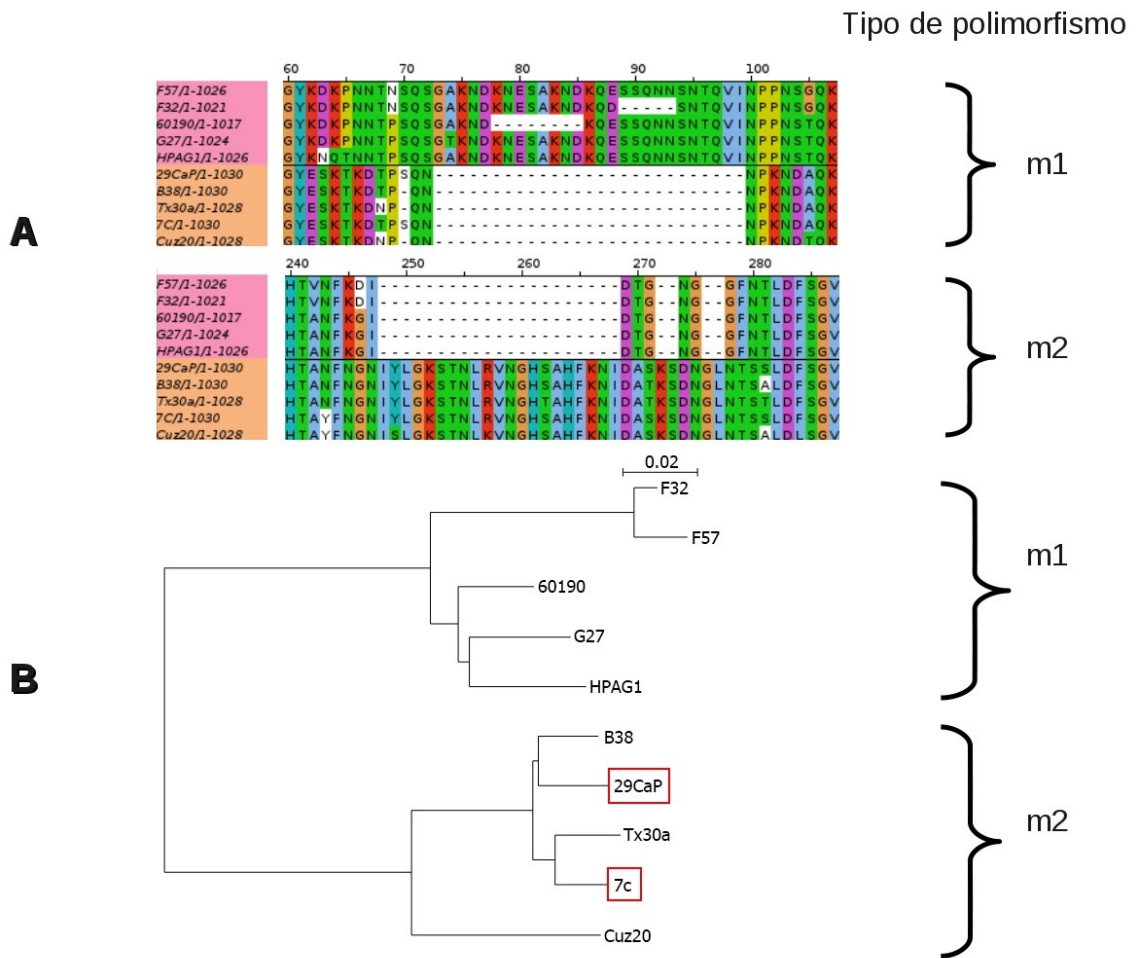


Figura 6.7: Tipificación de los polimorfismos de la región media de VacA de *H. pylori* 29CaP y 7C. **A)** Alineamiento múltiple en ClustalW2 de la región media de VacA de las cepas 29CaP y 7C con secuencias de referencia; **B)** Árbol filogenético realizado por el método Neighbor-Joining con la distancia del parámetro 2 de Kimura. En rectángulos rojos se muestran las cepas 29CaP y 7C.



Figura 6.8: Alineamiento en MAUVE de los contigs de *H. pylori* 29CaP y 7C con genomas de *H. pylori* disponibles en la base de datos de la NCBI.

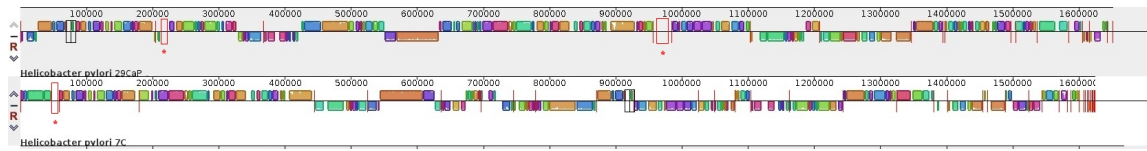


Figura 6.9: Regiones homólogas de los contigs de *H. pylori* 29CaP y 7C con genomas de *H. pylori* disponibles en la base de datos de la NCBI. Ampliación del alineamiento mostrado en la figura 6.8. Cada bloque de color corresponde a una región homóloga con al menos uno de los 46 genomas de *H. pylori* incluidos en el análisis. Las regiones que no tienen color como las mostradas con cajas rojas y un *, corresponden a regiones sin homología con otros genomas de *H. pylori*.

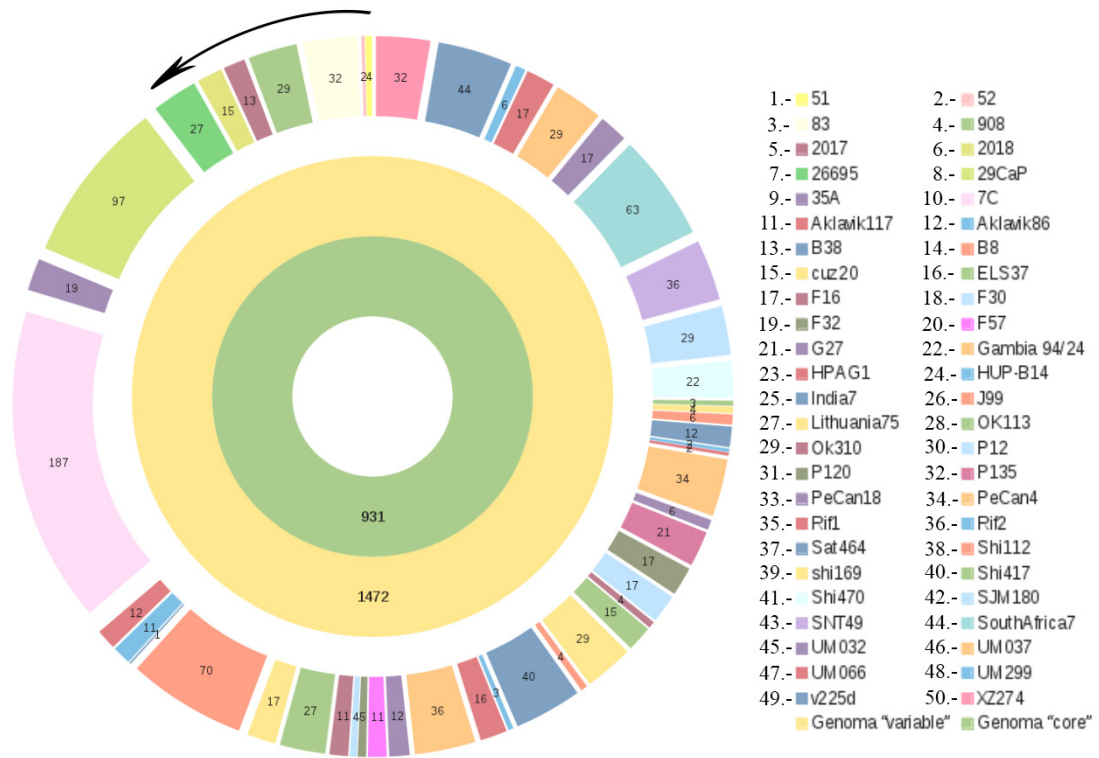


Figura 6.10: Representación del genoma "core" de *H. pylori*. El círculo interno representa el genoma "core" definido como el número de genes comunes en 50 genomas de *H. pylori*. El círculo medio representa el genoma variable, definido como los genes compartidos entre 2 a 49 cepas. El círculo externo representa los genes "cepa específicos" donde cada bloque color corresponde a una cepa de *H. pylori* de acuerdo a la leyenda de la derecha. El orden de las cepas está indicado por el número y la dirección de la flecha

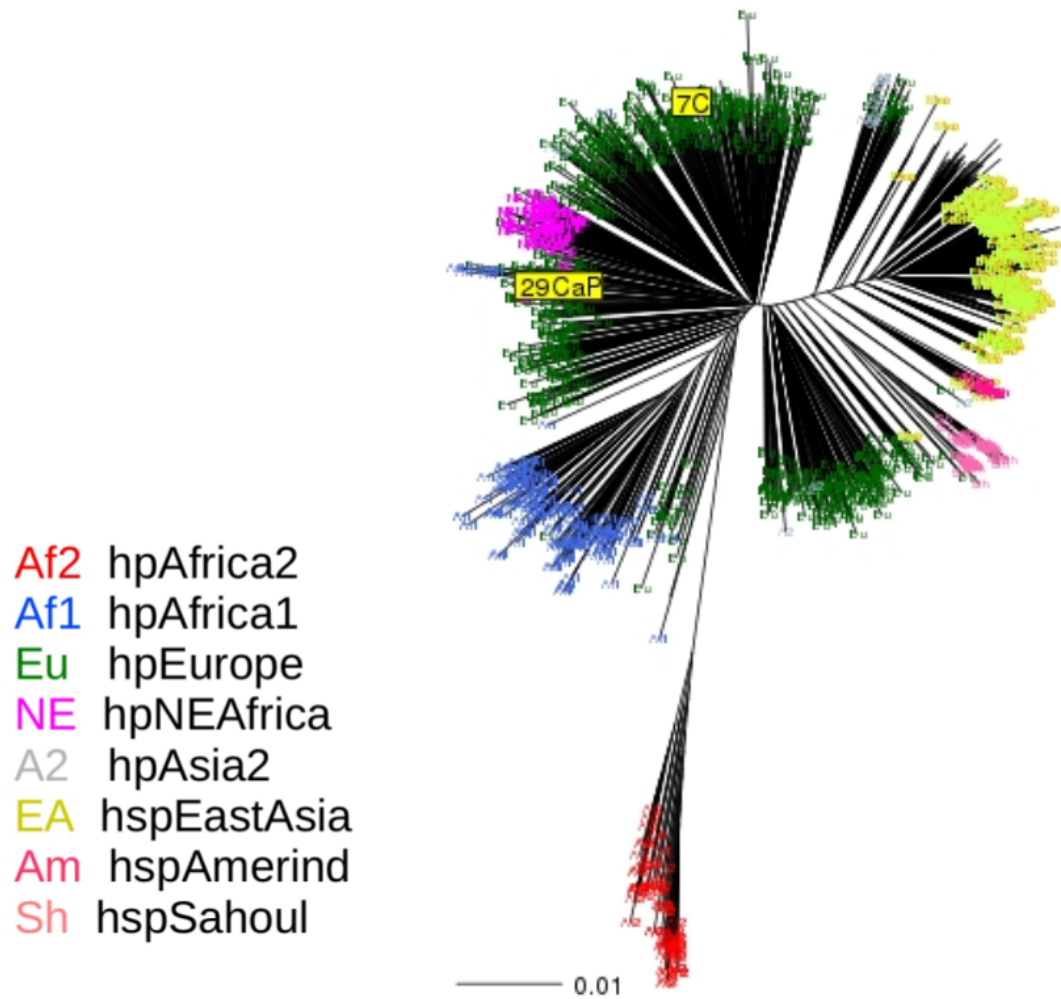


Figura 6.11: MLST de *H. pylori*

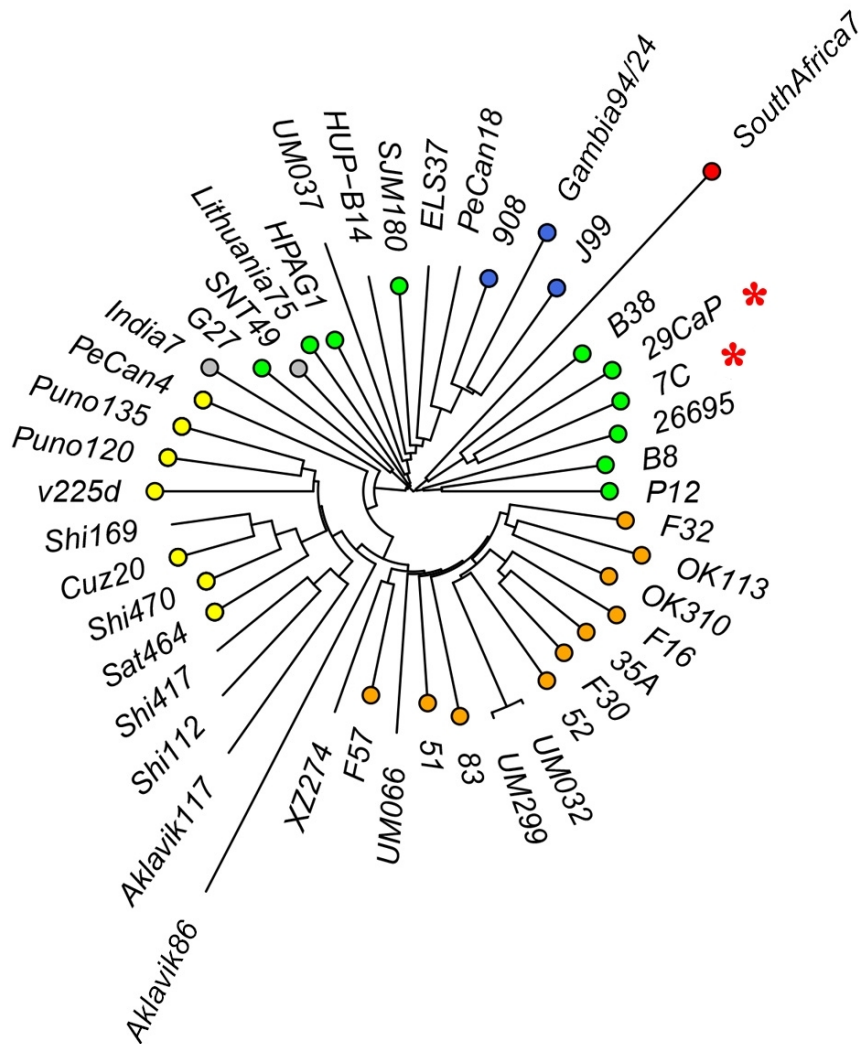


Figura 6.12: Filogenia de *H. pylori* generada con las regiones homólogas del alineamiento de genomas completos en Mauve. En círculos de colores se muestra el origen filogeográfico determinado por MLST reportado en la literatura. Colores: Rojo, hpAfrica2; Verde, hpEuropa; Azul, hspWAfrica; Gris, hpAsia2; Amarillo, hspAmerind; Naranja, hspEAsia. Las ramas sin círculos corresponden a cepas cuyo MLST no ha sido reportado. Con asteriscos (*) rojos se señalan las cepas *H. pylori* 29CaP y 7C

Capítulo 7

Discusión

Se secuenciaron los genomas de dos cepas de *H. pylori* aisladas en México. La cepa 7C se aisló de un paciente con gastritis crónica, es una cepa resistente a los fármacos de elección al tratamiento de eradicación y en un estudio anterior se determinó que esta cepa incrementa la apoptosis de las células epiteliales gástricas humanas de la línea celular AGS, fenómeno asociado a la sobreexpresión de proteínas citosólicas y de membrana de la bacteria [12]. La cepa 29CaP, se aisló de la biopsia tomada de la periferia de la lesión tumoral de un paciente con adenocarcinoma gástrico avanzado, esto resulta de interés en el grupo de trabajo ya que existen pocos reportes de aislamiento de la bacteria a partir de lesiones tumorales debido a que se considera que *H. pylori* no se desarrolla en la mucosa atrofiada y que la metaplasia intestinal es un medio desfavorable [86]. Adicionalmente, ambas cepas se están estudiando en un modelo murino de infección.

Para la secuenciación se utilizó la tecnología de pirosecuenciación que brinda lecturas de más de 300 pares de bases, lo que facilitó el ensamble *de novo* de los genomas de las cepas de *H. pylori* 29CaP y 7C, con una profundidad mayor a 100x. Esto dió la posibilidad de corregir errores de secuenciación al extender los contigs. El número final de contigs del genoma de *H. pylori* 29CaP resultó menor y de mayor longitud que el de *H. pylori* 7C, probablemente debido a que se incluyeron en el ensamble del genoma de esta cepa un conjunto adicional de lecturas de longitud mayor que permitió unir un mayor número de “contigs”. Sin embargo, el genoma de ninguna de las dos cepas se pudo cerrar. Una de las limitantes para el cierre

del genoma fue el empleo en el ensamble de lecturas no pareadas, pero además la presencia de secuencias repetidas y homopolímeros en la secuencia nucleotídica en los extremos de los contigs. En particular se ha estudiado que la presencia de múltiples homopolímeros es frecuente y casi característico del género *Helicobacter*, generando del 60 al 90 % de cambios de marco de lectura en el genoma [44].

Se estimó que las secuencias obtenidas corresponden a un genoma tipo “draft”, ya que cubren cerca del 92 % del genoma y la identidad de las secuencias comparada con los genomas de *H. pylori* fue igual o superior al 96 % en todos los casos (Tablas 6.4 y 6.5). Cada “contig” tuvo similitud con diferentes regiones de los genomas de *H. pylori*, por este motivo, para la anotación de los genomas con el algoritmo RATT [58] se utilizó un genoma de referencia para cada “contig”. Las anotaciones transferidas a los genomas de las cepas 29CaP y 7C, concuerdan con el número de genes promedio en los genomas de *H. pylori* publicados [38], pero no se descarta la posibilidad de que este número se modifique debido a que las secuencias aún están fragmentadas, lo que significa que existen genes repetidos o divididos en los extremos de los contigs y que pueden ser la razón de la variación en el número de elementos transferidos mostrados en la Tabla 6.8. Adicionalmente la presencia de múltiples bloques sinténicos que varían en orden y orientación entre los genomas de *H. pylori*, concuerda con las observaciones de la plasticidad del genoma de *H. pylori* debido a su competencia natural y a su alta tasa de recombinación.

Una de las características del genoma de la cepa 29CaP es la presencia de un profago. En los últimos años el interés en el estudio de los bacteriófagos de *H. pylori* ha surgido ya que podrían contribuir en la evolución de la virulencia de la bacteria [46, 50], además de que representan una nueva terapia alternativa en el tratamiento de infecciones causadas por cepas de *H. pylori* resistentes a antibióticos [50, 81]. Hasta la fecha se ha secuenciado el genoma de cuatro bacteriófagos de *H. pylori* y las predicciones *in-silico* revelan la existencia de secuencias fágicas crípticas y de profagos intactos en algunos genomas de *H. pylori* (Ver Tablas A.1 y A.2). El profago de *H. pylori* 29CaP tuvo similitud con los bacteriófagos hasta ahora secuenciados y con algunos profagos previamente reportados. De acuerdo a la predicción realizada el profago presente en la cepa 29CaP se encuentra intacto, lo que lo hace potencialmente inducible.

Por otra parte, la presencia de un plásmido en el genoma de *H. pylori* 7C resultó ser una característica de esta cepa. La presencia de los genes *mobA-D* sugiere que es un plásmido movilizable al igual que otros plásmidos de *H. pylori* reportados en la literatura [36, 66]. Se reporta que los plásmidos de *H. pylori* están involucrados en la transferencia conjugativa de DNA entre diferentes cepas, participando en la generación de variación genética de *H. pylori* [66]. Interesantemente, el plásmido de *H. pylori* 7C contiene el elemento de inserción IS605, el cual se presenta comunmente en una o varias copias en los genomas de *H. pylori* y se asocia a la generación de rearrreglos genómicos incluyendo la delección parcial o total de la isla de patogenicidad *cag* [13, 26].

Una característica que resultó de interés en ambas cepas es la ausencia de la isla de patogenicidad *cag*. La delección total de *cag*-PAI originada por el elemento de inserción IS605 se ha reportado[13]. Así mismo, se publicó en el año 2010 el genoma de la cepa *H. pylori* B38 aislada de un paciente con linfoma tipo MALT [76], la cual carece de *cag*-PAI al igual que las cepas utilizadas en este estudio. Sin embargo, no se sabe con certeza la frecuencia de los aislamientos de *H. pylori* que carecen de esta isla, pero existe evidencia de que del 40 al 70 % de las cepas a nivel mundial presentan delecciones parciales o rearrreglos en la secuencia de esta región [37], por lo tanto es posible que cepas de *H. pylori* *cag*-PAI positivas que inicialmente infectaron a un individuo presenten rearrreglos en la isla durante el curso de la infección, hasta el grado de perderla completamente.

Recientemente, Perkins y cols. [61] reportaron que *cag*-PAI es difícil de ensamblar *de novo* con las tecnologías de Illumina y Ion Torrent, no conocemos reportes al respecto con el uso del método de secuenciación que empleamos en este trabajo. Sin embargo, se reporta que para la tecnología de pirosecuenciación (Roche 454) existe un sesgo de cobertura en regiones con alto contenido de GC o AT, donde estas regiones se encuentran sub-representadas en los datos de la corrida [9, 35, 62]. Esta situación es importante al considerar que la isla de patogenicidad tiene menor contenido de GC (35 %) con respecto al resto del genoma (38-45 %)[13]. El alineamiento de las lecturas de cada genoma contra la secuencia de la isla de patogenicidad de cepas de referencia nos permitió descartar la posibilidad de un mal ensamble de esta región, aunque no se descarta la posibilidad de que no se haya captado la secuencia debido a

un sesgo de contenido de GC, por lo que aún se requiere comprobar mediante PCR la presencia o ausencia de esta isla en los genomas.

Con respecto a VacA, ambas cepas presentaron el genotipo s2, i2, m2, el cual es poco toxigénico en cultivos celulares y el cual se asocia a un menor riesgo de producir patología gástrica. Este hallazgo concuerda con el genotipo de la cepa *H. pylori* B38, que como se mencionó anteriormente es *cag*-PAI negativa y que además presenta el genotipo s2,m2. De la misma manera, se correlaciona con los resultados obtenidos en un estudio previo realizado en el grupo de trabajo en donde se compararon los genotipos de VacA de aislamientos de *H. pylori* de biopsias de pacientes con cáncer gástrico y de individuos sin alteraciones patológicas en la mucosa, en el que se identificó que en el grupo de cáncer gástrico existe un incremento en la diversidad de genotipos de las regiones señal y media de la citotoxina, con un aumento en la frecuencia de genotipos poco toxigénicos [49]. El genotipo s2 se presentó únicamente en cepas aisladas de biopsias de pacientes con cáncer gástrico con una frecuencia del 12% y en combinación con el genotipo s1b (s1b/s2) con una frecuencia del 20%, de igual manera el genotipo m2 se observó con 4.2 veces mayor frecuencia (21%/5%) en el grupo de cáncer con respecto al grupo control [49]. Estos resultados en conjunto con la ausencia de la isla en los genomas de *H. pylori*, resaltan la importancia de estudiar un mayor número de aislamientos a partir de un mismo individuo para poder establecer la asociación de otro(s) factor(es) de virulencia con patogenicidad, asimismo es necesario incluir cepas de diferentes orígenes geográficos, ya que existe evidencia de que no hay correlación entre los polimorfismos de VacA (incluyendo la región intermedia) con la patología gástrica cuando se incluyen en el estudio cepas provenientes de diferentes regiones geográficas, aún en países en los que el cáncer gástrico es un problema de salud importante [57]. Esto sugiere que otros factores bacterianos participan en el desarrollo de patología gástrica y que los polimorfismos de CagA y VacA reflejan mayoritariamente la adaptación de la bacteria a su hospedero que la causalidad de una enfermedad gástrica, sin excluirlos como agentes potenciadores de riesgo.

De hecho, las observaciones recientes sobre diferencias en la virulencia de cepas de *H. pylori* de diferentes orígenes geográficos sugieren que la agrupación filogeográfica de las cepas puede ser un predictor de riesgo para el desarrollo de cáncer

gástrico [18, 70]. Hasta la fecha no se tienen reportes del origen filogeográfico de *H. pylori* aislado de población mexicana genéticamente mezclada. Sin embargo, existe evidencia de que las cepas de *H. pylori* de población indígena mexicana se incluyen principalmente en el grupo hpEurope, junto con otras cepas de países Latinoamericanos, y que algunas cepas presentan ancestría amerindia, asiática y africana [11]. La tipificación por MLST como la filogenia de genomas completos permitió establecer que la ancestría de ambas cepas secuenciadas es europea (hpEurope), aunque la filogenia inferida por MLST sugiere que la cepa 29CaP es filogenéticamente cercana al grupo de cepas de origen africano (hpAfrica1 y hpNEAfrica). Esto correlaciona con lo anteriormente descrito para las cepas de *H. pylori* de población mexicana indígena [11] y con lo reportado para la población mexicana genéticamente mezclada mediante la tipificación de los haplotipos del DNA mitocondrial, que muestra que aún conserva marcadores genéticos amerindios con algunos componentes europeos y africanos [32], lo que refleja la historia de la colonización de México prehispánico en la que la población amerindia autóctona fue desplazada por inmigrantes europeos, principalmente de España, y en menor grado por esclavos africanos durante el siglo XVII [24]. Aún resulta interesante saber si en México el origen filogeográfico hpEurope se relaciona con el desarrollo de cáncer gástrico, como se reportó en Colombia [18] por lo que se requieren estudios en donde se analice un mayor número de cepas de diferentes padecimientos gástricos.

Capítulo 8

Conclusiones

- Se obtuvo la secuencia “draft” de los genomas de las cepas de *H. pylori* 29CaP y 7C, que son las primeras cepas de *H. pylori* aisladas en México cuyo genoma se secuenció.
- El origen filogeográfico de las cepas estudiadas fue hpEurope, lo que correlaciona con las características genéticas de la población Mexicana.
- No se evidenció la presencia de alguna característica genómica específica de aislamientos de cáncer gástrico.
- Se logró mejorar el ensamblaje de los genomas mediante la utilización de otros algoritmos para el análisis de las secuencias.
- Actualmente se está trabajando en el cierre de los genomas analizados en este trabajo.

Apéndice A

Bacteriófagos y profagos de *H. pylori*

Hasta la fecha se han secuenciado cuatro bacteriófagos de *H. pylori*, los números de acceso de éstos se enlistan en la Tabla A.1. Adicionalmente se realizó la predicción de profagos en los 46 genomas de *H. pylori* (Tabla 5.4) con el programa PHAST [87]. Únicamente 15 genomas presentaron regiones de profagos, de los cuales 4 se predijeron como intactos, 8 como incompletos y 3 como cuestionables (ver Tabla A.2).

En la Figura A.1 se presenta el alineamiento múltiple de los bacteriófagos y profagos de *H. pylori* incluyendo el profago de la cepa 29CaP. El alineamiento se realizó con el algoritmo PROmer del paquete Mummer3 [42], el cual encuentra regiones conservadas de aminoácidos en la secuencia de nucleótidos traducida en seis marcos de lectura. En el alineamiento se consideró como referencia el profago de la cepa 29CaP.

Tabla A.1: Genomas de bacteriófagos de *H. pylori* secuenciados

Cepa	No. de acceso	Referencia
phiHP33	NC_016568	[46]
KHP30	NC_019928	[79]
KHP40	NC_019931	[79]
1961P	NC_019512	[50]

Tabla A.2: Profagos de *H. pylori* predichos con PHAST en genomas secuenciados

Cepa	No. de Acceso	posición	Integridad
51	NC_017382	667653-684059	incompleto
83	NC_017375	78964-103565	incompleto
B38	NC_012973	533436-556398	incompleto
		1503699-1529021	incompleto
Cuz20	NC_017358	185385-217197	intacto
ELS37	NC_017063	182778-195077	cuestionable
F16	NC_017368	980907-1011083	cuestionable
F57	NC_017367	734227-754727	incompleto
Gambia94/24	NC_017371	201182-219928	intacto
HUP-B14	NC_017733	660566-673624	incompleto
India7	NC_017372	1216115-1249186	intacto
OK310	NC_020509	680257-695333	incompleto
Sat464	NC_017359	866117-876390	incompleto
UM032	NC_021215	1501469-1515140	cuestionable
UM037	NC_021217	1185213-1215346	intacto
UM299	NC_021216	488504-502176	cuestionable

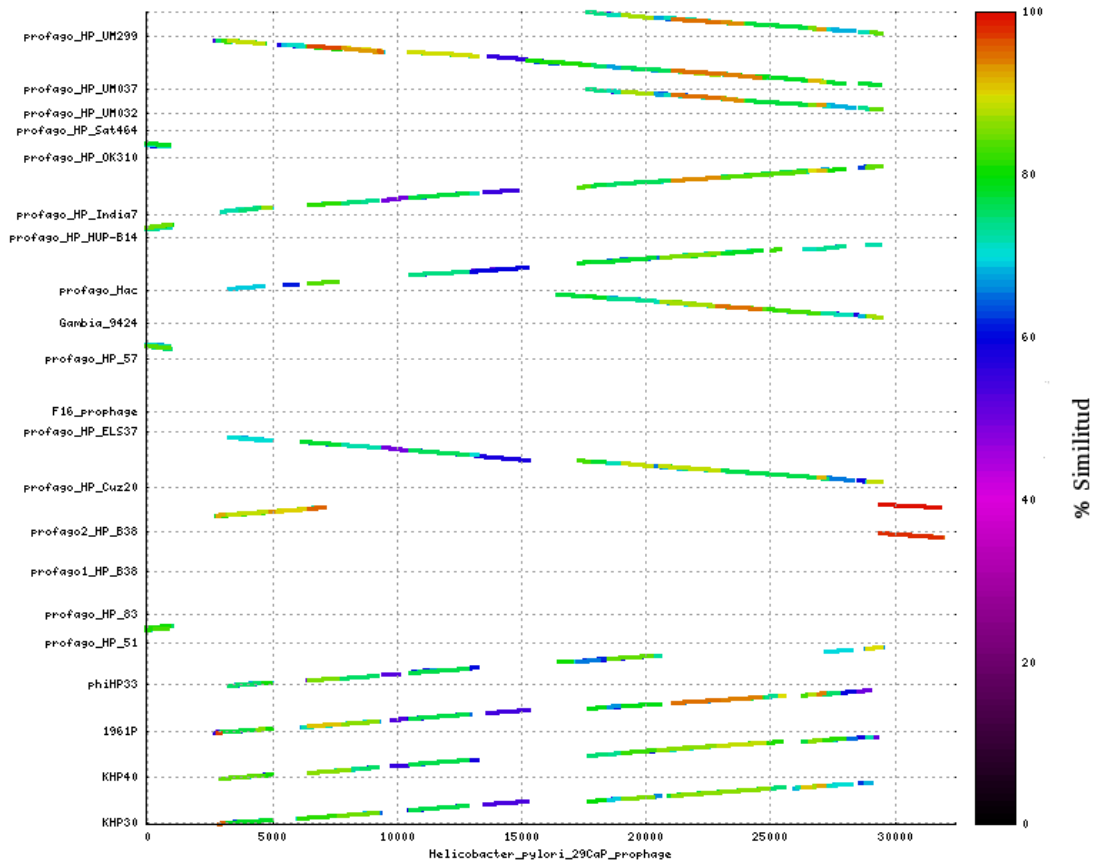


Figura A.1: Alineamiento de los bacteriófagos y profagos de *H. pylori* contra la cepa 29CaP. En el gráfico se representa con barras de colores la similitud del profago de *H. pylori* 29CaP con los bacteriófagos y profagos de *H. pylori* presentados en las tablas A.1 y A.2. Los colores corresponden al porcentaje de similitud mostrado en la barra de la derecha.

Apéndice B

Publicaciones

- Morales-Guerrero S. E., **Mucito-Varela E.**, Aguilar-Gutiérrez R.G., López-Vidal Y., Castillo-Rojas G. (2013). The Role of CagA Protein Signaling in Gastric Carcinogenesis – CagA Signaling in Gastric Carcinogenesis, *Current Topics in Gastritis - 2012*, Prof. Gyula Mozsik (Ed.), ISBN: 978-953-51-0907-5, InTech, DOI: 10.5772/53136. Available from: <http://www.intechopen.com/books/current-topics-in-gastritis-2012>
- **Mucito-Varela E. J.**, Flores-López V., Castillo-Rojas G., Cevallos-Gaos M.A., López-Vidal Y. Abstract no.: P02. Genome sequencing of *Helicobacter pylori* strains isolated in Mexico from gastric cancer and chronic gastritis Lesions. (2013), Abstracts of the European *Helicobacter* Study Group XXVIth International Workshop on *Helicobacter* and Related Bacteria in Chronic Digestive Inflammation and Gastric Cancer. *Helicobacter*, 18: 77–154. doi:10.1111/hel.12079 Impact Factor: 3.511

The Role of CagA Protein Signaling in Gastric Carcinogenesis — CagA Signaling in Gastric Carcinogenesis

Stephanie E. Morales-Guerrero,
Eduardo Mucito-Varela,
Germán Rubén Aguilar-Gutiérrez,
Yolanda Lopez-Vidal and Gonzalo Castillo-Rojas

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/53136>

1. Introduction

Even though chronic gastritis was well established as a risk factor for the development of several gastric pathologies, such as peptic ulceration, gastric carcinoma, and lymphoma, before the discovery of *Helicobacter pylori* by Warren and Marshall in 1983 (Warren, 1983), the development of chronic gastritis was thought to be due to a plethora of possible causes such as stress, excessive alcohol consumption, ingestion of hot beverages and spicy food, among others (Mobley et al., 2001). With the association of *Helicobacter pylori* as the cause of chronic gastritis, the understanding of bacterial pathogenicity and the host response against the bacterium take enormous importance to understand the pathophysiology of the infection.

It is estimated that almost half of the world's population is infected with *Helicobacter pylori* and although the majority of colonized individuals remain asymptomatic, the infection represents the primary cause of chronic gastritis and it is a risk factor for the development of gastric cancer; it was the first bacterium classified as a type 1 carcinogen by the IARC in 1994 (IARC, 1994). Gastric cancer, despite its declining incidence rate, it remains as the fourth most common cancer and the second cause of cancer-related death worldwide (Ferlay J, 2010) and epidemiological studies have determined that attributable risk for gastric cancer conferred by *H. pylori* infection is approximately 75% (Peek & Blaser, 2002).

The development of gastritis associated with the infection is a multifactorial process, where the bacterium and host factors influence the pathogenesis of the disease. Colonization by *H. pylori* causes chronic gastritis, that can remain asymptomatic, produces gastric epithelial damage (Ricci et al., 2011). In the initial phase of infection, the bacteria penetrate through the mucous layer to multiply in the proximity of epithelial cells. The gastric epithelium responds to the infection by mucin depletion and cellular regenerative changes (Mobley et al., 2001); it also occurs the infiltration of polymorphonuclear cells into surface epithelium and hypochlorhydria and a deficient ascorbic acid secretion are also observed (Sobala et al., 1993). Later, bacterial components penetrate through the damaged epithelium and stimulates the production of IL-8 and IL-1 by macrophages, the activation and degranulation of mast cells and the release of inflammatory mediators that increase vascular permeability and up-regulate the expression of leukocyte adhesion molecules (Mobley et al., 2001). As the immune response fails to eliminate the bacteria, it occurs a gradual accumulation of inflammatory cells, which represents an active chronic phase of gastritis.

2. *Helicobacter pylori* virulence factors in gastritis and gastric cancer

Researchers have studied *H. pylori*'s virulence factors, to elucidate their role in the development of gastritis and gastric cancer, the most studied are the pathogenicity island (*cag*-PAI) and its effector, the cytotoxin associated protein (CagA) and the vacuolating cytotoxin (VacA). However, there are other virulence factors (Figure 1), including the urease, lipopolysaccharide, peptidoglycan, several adhesins, such as outer inflammatory protein (OipA), sialic acid-binding adhesin (SabA), blood group antigen-binding adhesin (BabA/B), that contribute to the damage in the gastric epithelium (Backert et al., 2010).

3. Urease

To colonize the stomach, *H. pylori* has to survive in the presence of acid produced in the stomach. To counteract this acidic environment, *H. pylori* produces an important enzyme, urease, which hydrolyses urea into NH_3 and CO_2 . This enzyme has an essential role in the *H. pylori* infection as observed in urease-defective bacteria mutants which cannot colonize the stomach (Montecucco & Rappuoli, 2001). Urease causes damage to the epithelium through the production of ammonia, that in conjunction with neutrophil metabolites (Megraud et al., 1992), form carcinogenic agents that might participate in the development of gastric malignancies (Suzuki et al., 1992). Ammonia is capable of cause different cell alterations, including swelling of intracellular acidic compartments, alterations of vesicular membrane transport, repression of protein synthesis and ATP production, and cell-cycle arrest (Montecucco & Rappuoli, 2001). Urease might also help to the recruitment of neutrophils and monocytes in the mucosa and to the production of proinflammatory cytokines (Harris et al., 1996).

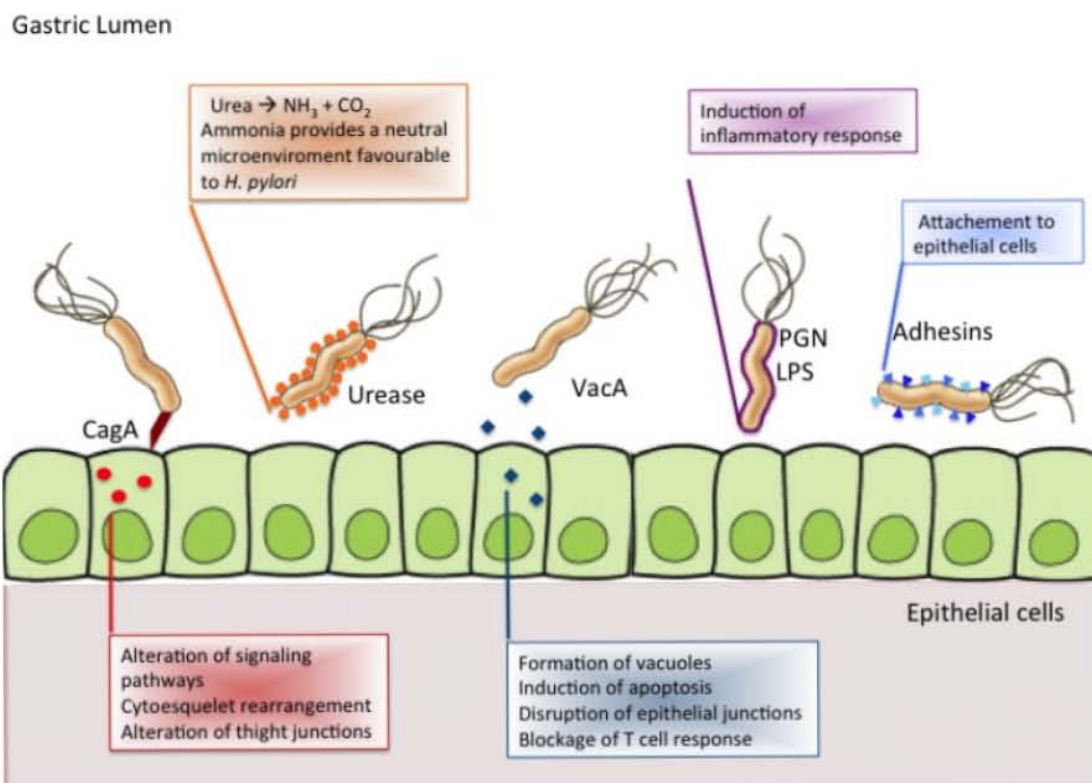


Figure 1. *Helicobacter pylori* virulence factors activities. Although CagA and VacA are the most studied virulence factors, there are more bacterial proteins that promote damage to the host cells.

4. Lipopolysaccharides (LPS) and peptidoglycan

H. pylori LPS are essential components of the bacterial outer membrane that induces a low immunological response in contrary to other bacteria (Moran, 1995). However, despite the low immunological activity of LPS, *H. pylori* colonization is associated with an inflammatory response, because *H. pylori* has the ability of activate mononuclear cells by LPS-independent mechanisms as well as other bacterial surface molecules (Mai et al., 1991, Moran, 1998). Another important component of the cell wall of *H. pylori*, peptidoglycan, interacts with the intracellular pattern recognition receptor Nod1 (an intracellular sensor for peptidoglycan from Gram-negative bacteria), this interaction leads to activation of NF- κ B signaling, with the subsequent secretion of the inflammatory molecules IL-8 (Viala et al., 2004) and β -defensin-2 (Boughan et al., 2006). On the other hand, peptidoglycan translocation of *H. pylori*, leads to enhanced PI3K-AKT signaling, which mediates protection from apoptosis and cell migration, both phenotypes related to carcinogenesis (Nagy et al., 2009).

5. Adhesins

Approximately 4% of the *H. pylori* genome is predicted to encode outer membrane proteins (OMPs) many of which serve as adhesins (Wroblewski et al., 2010), epithelial adhesion provides to *H. pylori* better access to epithelial surface and in this way contributes to deliver bacterial toxins such as CagA (Yamaoka, 2010). Fucosylated ABO blood group antigens and sialyl-Lewis^x and sialyl-Lewis^a antigens (sLe^x and sLe^a) have been identified as functional receptors for *H. pylori* (Boren et al., 1993).

The outer membrane inflammatory protein (OipA), identified in 2000 (Yamaoka et al., 2000), is involved in the attachment of *H. pylori* to gastric epithelial cells *in vitro* (Yamaoka, 2010). OipA is also involved in up-regulation of matrix metalloproteinase 1 (MMP-1), in inhibition of glycogen synthase kinase 3 β (GSK-3 β) and in β -catenin translocation to the nucleus, this protein also induces an inflammatory response and actin rearrangement through phosphorylation signaling pathways (Wroblewski et al., 2010), although, most of these are also involved in *cag*-PAI signaling alterations, suggesting that there might be some interaction between OipA and *cag*-PAI, it is important to note that it has been found a correlation coefficient of 0.82 between the *oipA* and *cagA* positivity in some virulent strains (Yamaoka et al., 2002).

Blood group antigen binding adhesin (BabA), has been thought to be the primary protein involved in adherence to the gastric mucosa (Boren et al., 1993). It is encoded by the *babA2* gene (the *babA1* although is present, is silent because of lack of an initiation codon (Mobley et al., 2001) and binds to fucosylated Lewis^b antigen. Some studies report that the presence of *babA2* is associated with duodenal ulcer and gastric cancer and found in conjunction with *cagA* and *vacA* s1 allele, provides a greater risk of develop a severe gastric disease (Gerhard et al., 1999).

Sialic acid-binding adhesin (SabA) is responsible of *H. pylori* binding to sialylated Le^x and Le^a in epithelial cells. Infiltration of neutrophils into the gastric mucosa is a characteristic feature of chronic gastritis caused by *H. pylori* and SabA binds to neutrophils through this sialylated carbohydrates and induces oxidative burst in these cells, which has a consequence of produce oxidative damage in gastric epithelium. It has been proposed that SabA-positive status is associated with gastric cancer, intestinal metaplasia, and corpus atrophy and negatively associated with duodenal ulcer and neutrophil infiltration (Yamaoka, 2008).

6. Vacuolating cytotoxin (VacA)

VacA is a pore-forming cytotoxin identified in supernatants of *H. pylori* broth cultures that cause aberrant vacuolation of cultured cells (Leunk et al., 1988). This cytotoxin is secreted from the bacteria as a large 140-kilodalton polypeptide and latter trimmed at both ends to finally deliver it in an active form to host cells, where it exerts its activity.

VacA induces multiple cellular activities, the best studied is the alteration on endosomal maturation which consequently leads to vacuolation of epithelial cells, VacA is also capable

of induce membrane-channel formation, cytochrome c release from mitochondria, and binding to cell-membrane receptors activating a proinflammatory response (Amieva & El-Omar, 2008). VacA has the ability to cause leakage of ions and small molecules, by disrupting the barrier function of tight junctions (Papini et al., 1998) and also inhibit T-cell activation and proliferation (Gebert et al., 2003).

All strains of *H. pylori* contain the *vacA* gene, although, there is considerable genetic diversity among them, and consequently the cytotoxicity activity of the toxin varies between strains (Amieva & El-Omar, 2008). The diversity observed is attributed to variations in *vacA* gene structure within the signal (s) region, the middle (m) region, and the intermediate (i) region (Wroblewski et al., 2010).

The amino terminus contains the signal sequence that shows allelic variability and has been classified into different types. Strains with s1 allele secrete an active toxin and are also highly associated with ulcers and gastric cancer (Atherton et al., 1995), however, s1/s2 combination or s2 genotypes are found in patients with gastric cancer (Lopez-Vidal et al., 2008). The middle region of the gene also shows allelic variation, with m1 subtype having stronger vacuolating activity and it is associated with an increased risk for development of gastric epithelial injury and gastric cancer (Yamaoka, 2010).

The clinical isolates of *H. pylori* have been grouped into two broad families defined as type I and type II on the basis of whether they have a functional pathogenicity island (*cag*-PAI), and secrete an active vacuolating cytotoxin (VacA). Type I strains are positive for all these characteristics. In contrast, type II strains do not have a functional *cag*-PAI and possess the *vacA* gene that encodes for non-toxic protein (Censini et al., 1996, Tomb et al., 1997).

7. Cytotoxin associated gen-Pathogenicity Island (*cag*-PAI)

Colonization of the human stomach by *H. pylori* virulent strains is associated with a significantly increased risk for the development of several gastric diseases, such as gastric cancer. These highly virulent *H. pylori* strains harbor a (cytotoxin-associated genes) pathogenicity island (*cag*-PAI) that encode proteins that are components of a type IV secretion system (T4SS) apparatus and the CagA effector protein into host target cells (Castillo-Rojas et al., 2004). T4SS translocates the bacterial oncoprotein CagA and peptidoglycan that induces proinflammatory chemokine and cytokine secretion, including interleukin-8 (IL-8) in gastric epithelial cells (Figueiredo et al., 2005). The hallmark of *H. pylori* virulent strains is the presence of *cag*-PAI that induce of gastric mucosal inflammation, which is a risk factor for developing severe gastric pathologies (Antonio-Rincon et al., 2011).

The presence of flanking direct repeated sequences 605 (IS605) in the *cag*-PAI have been reported to cause rearrangements and partial or total deletions of it. This has generated *H. pylori* strains with varying virulence, and consequently, the clinical outcome of the infection (Censini et al., 1996).

T4SSs are a large group of highly versatile secretion machineries in many Gram-negative pathogenic bacteria that are evolutionarily related to bacterial conjugative systems (Fronzes et al., 2009). Although the *H. pylori* *cag*-PAI has been evolutionarily related to others T4SSs based on their common features like structural and sequence similarities, only a few *cag* PAI proteins (also termed as HP α VirB/D) show clear sequence similarities to the prototypical system of *A. tumefaciens* and other bacterial species.

A systematic mutagenesis study with isogenic mutants in each of the *cag*-PAI genes, reported that 14 genes out of 27 of *cag*-PAI are essential for CagA translocation and induction of IL-8 secretion in gastric epithelial cells (Aguilar et al., 2001).

Twelve genes out of 27 of *cag*-PAI encoded proteins share sequence similarities with components of the prototypical T4SS system VirB/D4 of *Agrobacterium tumefaciens*, which contains 11 essential proteins of the secretion apparatus (VirB1–VirB11) encoded by the operon *virB* and a coupling protein (VirD4) that mediates substrate recognition (Fronzes et al., 2009). The majority of these proteins are conserved among the known T4SSs of other bacteria (Alvarez-Martinez & Christie, 2009).

The prototypical (macromolecular complex) T4SS of *A. tumefaciens*, is among the best-studied T4SS and is composed of two major structural complexes: the T-pilus and a membrane-associated complex responsible for the translocation of substrates across the inner and the outer membrane (Fronzes et al., 2009, Alvarez-Martinez & Christie, 2009). The membrane-associated complex includes the energetic components VirB4, VirB11 and VirD4, which are conserved among the known T4SSs of other bacteria. These proteins contain Walker nucleotide binding, and hydrolysis motifs and are known to energize T4SS for its function. The membrane-associated complex also includes the proteins VirB1, VirB3, VirB6, VirB7, VirB8, VirB9 and VirB10 that form the substrate translocation channel that spans both bacterial membranes. VirB6, VirB8, and VirB10 are part of the subcomplex of proteins attached to the inner membrane with domains spanning the periplasm, while VirB7 and VirB9 are part of the outer membrane subcomplex. The T4SS-determined T-pilus is an extracellular appendage that is thought to initiate cell–cell contact with plant target cells before the initiation of T-DNA transfer. This structure includes the proteins VirB2 and VirB5 as major and minor components, respectively (Backert et al., 2008).

H. pylori and *A. tumefaciens* share similarity in several components of their T4SSs and probably function in a similar manner (Backert & Meyer, 2006). Nevertheless, the molecular mechanisms responsible for assembly, activity and specific recognition of host target cells of these macromolecular complexes remains unclear. It has been suggested that after the induction by cell contact, *H. pylori* forms T4SS pili located at one bacterial pole for the injection of virulence factors into target cells such as the oncprotein CagA.

8. Cytotoxin-associated protein (CagA)

CagA protein is the only known effector protein translocated to the host cell by the T4SS (Backert et al., 2010). It was firstly described as a very immunogenic protein in humans infected

with cytotoxin-producing strains (Covacci et al., 1993). After that, its association with gastric cancer and peptic ulcer was well established.

Recent reports suggest that CagA is able to down regulate the vacuolation effects of VacA on host cells, and conversely, VacA is able to down regulates CagA activity (Oldani et al., 2009, Tegtmeyer et al., 2009). Clinical isolates that contain the *cag*-PAI generally are *babA*, *oipA* and *vacA* s1 positive, Yamaoka proposes that these virulence factors may interact with each other with a certain biological significance; therefore, these factors interact synergistically with each other to induce gastric diseases (Yamaoka, 2010).

CagA is encoded by the *cagA* gene, located at one end of the *cag*-PAI (Censini et al., 1996). One copy per genome is present in most of the strains whose genome has been sequenced but strains with two gene copies might exist, as revealed by the genome sequence of the Amerindian strains *H. pylori* Shi470 and V225d, although functionality of both copies in these strains is unknown (Kersulyte et al., Mane et al., 2010). It has been estimated that 60-95% of strains worldwide carry the *cagA* gene, which is different between geographic regions (Hatakeyama, 2004). However, the complete *cag*-PAI is important for the translocation of this protein and the completeness of the island may impact on the association of CagA presence and the outcome of the disease.

Other main feature of CagA protein is its high variable size, which range from 120-145 kDa (Covacci et al., 1993, Tummuru et al., 1993). This variation is due to the presence of polymorphism at the carboxy-terminal region given by the presence of repeat sequences called EPIYA motifs, which are present in a copy number varying from one to seven (Xia et al., 2009). Four major types of EPIYA motifs (A, B, C, and D) have been described based on the specific amino acid sequence that flanks the Glu-Pro-Ile-Tyr-Ala motif at both sides (Hatakeyama, 2004).

The EPIYA-repeat region of CagA is made of tandem alignment of the four distinct EPIYA segments (A, B, C, D) in various combinations (Hatakeyama, 2011). CagA molecules containing the EPIYA-A, EPIYA-B and EPIYA-C segments in tandem are denoted as “western type”, and are found in many *H. pylori* isolates from Europe, North America, and Australia. On the other hand, CagA carrying EPIYA-A, EPIYA-B and EPIYA-D segments are called “East Asian type” and have been isolated in East Asian countries such as Japan, Korea and China (Furuta et al., 2011).

The EPIYA-C segment in “Western type” isolates can be present in up to five copies which might be due to recombination mediated by a 16 amino acid sequence designated CagA multimerization sequence that is present at both sides of EPIYA-C motif. The CagA multimerization sequence is also present in EPIYA-D but only at one end, therefore cannot duplicate by this mechanism (Furuta et al., 2011). Other implications of CagA multimerization sequence in the bacterial pathogenicity will be discussed later in the text.

It has been proposed that EPIYA motifs of CagA act as “master keys” that can potentially be tyrosine-phosphorylated by host kinases and are capable to interact with a wide range of different host cell proteins (Backert et al., 2010). In the same way, the observation that the EPIYA-region is unstructured and can adopt multiple conformations mimicking host sub-

strates of kinase families, thus manipulating eukaryotic cellular biochemistry during infection, provides further support for this "master key" hypothesis (Nesic et al., 2010).

In contrast to the polymorphic C-terminal EPIYA region of CagA, N-terminal region is well conserved among CagA proteins (Hatakeyama, 2011). The N-terminal CagA region is required for the membrane association of CagA in polarized epithelial cells (Bagnoli et al., 2005) and is important for its translocation into host cells (Murata-Kamiya et al., 2010), the first 200 amino acids at the N-terminus of CagA form a membrane-binding domain in *H. pylori*, mediating proper translocation of the CagA protein via the T4SS (Steininger et al., 2011).

9. CagA tridimensional structure and membrane localization

It has been shown that CagA targeting to the epithelial cell membrane is important for host signaling (Higashi et al., 2002, Bagnoli et al., 2005, Higashi et al., 2005) and there is evidence that membrane binding of CagA is important for its translocation into host cells (Murata-Kamiya et al., 2010, Steininger et al., 2011).

Recently, it has been proposed that interaction of CagA with host cell proteins depends on the localization of CagA into distinct compartments in host cells (Pelz et al., 2011). The model proposed states that membrane localization is mediated by two membrane domains, one composed by residues 1-200 (N-terminal domain) and the other from 200-1216 (C-terminal domain), detailed description of the function of these domains are described in the Figure 2. In addition, C-terminal domain is comprised of two regions that interact with each other and are located in the membrane together when linked to N terminal domain, the first region is from amino acids 200-800 containing phosphatidyl serine binding domain and the second from 800-1216 containing EPIYA and CagA multimerization motifs. The region from 800-1216 when expressed alone localizes in cytosol and produces a strong alteration in cell morphology (Steininger et al., 2011).

According with the above-mentioned model, CagA spatial conformation may play an important role for the proper translocation and compartmentalization, so a tridimensional structure prediction of CagA was made using I-TASSER, which has been ranked as the best method for the automated protein structure prediction in the last two community-wide critical assessment of protein structure prediction (CASP) experiments (Roy et al., 2010). The best predicted model for both East-Asia and Western type of CagA consists of a sinuous super helical structure in which N-terminal region protrudes at one side and C-terminal region forms a loop (Figure 3). These predicted structural models agree with the model proposed of Steininger et al., in which amino acids 200-1216 forms two domains that interact with each other (Steininger et al., 2011). According to the results obtained with I-TASSER the most similar structure available in the PDB database (www.rcsb.org/pdb/home/home.do) is from Cand1 protein, however the only function similarity shared with CagA is the capacity of protein binding.

Interestingly, it is also proposed that N-terminal membrane-binding domain functions as an inhibitory domain of the C-terminal region effects, and that region from 800-1216 when ex-

pressed alone induces the highest percentage of epithelial cell elongation (Pelz et al., 2011). Also it has been proposed that CagA is cleaved by proteases in human cells yielding two fragments of about 100-105 kDa and 35-40 kDa (Moese et al., 2001), the last would correspond to the region from amino acid 800-1216. This leads to the question if some structural variants of CagA are better cleaved inside the host cell so that can not be totally inhibited by the N-terminal membrane-binding domain, but this requires investigation.

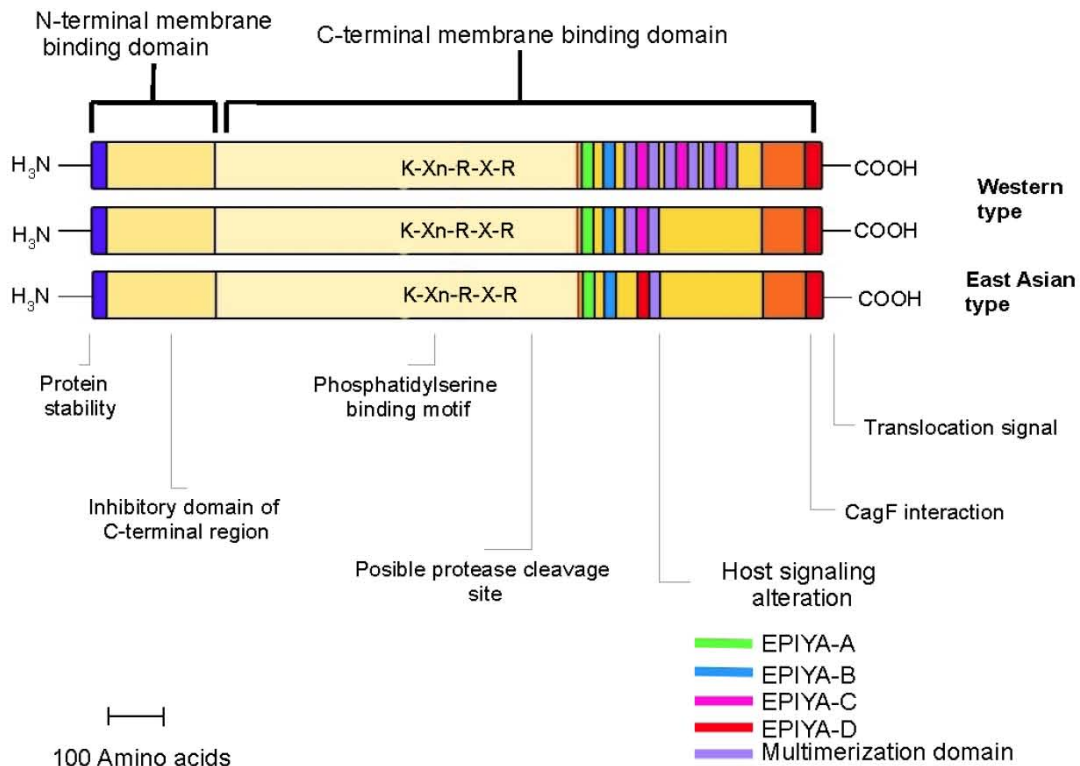


Figure 2. CagA structure in schematic representation. The most common variants of western and East Asian type are depicted. CagA possess at least two membrane-binding domains (Steininger et al., 2011), which are marked with quotation. Main known functions of certain regions are mentioned.

10. Phosphorylation-dependent cell signaling of CagA

Phosphorylation of CagA deregulates several signaling pathways leading to modification in host cell shape and adhesion, contributing to the transformation of cells (Hatakeyama, 2011, Jones et al., 2010). All of the EPIYA motifs in the four distinct EPIYA segments serve as tyrosine phosphorylation sites of CagA in host cells (Hatakeyama, 2011). Enzymes that can phosphorylate CagA *in vitro* and during infection *in vivo* are members of the Src and Abl families, which have been, recognized as oncogenic tyrosine kinases (Selbach et al., 2002, Stein et al., 2002,

Poppe et al., 2007, Tammer et al., 2007). The phosphorylation is a hierarchic regulated process in which c-Src phosphorylates EPIYA-C or EPIYA-D early in infection, while c-Abl phosphorylates either EPIYA-A, EPIYA-B, EPIYA-C, or EPIYA-D later in infection but only 2 EPIYA sites can be simultaneously phosphorylated per CagA molecule (Mueller et al., 2012). Furthermore, simultaneous phosphorylation of EPIYA-AC or EPIYA-BD confers strongest elongation phenotype in an AGS cellular model, even in a single CagA protein with both of the phosphorylatable EPIYA motifs (preferably EPIYA-AC), or in 2 independent CagA proteins, each with a single phosphorylatable EPIYA motif, which may form dimers to trigger this response (Mueller et al., 2012).

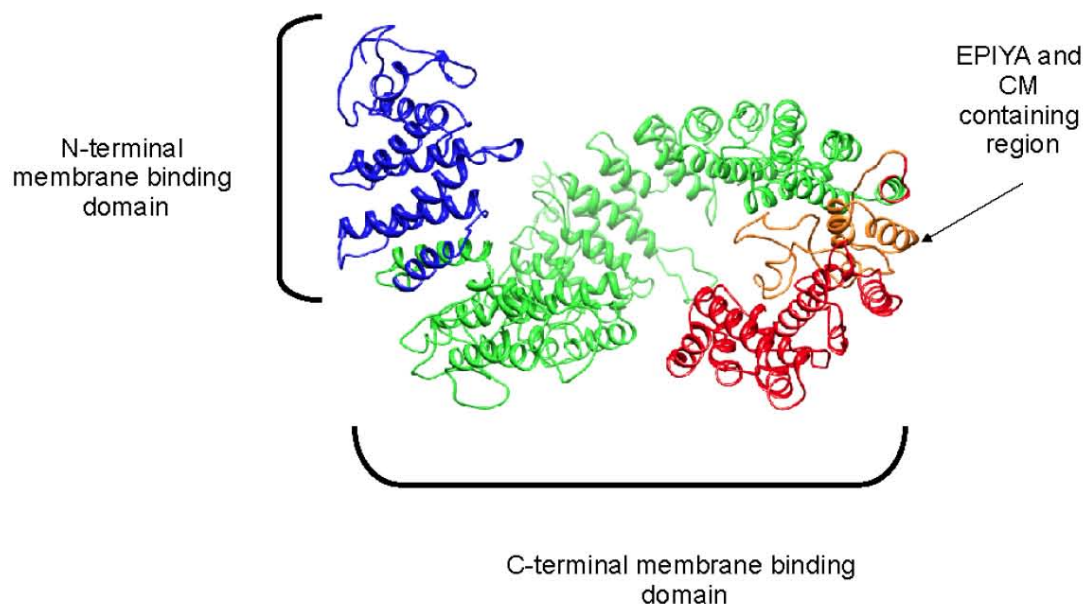


Figure 3. Tridimensional structure of CagA. The model was predicted using I-TASSER, image was obtained using UCSF CHIMERA software (Pettersen et al., 2004).

Once tyrosine phosphorylated, EPIYA-C or -D can bind to SHP-2 phosphatase (Hatakeyama, 2011) and activates the phosphatase activity of Shp2 producing an extremely elongated cell shape known as the hummingbird phenotype and altered cell motility (Backert et al., 2010, Higashi et al., 2004, Tsutsumi et al., 2006). SHP2 is required for full activation of the MAPK/ERK pathway, which conveys a potent mitogenic signal (Matozaki et al., 2009) and for direct dephosphorylation and inactivation of focal adhesion kinase, FAK, that deregulates integrin signaling (Tsutsumi et al., 2006).

Carrying the type EPIYA-D or multiple EPIYA-C repeats is associated with increased SHP-2 phosphatase activity induced by CagA, increasing the risk of precancerous and cancer lesions. This is because East-specific D segment shows higher binding affinity to SHP2 than the West-specific C segment but the C-type motif is variable in copy number and sequence, which can provide more sites for SHP2 binding.

On the other hand, phosphorylated EPIYA-A and EPIYA-B have been shown ability to bind to Csk (Tsutsumi et al., 2003). Interaction of CagA with Csk stimulates the kinase activity of Csk, which in turn inhibits Src in a negative feedback loop that constrains the phosphorylation-dependent CagA activity below a certain threshold to ensure long-term colonization of *H. pylori* in the stomach without causing fatal damages (Hatakeyama, 2011). In addition, phosphorylated EPIYA-B can interact with phosphatidylinositol 3-kinase (PI3K) that regulates diverse cellular functions including proliferation and survival via AKT pathway (Nagy et al., 2009).

11. Phosphorylation-independent cell signaling of CagA

Recent studies have shown that not all cellular interactions exerted by CagA depend on its tyrosine phosphorylation so far more than 13 cellular interaction partners of non-phosphorylated CagA have been identified (Backert et al., 2010, Tegtmeyer et al., 2009, Mimuro et al., 2002). It has been shown that these interactions exert effects such as pro-inflammatory and mitogenic responses, disruption of cell-to-cell junctions or loss of cell polarity.

The first cellular interaction partner described of non-phosphorylated CagA was the adapter protein Gbr2 (Mimuro et al., 2002). Recently, it was reported that Gbr2 is the only host factor that interacts with both phosphorylated and non-phosphorylated EPIYA motifs (Selbach et al., 2009). Non-phosphorylated CagA was shown to interact with Gbr2 *both in vitro* and *in vivo*, which provides a mechanism by which Gbr2-associated SOS (son of sevenless) is recruited to the plasma membrane, this complex (CagA-Gbr2-SOS) can promote Ras-GTP formation, leading to cell scattering by stimulation of the MAPK/ERK signaling pathway as well as activation of nuclear transcription factors involved in cell proliferation and expression of the anti-apoptotic myeloid cell leukemia sequence-1 protein (Mimuro et al., 2007).

On the other hand, Chang YJ, et al., 2006, reported that *cagA*-positive *H. pylori* strains affect the cell cycle during infection by stimulating cyclin D1 expression, G1-S phase progression and host cell survival, suggesting the presence of distinct functional domains within CagA that play essential roles in protein targeting and alteration of host-cell transcription signaling pathways (Chang et al., 2006).

Yokoyama et al., 2005, found that CagA can stimulate the calcium-dependent serine/threonine phosphatase calcineurin in gastric epithelial cells, inducing translocation of the nuclear factor of activated T cells (NFAT) from the cytoplasm to the nucleus, where activates NFAT-dependent genes (Yokoyama et al., 2005). It is possible that the reported CagA-PLC interaction triggers Ca^{2+} mobilization and subsequent activation of calcineurin.

One of the NFAT-dependent genes activated by CagA in gastric epithelial cells is p21Cip1, a cyclin-dependent kinase inhibitor. Accordingly, although CagA activates a growth-promoting signal via the SHP-2-MAP kinase pathway or the Grb2-Ras-MAP kinase pathway, it simultaneously inhibits cell proliferation through NFAT-dependent p21Cip1 induction. Intriguingly, the *H. pylori*-vacuolating toxin *vacA* counteracts the activity of CagA to stimulate NFAT. Thus,

VacA has a role in determining the magnitude of NFAT deregulation in gastric epithelial cells expressing CagA. Such a functional interplay between CagA and VacA has already been suggested from the observation that secretion of VacA protein is associated with the presence of CagA despite the presence of *vacA* gene in all *H. pylori* strains (Yokoyama et al., 2005).

The CagA multimerization sequence mediates a stable binding with SHP-2 protein potentiating CagA virulence and is also responsible of the phosphorylation-independent activity of CagA (Ren et al., 2006, Nagase et al., 2011). Furthermore, it has been shown that CagA binds and inhibits the polarity-regulating kinase partitioning-defective 1 (PAR1)/microtubule affinity-regulating kinase (MARK) through CagA multimerization sequence (Nagase et al., 2011). Kurashima et al., showed that CagA multimerization sequence destabilizes the E-cadherin/beta-catenin complex to trigger aberrant activation of the beta-catenin signal that underlies intestinal metaplasia (Kurashima et al., 2008). Recently, it was shown that the diversity in the structure and number of CagA multimerization sequences influence the strength of pathological activities of individual CagA with its cellular targets (PAR1b and SHP-2), emphasising its importance as a determinant of the degree of virulence of CagA (Lu et al., 2008) for development of gastric cancer.

A recent study shows that CagA can also interact with runt-related transcription factor gene 3 (RUNX3, a tumor suppressor that is frequently inactivated in gastric cancer) by a novel identified WW domain in the amino-terminal region of CagA (Tsang et al., 2010). In particular, CagA induces the ubiquitination and degradation of RUNX3, thereby extinguishing its ability to inhibit the transcriptional activation of RUNX3; currently, functional inactivation of RUNX3 can be caused by mutation, epigenetic gene silencing (promoter methylation) or cytoplasmic protein mislocalization (Ito, 2011). Recent evidence shows a strong association between methylation of the RUNX3 promoter and gastric cancer (Fan et al., 2011, Li et al., 2011). Recently, Lin et al., found that loss of RUNX3 in gastric cancer prevents G1 phase arrest and promotes tumorigenesis through Akt1/ β -catenin/cyclin D1 signaling pathway (Lin et al., 2012).

A consequence of non-phosphorylated CagA interactions in polarized epithelial cells is the disruption of cell junctions. In particular, tight and adherence junctions are essential for the integrity of the gastric epithelium (Wessler & Backert, 2008). CagA interferes with these intercellular junctions via several pathways. 1) Injected CagA associates with the epithelial tight-junction scaffolding protein, zone occludens-1 (ZO-1), and the transmembrane protein, junction adhesion molecule (JAM2), causing an ectopic assembly of tight-junction components at sites of bacterial attachment (Amieva & El-Omar, 2008), leading to alterations of cell polarity. 2) Non-phosphorylated CagA can interact with the transmembrane cell-cell junction protein E-cadherin (Hatakeyama, 2008). Subsequently, it was found that CagA forms a complex with c-Met recruiting E-cadherin and the Armadillo-domain protein, catenin delta, indicating that the interaction between CagA and E-cadherin is not direct (Oliveira et al., 2009). Baek et al., in 2007, reported that the interaction of alpha-Pix (PAK-interactive exchange factor) with CagA activates PAK1, ERK and NF- κ B, which induces IL-8 expression (Baek et al., 2007, Lim et al., 2009); beta 1 integrin (integrin β 1) is another CagA-independent signaling that promotes cell

motility (Snider et al., 2008), although the functional importance of this interaction needs to be investigated.

12. Concluding remarks

Studies host interactions with *H. pylori* virulence factors, such as CagA and the T4SS, have provided us many fundamental insights into the processes leading to *H. pylori* pathogenesis. The current hypothesis implies a model with translocated CagA as an "eukaryotic" signaling mimetic molecule either present in a large multiprotein complex or simultaneously in separated locations within infected target cells, which may have an important impact on the multistep pathogenesis of *H. pylori*. In the future, it will be important to search for additional injected proteins because it is rather unlikely that *cag*-PAI acquisition during evolution was intended to inject few effectors like CagA and peptidoglycan. Future studies are required to uncover the complex mechanisms that explain how *H. pylori* interacts with host cells at the molecular level to accomplish its persistence through host life.

Author details

Stephanie E. Morales-Guerrero¹, Eduardo Mucito-Varela¹, Germán Rubén Aguilar-Gutiérrez², Yolanda Lopez-Vidal¹ and Gonzalo Castillo-Rojas¹

¹ Programa de Inmunología Molecular Microbiana, Departamento de Microbiología y Parasitología, Facultad de Medicina, Universidad Nacional Autónoma de México (UNAM), México City, México

² Centro de Investigación Sobre Enfermedades Infecciosas (CISEI), Instituto Nacional de Salud Pública (INSP). Cuernavaca, Morelos, México

References

- [1] Aguilar, G. R., G. Ayala & G. Fierros-Zarate, (2001) *Helicobacter pylori*: recent advances in the study of its pathogenicity and prevention. *Salud Publica Mex* 43: 237-247.
- [2] Alvarez-Martinez, C. E. & P. J. Christie, (2009) Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev* 73: 775-808.
- [3] Amieva, M. R. & E. M. El-Omar, (2008) Host-bacterial interactions in *Helicobacter pylori* infection. *Gastroenterology* 134: 306-323.
- [4] Antonio-Rincon, F., Y. Lopez-Vidal, G. Castillo-Rojas, E. C. Lazcano-Ponce, S. Ponce-de-Leon, M. L. Tabche-Barrera & G. R. Aguilar-Gutierrez, (2011) Pathogenicity island

- cag*, *vacA* and IS605 genotypes in Mexican strains of *Helicobacter pylori* associated with peptic ulcers. *Ann Clin Microbiol Antimicrob* 10: 18.
- [5] Atherton, J. C., P. Cao, R. M. Peek, Jr., M. K. Tummuru, M. J. Blaser & T. L. Cover, (1995) Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. Association of specific *vacA* types with cytotoxin production and peptic ulceration. *J Biol Chem* 270: 17771-17777.
- [6] Backert, S., R. Fronzes & G. Waksman, (2008) VirB2 and VirB5 proteins: specialized adhesins in bacterial type-IV secretion systems? *Trends Microbiol* 16: 409-413.
- [7] Backert, S. & T. F. Meyer, (2006) Type IV secretion systems and their effectors in bacterial pathogenesis. *Curr Opin Microbiol* 9: 207-217.
- [8] Backert, S., N. Tegtmeyer & M. Selbach, (2010) The versatility of *Helicobacter pylori* CagA effector protein functions: The master key hypothesis. *Helicobacter* 15: 163-176.
- [9] Baek, H. Y., J. W. Lim & H. Kim, (2007) Interaction between the *Helicobacter pylori* CagA and alpha-Pix in gastric epithelial AGS cells. *Ann N Y Acad Sci* 1096: 18-23.
- [10] Bagnoli, F., L. Buti, L. Tompkins, A. Covacci & M. R. Amieva, (2005) *Helicobacter pylori* CagA induces a transition from polarized to invasive phenotypes in MDCK cells. *Proc Natl Acad Sci U S A* 102: 16339-16344.
- [11] Boren, T., P. Falk, K. A. Roth, G. Larson & S. Normark, (1993) Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens. *Science* 262: 1892-1895.
- [12] Boughan, P. K., R. H. Argent, M. Body-Malapel, J. H. Park, K. E. Ewings, A. G. Bowie, S. J. Ong, S. J. Cook, O. E. Sorensen, B. A. Manzo, N. Inohara, N. J. Klein, G. Nunez, J. C. Atherton & M. Bajaj-Elliott, (2006) Nucleotide-binding oligomerization domain-1 and epidermal growth factor receptor: critical regulators of beta-defensins during *Helicobacter pylori* infection. *J Biol Chem* 281: 11637-11648.
- [13] Castillo-Rojas, G., M. Mazari-Hiriart & Y. Lopez-Vidal, (2004) [*Helicobacter pylori*: focus on CagA and VacA major virulence factors]. *Salud Publica Mex* 46: 538-548.
- [14] Censini, S., C. Lange, Z. Xiang, J. E. Crabtree, P. Ghiara, M. Borodovsky, R. Rappuoli & A. Covacci, (1996) *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc Natl Acad Sci U S A* 93: 14648-14653.
- [15] Chang, Y. J., M. S. Wu, J. T. Lin, R. G. Pestell, M. J. Blaser & C. C. Chen, (2006) Mechanisms for *Helicobacter pylori* CagA-induced cyclin D1 expression that affect cell cycle. *Cell Microbiol* 8: 1740-1752.
- [16] Covacci, A., S. Censini, M. Bugnoli, R. Petracca, D. Burrone, G. Macchia, A. Massone, E. Papini, Z. Xiang, N. Figura & et al., (1993) Molecular characterization of the 128-kDa immunodominant antigen of *Helicobacter pylori* associated with cytotoxicity and duodenal ulcer. *Proc Natl Acad Sci U S A* 90: 5791-5795.

- [17] Fan, X. Y., X. L. Hu, T. M. Han, N. N. Wang, Y. M. Zhu, W. Hu, Z. H. Ma, C. J. Zhang, X. Xu, Z. Y. Ye, C. M. Han & W. S. Pan, (2011) Association between RUNX3 promoter methylation and gastric cancer: a meta-analysis. *BMC Gastroenterol* 11: 92.
- [18] Ferlay J, S. H., Bray F, Forman D, Mathers C and Parkin DM, (2010) GLOBOCAN 2008, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10. In: Lyon, France: International Agency for Research on Cancer. pp.
- [19] Figueiredo, C., J. C. Machado & Y. Yamaoka, (2005) Pathogenesis of *Helicobacter pylori* Infection. *Helicobacter* 10 Suppl 1: 14-20.
- [20] Fronzes, R., P. J. Christie & G. Waksman, (2009) The structural biology of type IV secretion systems. *Nat Rev Microbiol* 7: 703-714.
- [21] Furuta, Y., K. Yahara, M. Hatakeyama & I. Kobayashi, (2011) Evolution of *cagA* oncogene of *Helicobacter pylori* through recombination. *PLoS One* 6: e23499.
- [22] Gebert, B., W. Fischer, E. Weiss, R. Hoffmann & R. Haas, (2003) *Helicobacter pylori* vacuolating cytotoxin inhibits T lymphocyte activation. *Science* 301: 1099-1102.
- [23] Gerhard, M., N. Lehn, N. Neumayer, T. Boren, R. Rad, W. Schepp, S. Miehle, M. Classen & C. Prinz, (1999) Clinical relevance of the *Helicobacter pylori* gene for blood-group antigen-binding adhesin. *Proc Natl Acad Sci U S A* 96: 12778-12783.
- [24] Harris, P. R., H. L. Mobley, G. I. Perez-Perez, M. J. Blaser & P. D. Smith, (1996) *Helicobacter pylori* urease is a potent stimulus of mononuclear phagocyte activation and inflammatory cytokine production. *Gastroenterology* 111: 419-425.
- [25] Hatakeyama, M., (2004) Oncogenic mechanisms of the *Helicobacter pylori* CagA protein. *Nat Rev Cancer* 4: 688-694.
- [26] Hatakeyama, M., (2008) SagA of CagA in *Helicobacter pylori* pathogenesis. *Curr Opin Microbiol* 11: 30-37.
- [27] Hatakeyama, M., (2011) Anthropological and clinical implications for the structural diversity of the *Helicobacter pylori* CagA oncoprotein. *Cancer Sci* 102: 36-43.
- [28] Higashi, H., A. Nakaya, R. Tsutsumi, K. Yokoyama, Y. Fujii, S. Ishikawa, M. Higuchi, A. Takahashi, Y. Kurashima, Y. Teishikata, S. Tanaka, T. Azuma & M. Hatakeyama, (2004) *Helicobacter pylori* CagA induces Ras-independent morphogenetic response through SHP-2 recruitment and activation. *J Biol Chem* 279: 17205-17216.
- [29] Higashi, H., R. Tsutsumi, S. Muto, T. Sugiyama, T. Azuma, M. Asaka & M. Hatakeyama, (2002) SHP-2 tyrosine phosphatase as an intracellular target of *Helicobacter pylori* CagA protein. *Science* 295: 683-686.
- [30] Higashi, H., K. Yokoyama, Y. Fujii, S. Ren, H. Yuasa, I. Saadat, N. Murata-Kamiya, T. Azuma & M. Hatakeyama, (2005) EPIYA motif is a membrane-targeting signal of *Helicobacter pylori* virulence factor CagA in mammalian cells. *J Biol Chem* 280: 23130-23137.

- [31] IARC, (1994) Schistosomes, liver flukes and *Helicobacter pylori*. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7-14 June 1994. *IARC Monogr Eval Carcinog Risks Hum* 61: 1-241.
- [32] Ito, K., (2011) RUNX3 in oncogenic and anti-oncogenic signaling in gastrointestinal cancers. *J Cell Biochem* 112: 1243-1249.
- [33] Jones, K. R., J. M. Whitmire & D. S. Merrell, (2010) A tale of two toxins: *Helicobacter pylori* CagA and VacA modulate host pathways that impact disease. *Frontiers in Microbiology* 1: 1-17.
- [34] Kersulyte, D., A. Kalia, R. H. Gilman, M. Mendez, P. Herrera, L. Cabrera, B. Velapatino, J. Balqui, F. Paredes Puente de la Vega, C. A. Rodriguez Ulloa, J. Cok, C. C. Hooper, G. Dailide, S. Tamma & D. E. Berg, (2010) *Helicobacter pylori* from Peruvian amerindians: traces of human migrations in strains from remote Amazon, and genome sequence of an Amerind strain. *PLoS One* 5(11): e15076.
- [35] Kurashima, Y., N. Murata-Kamiya, K. Kikuchi, H. Higashi, T. Azuma, S. Kondo & M. Hatakeyama, (2008) Deregulation of beta-catenin signal by *Helicobacter pylori* CagA requires the CagA-multimerization sequence. *Int J Cancer* 122: 823-831.
- [36] Leunk, R. D., P. T. Johnson, B. C. David, W. G. Kraft & D. R. Morgan, (1988) Cytotoxic activity in broth-culture filtrates of *Campylobacter pylori*. *J Med Microbiol* 26: 93-99.
- [37] Li, W. Q., K. F. Pan, Y. Zhang, C. X. Dong, L. Zhang, J. L. Ma, T. Zhou, J. Y. Li & W. C. You, (2011) RUNX3 methylation and expression associated with advanced precancerous gastric lesions in a Chinese population. *Carcinogenesis* 32: 406-410.
- [38] Lim, J. W., K. H. Kim & H. Kim, (2009) alphaPix interacts with *Helicobacter pylori* CagA to induce IL-8 expression in gastric epithelial cells. *Scand J Gastroenterol* 44: 1166-1172.
- [39] Lin, F. C., Y. P. Liu, C. H. Lai, Y. S. Shan, H. C. Cheng, P. I. Hsu, C. H. Lee, Y. C. Lee, H. Y. Wang, C. H. Wang, J. Q. Cheng, M. Hsiao & P. J. Lu, (2012) RUNX3-mediated transcriptional inhibition of Akt suppresses tumorigenesis of human gastric cancer cells. *Oncogene*.
- [40] Lopez-Vidal, Y., S. Ponce-de-Leon, G. Castillo-Rojas, R. Barreto-Zuniga & A. Torre-Delgadillo, (2008) High diversity of *vacA* and *cagA* *Helicobacter pylori* genotypes in patients with and without gastric cancer. *PLoS One* 3: e3849.
- [41] Lu, H. S., Y. Saito, M. Umeda, N. Murata-Kamiya, H. M. Zhang, H. Higashi & M. Hatakeyama, (2008) Structural and functional diversity in the PAR1b/MARK2-binding region of *Helicobacter pylori* CagA. *Cancer Sci* 99: 2004-2011.
- [42] Mai, U. E., G. I. Perez-Perez, L. M. Wahl, S. M. Wahl, M. J. Blaser & P. D. Smith, (1991) Soluble surface proteins from *Helicobacter pylori* activate monocytes/macrophages by lipopolysaccharide-independent mechanism. *J Clin Invest* 87: 894-900.
- [43] Mane, S. P., M. G. Dominguez-Bello, M. J. Blaser, B. W. Sobral, R. Hontecillas, J. Skoneczka, S. K. Mohapatra, O. R. Crasta, C. Evans, T. Modise, S. Shallom, M. Shukla,

- C. Varon, F. Megraud, A. L. Maldonado-Contreras, K. P. Williams & J. Bassaganya-Riera, (2010) Host-interactive genes in Amerindian *Helicobacter pylori* diverge from their Old World homologs and mediate inflammatory responses. *J Bacteriol* 192: 3078-3092.
- [44] Matozaki, T., Y. Murata, Y. Saito, H. Okazawa & H. Ohnishi, (2009) Protein tyrosine phosphatase SHP-2: a proto-oncogene product that promotes Ras activation. *Cancer Sci* 100: 1786-1793.
- [45] Megraud, F., V. Neman-Simha & D. Brugmann, (1992) Further evidence of the toxic effect of ammonia produced by *Helicobacter pylori* urease on human epithelial cells. *Infect Immun* 60: 1858-1863.
- [46] Mimuro, H., T. Suzuki, S. Nagai, G. Rieder, M. Suzuki, T. Nagai, Y. Fujita, K. Nagamatsu, N. Ishijima, S. Koyasu, R. Haas & C. Sasakawa, (2007) *Helicobacter pylori* dampens gut epithelial self-renewal by inhibiting apoptosis, a bacterial strategy to enhance colonization of the stomach. *Cell Host Microbe* 2: 250-263.
- [47] Mimuro, H., T. Suzuki, J. Tanaka, M. Asahi, R. Haas & C. Sasakawa, (2002) Grb2 is a key mediator of *Helicobacter pylori* CagA protein activities. *Mol Cell* 10: 745-755.
- [48] Mobley, H. L., G. L. Mendz & S. L. Hazell, (2001) *Helicobacter pylori: physiology and genetics*, p. 626. ASM Press, Washington DC.
- [49] Moese, S., M. Selbach, U. Zimny-Arndt, P. R. Jungblut, T. F. Meyer & S. Backert, (2001) Identification of a tyrosine-phosphorylated 35 kDa carboxy-terminal fragment (p35CagA) of the *Helicobacter pylori* CagA protein in phagocytic cells: processing or breakage? *Proteomics* 1: 618-629.
- [50] Montecucco, C. & R. Rappuoli, (2001) Living dangerously: how *Helicobacter pylori* survives in the human stomach. *Nat Rev Mol Cell Biol* 2: 457-466.
- [51] Moran, A., (1998) The products of *Helicobacter pylori* that induce inflammation. *Eur. J. Gastroenterol. Hepatol.* 10 Suppl 1: S3-S8.
- [52] Moran, A. P., (1995) Cell surface characteristics of *Helicobacter pylori*. *FEMS Immunol Med Microbiol* 10: 271-280.
- [53] Mueller, D., N. Tegtmeyer, S. Brandt, Y. Yamaoka, E. De Poire, D. Sgouras, S. Wessler, J. Torres, A. Smolka & S. Backert, (2012) c-Src and c-Abl kinases control hierarchic phosphorylation and function of the CagA effector protein in Western and East Asian *Helicobacter pylori* strains. *J Clin Invest* 122: 1553-1566.
- [54] Murata-Kamiya, N., K. Kikuchi, T. Hayashi, H. Higashi & M. Hatakeyama, (2010) *Helicobacter pylori* exploits host membrane phosphatidylserine for delivery, localization, and pathophysiological action of the CagA oncoprotein. *Cell Host Microbe* 7: 399-411.
- [55] Nagase, L., N. Murata-Kamiya & M. Hatakeyama, (2011) Potentiation of *Helicobacter pylori* CagA protein virulence through homodimerization. *J Biol Chem* 286: 33622-33631.

- [56] Nagy, T. A., M. R. Frey, F. Yan, D. A. Israel, D. B. Polk & R. M. Peek, Jr., (2009) *Helicobacter pylori* regulates cellular migration and apoptosis by activation of phosphatidylinositol 3-kinase signaling. *J Infect Dis* 199: 641-651.
- [57] Nesic, D., M. C. Miller, Z. T. Quinkert, M. Stein, B. T. Chait & C. E. Stebbins, (2010) *Helicobacter pylori* CagA inhibits PAR1-MARK family kinases by mimicking host substrates. *Nat Struct Mol Biol* 17: 130-132.
- [58] Oldani, A., M. Cormont, V. Hofman, V. Chiozzi, O. Oregioni, A. Canonici, A. Sciallo, P. Sommi, A. Fabbri, V. Ricci & P. Boquet, (2009) *Helicobacter pylori* counteracts the apoptotic action of its VacA toxin by injecting the CagA protein into gastric epithelial cells. *PLoS Pathog* 5: e1000603.
- [59] Oliveira, M. J., A. M. Costa, A. C. Costa, R. M. Ferreira, P. Sampaio, J. C. Machado, R. Seruca, M. Mareel & C. Figueiredo, (2009) CagA associates with c-Met, E-cadherin, and p120-catenin in a multiproteic complex that suppresses *Helicobacter pylori*-induced cell-invasive phenotype. *J Infect Dis* 200: 745-755.
- [60] Papini, E., B. Satin, N. Norais, M. de Bernard, J. L. Telford, R. Rappuoli & C. Montecucco, (1998) Selective increase of the permeability of polarized epithelial cell monolayers by *Helicobacter pylori* vacuolating toxin. *J Clin Invest* 102: 813-820.
- [61] Peek, R. M., Jr. & M. J. Blaser, (2002) *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. *Nat Rev Cancer* 2: 28-37.
- [62] Pelz, C., S. Steininger, C. Weiss, F. Coscia & R. Vogelmann, (2011) A novel inhibitory domain of *Helicobacter pylori* protein CagA reduces CagA effects on host cell biology. *J Biol Chem* 286: 8999-9008.
- [63] Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng & T. E. Ferrin, (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605-1612.
- [64] Poppe, M., S. M. Feller, G. Romer & S. Wessler, (2007) Phosphorylation of *Helicobacter pylori* CagA by c-Abl leads to cell motility. *Oncogene* 26: 3462-3472.
- [65] Ren, S., H. Higashi, H. Lu, T. Azuma & M. Hatakeyama, (2006) Structural basis and functional consequence of *Helicobacter pylori* CagA multimerization in cells. *J Biol Chem* 281: 32344-32352.
- [66] Ricci, V., M. Romano & P. Boquet, (2011) Molecular cross-talk between *Helicobacter pylori* and human gastric mucosa. *World J Gastroenterol* 17: 1383-1399.
- [67] Roy, A., A. Kucukural & Y. Zhang, (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5: 725-738.
- [68] Selbach, M., S. Moese, C. R. Hauck, T. F. Meyer & S. Backert, (2002) Src is the kinase of the *Helicobacter pylori* CagA protein in vitro and in vivo. *J Biol Chem* 277: 6775-6778.

- [69] Selbach, M., F. E. Paul, S. Brandt, P. Guye, O. Daumke, S. Backert, C. Dehio & M. Mann, (2009) Host cell interactome of tyrosine-phosphorylated bacterial proteins. *Cell Host Microbe* 5: 397-403.
- [70] Snider, J. L., C. Allison, B. H. Bellaire, R. L. Ferrero & J. A. Cardelli, (2008) The beta1 integrin activates JNK independent of CagA, and JNK activation is required for *Helicobacter pylori* CagA+-induced motility of gastric cancer cells. *J Biol Chem* 283: 13952-13963.
- [71] Sobala, G. M., C. J. Schorah, S. Shires, D. A. Lynch, B. Gallacher, M. F. Dixon & A. T. Axon, (1993) Effect of eradication of *Helicobacter pylori* on gastric juice ascorbic acid concentrations. *Gut* 34: 1038-1041.
- [72] Stein, M., F. Bagnoli, R. Halenbeck, R. Rappuoli, W. J. Fantl & A. Covacci, (2002) c-Src/Lyn kinases activate *Helicobacter pylori* CagA through tyrosine phosphorylation of the EPIYA motifs. *Mol Microbiol* 43: 971-980.
- [73] Steininger, S., C. Pelz & R. Vogelmann, (2011) Purpose of recently detected inhibitory domain of the *Helicobacter pylori* protein CagA. *Gut Microbes* 2: 167-172.
- [74] Suzuki, M., S. Miura, M. Suematsu, D. Fukumura, I. Kurose, H. Suzuki, A. Kai, Y. Kudoh, M. Ohashi & M. Tsuchiya, (1992) *Helicobacter pylori*-associated ammonia production enhances neutrophil-dependent gastric mucosal cell injury. *Am J Physiol* 263: G719-725.
- [75] Tammer, I., S. Brandt, R. Hartig, W. Konig & S. Backert, (2007) Activation of Abl by *Helicobacter pylori*: a novel kinase for CagA and crucial mediator of host cell scattering. *Gastroenterology* 132: 1309-1319.
- [76] Tegtmeyer, N., D. Zabler, D. Schmidt, R. Hartig, S. Brandt & S. Backert, (2009) Importance of EGF receptor, HER2/Neu and Erk1/2 kinase signalling for host cell elongation and scattering induced by the *Helicobacter pylori* CagA protein: antagonistic effects of the vacuolating cytotoxin VacA. *Cell Microbiol* 11: 488-505.
- [77] Tomb, J. F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzgerald, N. Lee, M. D. Adams, E. K. Hickey, D. E. Berg, J. D. Gocayne, T. R. Utterback, J. D. Peterson, J. M. Kelley, M. D. Cotton, J. M. Weidman, C. Fujii, C. Bowman, L. Watthey, E. Wallin, W. S. Hayes, M. Borodovsky, P. D. Karp, H. O. Smith, C. M. Fraser & J. C. Venter, (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388: 539-547.
- [78] Tsang, Y. H., A. Lamb, J. Romero-Gallo, B. Huang, K. Ito, R. M. Peek, Jr., Y. Ito & L. F. Chen, (2010) *Helicobacter pylori* CagA targets gastric tumor suppressor RUNX3 for proteasome-mediated degradation. *Oncogene* 29: 5643-5650.

- [79] Tsutsumi, R., H. Higashi, M. Higuchi, M. Okada & M. Hatakeyama, (2003) Attenuation of *Helicobacter pylori* CagA x SHP-2 signaling by interaction between CagA and C-terminal Src kinase. *J Biol Chem* 278: 3664-3670.
- [80] Tsutsumi, R., A. Takahashi, T. Azuma, H. Higashi & M. Hatakeyama, (2006) Focal adhesion kinase is a substrate and downstream effector of SHP-2 complexed with *Helicobacter pylori* CagA. *Mol Cell Biol* 26: 261-276.
- [81] Tummuru, M. K., T. L. Cover & M. J. Blaser, (1993) Cloning and expression of a high-molecular-mass major antigen of *Helicobacter pylori*: evidence of linkage to cytotoxin production. *Infect Immun* 61: 1799-1809.
- [82] Viala, J., C. Chaput, I. G. Boneca, A. Cardona, S. E. Girardin, A. P. Moran, R. Athman, S. Memet, M. R. Huerre, A. J. Coyle, P. S. DiStefano, P. J. Sansonetti, A. Labigne, J. Bertin, D. J. Philpott & R. L. Ferrero, (2004) Nod1 responds to peptidoglycan delivered by the *Helicobacter pylori* cag pathogenicity island. *Nature immunology* 5: 1166-1174.
- [83] Warren, J. R. M., B, (1983) Unidentified curved bacilli on gastric epithelium in active chronic gastritis. *Lancet* 1: 1273-1275.
- [84] Wessler, S. & S. Backert, (2008) Molecular mechanisms of epithelial-barrier disruption by *Helicobacter pylori*. *Trends Microbiol* 16: 397-405.
- [85] Wroblewski, L. E., R. M. Peek, Jr. & K. T. Wilson, (2010) *Helicobacter pylori* and gastric cancer: factors that modulate disease risk. *Clin Microbiol Rev* 23: 713-739.
- [86] Xia, Y., Y. Yamaoka, Q. Zhu, I. Matha & X. Gao, (2009) A comprehensive sequence and disease correlation analyses for the C-terminal region of CagA protein of *Helicobacter pylori*. *PLoS One* 4: e7736.
- [87] Yamaoka, Y., (2008) Increasing evidence of the role of *Helicobacter pylori* SabA in the pathogenesis of gastroduodenal disease. *J Infect Dev Ctries* 2: 174-181.
- [88] Yamaoka, Y., (2010) Mechanisms of disease: *Helicobacter pylori* virulence factors. *Nat Rev Gastroenterol Hepatol* 7: 629-641.
- [89] Yamaoka, Y., S. Kikuchi, H. M. el-Zimaity, O. Gutierrez, M. S. Osato & D. Y. Graham, (2002) Importance of *Helicobacter pylori* oipA in clinical presentation, gastric inflammation, and mucosal interleukin 8 production. *Gastroenterology* 123: 414-424.
- [90] Yamaoka, Y., D. H. Kwon & D. Y. Graham, (2000) A M(r) 34,000 proinflammatory outer membrane protein (oipA) of *Helicobacter pylori*. *Proc Natl Acad Sci U S A* 97: 7533-7538.
- [91] Yokoyama, K., H. Higashi, S. Ishikawa, Y. Fujii, S. Kondo, H. Kato, T. Azuma, A. Wada, T. Hirayama, H. Aburatani & M. Hatakeyama, (2005) Functional antagonism between *Helicobacter pylori* CagA and vacuolating toxin VacA in control of the NFAT signaling pathway in gastric epithelial cells. *Proc Natl Acad Sci U S A* 102: 9661-9666.

Abstract no.: P02.05

ALTERED EXPRESSION OF PYK-REG-90 NON CODING RNA IN GASTRIC CANCER

S. Lee, J. Lee, C. Park, J. Park, H. Lee, S. Shin and Y. Lee

Yonsei University College of Medicine Severance Hospital, Seoul, Korea

Non-coding RNAs (ncRNAs) exceeds by large the number of protein-coding genes. However, for the vast majority of ncRNA, the function remains unknown. The long ncRNA gene pyk-reg-90 which map to the chromosomal 3 (p21.1) and located in the intergenic region and expression of this transcript was found to altered in colon cancer recently. We therefore explored their significance in gastric cancer.

Pyk-reg-90 and U6 expression level were determined by qRT-PCR in various gastric cancer cell lines and normal tissues, respectively. Fresh 50 gastric cancer tissues and adjacent normal tissues obtained by surgery and clinical data were collected prospectively. For qRT-PCR, total RNA was purified using TRIzol reagent. qRT-PCR analysis was carried out with iQ SYBR Green Supermix. Down regulation of pyk-reg-90 was obtained by three different siRNAs transfection and cell number and viability was checked after treatment of siRNAs. For apoptotic analysis, PARP was determined by western blot and caspase 3/7 were analyzed by luminescent assay.

Compared to normal tissue, expression of pyk-reg-90 was up-regulated in various gastric cancer cell lines. In gastric cancer tissue expression of pyk-reg-90 was significantly high in cancer than adjacent normal tissue ($p < 0.05$). Three different siRNAs for pyk-reg-90 down regulated expression of pyk-reg-90 in SNU-719 cells, and pool of siRNAs repressed expression of pyk-reg-90 by 50%. The number of SNU-719 cells was significantly reduced after 48 hours after siRNA treatment. Treatment with siRNAs increased expression of cleaved PARP-1 and caspase 3/7 compared to scramble. Our results suggested that pyk-reg-90 was upregulated in gastric cancer and down-regulation of a pyk-reg-90 can induce cell death.

Abstract no.: P02.06

GENE EXPRESSION PROFILING IN CHRONIC GASTRITIS ASSOCIATED WITH *HELICOBACTER PYLORI* INFECTIONS. E. Morales-Guerrero,* G. Castillo-Rojas,* C. Rangel-Escareño,[†] I. Imaz-Rosshandler,[‡] S. Ponce de León[‡] and Y. LópezVidal**Universidad Nacional Autónoma de Mexico, Mexico City, Mexico; [†]Instituto Nacional de Medicina Genómica, Mexico City, Mexico; [‡]Instituto Nacional de Ciencias Médicas y de la Nutrición SZ, Mexico City, Mexico

Helicobacter pylori is a gram negative bacterium that colonizes the stomach of almost half of the world's population and the infection usually persists for several decades. The infection with *H. pylori* produces a chronic gastritis, which is asymptomatic in the majority of the infected subjects; however, this chronic inflammation can progress to the development of gastric cancer, for which is considered as a type I carcinogen in humans.

Analysis of host gene expression in subjects with chronic gastritis could help to understand the first molecular mechanism altered in the host that could lead to development of more serious conditions.

Aim: Determine the gene expression profile in subjects infected with *H. pylori* and with chronic gastritis and functional dyspepsia.

Methodology: Samples from patients with chronic gastritis and samples from patients with functional dyspepsia were evaluated through microarray assays. Gene expression values for each gene across all samples were analyzed using the Gene Set Enrichment Analysis tool (GSEA), to determine the gene ontology terms enriched in each of the phenotypes evaluated.

Results and Conclusions: The GSEA analysis in functional dyspepsia and chronic gastritis shows that ontology terms related with mitochondria are enriched in functional dyspepsia, whereas ontology terms related with immune response are enriched in chronic gastritis samples. These results suggest that in chronic gastritis exists a down-regulation of genes associated with mitochondria, which can be related with *H. pylori* infection and could be the first alterations involved in the development of more serious diseases.

Abstract no.: P02.07

GENOME SEQUENCING OF *HELICOBACTER PYLORI* STRAINS ISOLATED IN MEXICO FROM GASTRIC CANCER AND CHRONIC GASTRITIS LESSIONSE. J. Mucito-Varela,* V. Flores-López,[†] G. Castillo-Rojas,* M. A. Cevallos-Gaos* and Y. López-Vidal**Universidad Nacional Autónoma de Mexico, Mexico City, Mexico; [†]CINVESTAV, IPN, Mexico City, Mexico

H. pylori infection is associated to gastritis, peptic ulcer and gastric cancer development. It has been proposed that virulence factors like CagA and VacA containing specific polymorphisms increase the risk of disease development but distribution of these polymorphisms worldwide doesn't always correlate to disease acquisition suggesting that other bacterial factors play an important role in gastric cell damage and transformation. Whole genome sequencing provides information of the total gene content allowing the search for new genetic markers associated to virulence. Representative sequences from different parts of the world must be included to gain insights in common processes of *H. pylori* disease-associated strains. Until now it had not been sequenced any genome of *H. pylori* isolated in Mexico so we obtained the whole genome sequence of two *H. pylori* strains by pyrosequencing. Assembly was done de novo and we finally constructed 17 and 39 contigs that cover more than 92% of the genome and 1.6 Mb combined length. One thousand four hundred genes have been annotated for each genome. The genome sequence of these strains allowed us to do comparative genomics to better understand gastric cancer pathogenesis.

Abstract no.: P02.08

SODIUM BUTYRATE INDUCES FOXO3 DOWNSTREAM SIGNAL PATHWAYS BY PTEN MODULATION IN GASTRIC CANCER CELL

Y. Lee, D. Yu, S. Kim, Y. Lee and D. Lee

Yonsei University College of Medicine, Seoul, Korea

Background/Aims: Sodium Butyrate (NaB) has various effects including the regulation of cell growth and differentiation. PTEN (phosphatase and tensin homologue deleted on chromosome 10) is known to promote gastrointestinal cell differentiation. Forkhead transcription factors of class O (FoxO) are implicated in the regulation of apoptosis, cell survival, and pathogenesis. In this study, we aimed to determine whether NaB regulated the expression of the PTEN through the PI3K/Akt pathway and examine the relationship between PTEN and FoxO3a in gastric cancer.

Method: Gastric cancer cell line AGS was used. NaB 2 μ mol/L treated cells for respective hours were used. The cells were used for proliferation assay or were lysed directly for protein analysis. The effects of NaB on gastric cancer cells in PI3K/PTEN/AKT signal pathways were analyzed by using MTS assay, western blotting, and qRT-PCR. siPTEN or inhibitors for PI3K/AKT signal pathways were used respectively. FHRE luciferase assay were employed to measure FOXO signal activity.

Results: AGS cells treated with NaB for 24–72 hours showed the inhibition of cell growth in a time dependent manner. Cell growth was significantly inhibited by 50% at 72 hours. NaB treatment markedly increased the expression of PTEN, while down-regulating the PI3K/Akt pathway. NaB treatment resulted in the enhancement the translocation of FoxO3a to the nucleus. The induction of FoxO3a's downstream target genes, GADD45a and p27 was confirmed by the western blotting and qRT-PCR.

Conclusion: The increase of PTEN activity by NaB induce the induction of FoxO3a through the down regulation of PI3K/Akt signaling pathway.

Bibliografía

- [1] N. Alikhan, N.K Petty, N.L.B. Zakour, y *et al.* BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC genomics*, 12(1):402, 2011.
- [2] R.A. Alm, L.S. Ling, D.T. Moir, y *et al.* Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature*, 397(6715):176–80, 1999.
- [3] S.F. Altschul, T.L. Madden, A. Schäffer, y *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res*, 25(17):3389–3402, 1997.
- [4] M.R. Amieva y E.M. El-Omar. Host-bacterial interactions in *Helicobacter pylori* infection. *Gastroenterology*, 134(1):306–23, 2008.
- [5] F. Antonio-Rincón, Y. López-Vidal, G. Castillo-Rojas, y *et al.* Pathogenicity island *cag*, *vacA* and *is605* genotypes in mexican strains of *Helicobacter pylori* associated with peptic ulcers. *Ann Clin Microbiol Antimicrob*, 10:18, 2011.
- [6] J.C Atherton, P. Cao, R.M. Peek, y *et al.* Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori* association of specific *vacA* types with cytotoxin production and peptic ulceration. *J Biol Chem*, 270(30):17771–17777, 1995.
- [7] S. Backert, T. Schwarz, S. Miehle, y *et al.* Functional analysis of the *cag* pathogenicity island in *Helicobacter pylori* isolates from patients with gastritis, peptic ulcer, and gastric cancer. *Infect Immun*, 72(2):1043–56, 2004.

-
- [8] M. Baele, F. Pasmans, B. Flahou, y *et al.* Non-*Helicobacter pylori* helicobacters detected in the stomach of humans comprise several naturally occurring *Helicobacter* species in animals. *FEMS Immunol Med Microbiol*, 55(3):306–313, 2009.
- [9] Y. Benjamini y T. P. Speed. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic acids res*, 40(10):e72, 2012.
- [10] E.J. Beswick, G. Suarez, y V.E. Reyes. *H. pylori* and host interactions that influence pathogenesis. *World J Gastroenterol*, 12(35):5599–605, 2006.
- [11] M. Camorlinga-Ponce, G. Perez-Perez, G. Gonzalez-Valencia, y *et al.* *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.
- [12] G. Castillo-Rojas, A. Vega-Belmont, Y. López-Vidal, y *et al.* Comparative proteomic analysis of *Helicobacter pylori*-expressed proteins in gastric epithelial cell apoptosis. *Curr Proteomics*, 6(3):187–197, 2009.
- [13] S. Censini, C. Lange, Z. Xiang, y *et al.* *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type i-specific and disease-associated virulence factors. *Proc Natl Acad Sci USA*, 93(25):14648–53, 1996.
- [14] I. Chen, P.J. Christie, y D. Dubnau. The ins and outs of DNA transfer in bacteria. *Science*, 310(5753):1456–60, 2005.
- [15] B. Chevreux, T. Wetter, y S. Suhai. Genome sequence assembly using trace signals and additional sequence information. *JCSB(GCB)*, págs. 45–56, 1999.
- [16] A.E. Darling, B. Mau, y N.T. Perna. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS one*, 5(6):e11147, 2010.
- [17] Polk D.B. y R.M. Peek Jr. *Helicobacter pylori*: gastric cancer and beyond. *Nat Rev Cancer*, 10:403–414, 2010.

- [18] T. de Sablet, M.B. Piazuolo, C.L. Shaffer, y *et al.* Phylogeographic origin of *Helicobacter pylori* is a determinant of gastric cancer risk. *Gut*, 60(9):1189–95, 2011.
- [19] J. Dicksved, M. Lindberg, M. Rosenquist, y *et al.* Molecular characterization of the stomach microbiota in patients with gastric cancer and controls. *J Med Microbiol*, 58:509–516, 2009.
- [20] P Doig, B.L. de Jonge, Richard A Alm, y *et al.* *Helicobacter pylori* physiology predicted from genomic comparison of two strains. *Microbiol Mol Biol Rev*, 63(3):675–707, 1999.
- [21] M.S. Dorer, S. Talarico, y N.R. Salama. *Helicobacter pylori*'s unconventional role in health and disease. *PLoS Pathog*, 5(10):e1000544, 2009.
- [22] D. Falush, T. Wirth, B. Linz, y *et al.* Traces of human migrations in *Helicobacter pylori* populations. *Science*, 299(5612):1582–5, 2003.
- [23] J. Felsenstein. PHYLIP (Phylogeny Inference Package) version 3.69. *Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.*
- [24] I.F. Fernandez. Historia de México. *Pearson Educación*, 2006.
- [25] S. Fischer, B.P. Brunk, F. Chen, y *et al.* Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Curr Protoc Bioinform*, págs. 6–12, 2011.
- [26] W. Fischer, L. Windhager, S. Rohrer, y *et al.* Strain-specific genes of *Helicobacter pylori*: genome evolution driven by a novel type iv secretion system and genomic island transfer. *Nucleic Acids Re*, 38(18):6089–6101, 2010.
- [27] A.C. Ford y A.T.R. Axon. Epidemiology of *Helicobacter pylori* infection and public health implications. *Helicobacter*, 15 Suppl 1:1–6, 2010.
- [28] D.B. Schauer Ge, Z. Genomics of *Helicobacter* species. En Philip M.; Bourke Billy Chan, Ricky V.L; Sherman, ed., *Bacterial Genomes and Infectious Diseases*, págs. 223–233. Springer, 2006.

-
- [29] Z. Ge y D.E. Taylor. Contributions of genome sequencing to understanding the biology of *Helicobacter pylori*. *Annu Rev Microbiol*, 53:353–387, 1999.
- [30] M. Gouy, S. Guindon, y O. Gascuel. Seaview version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*, 27(2):221–224, 2010.
- [31] D.Y. Graham, A.R. Opekun, M.S. Osato, y *et al.* Challenge model for *Helicobacter pylori* infection in human volunteers. *Gut*, 53(9):1235–1243, 2004.
- [32] M. Guardado-Estrada, E. Juarez-Torres, I. Medina-Martinez, y *et al.* A great diversity of Amerindian mitochondrial DNA ancestry is present in the Mexican mestizo population. *J Hum Genet*, 54(12):695–705, 2009.
- [33] L.J. Harmon, J.T. Weir, C.D. Brock, y *et al.* GEIGER: investigating evolutionary radiations. *Bioinformatics*, 24(1):129–131, 2008.
- [34] D. Hofreuter y R. Haas. Characterization of two cryptic *Helicobacter pylori* plasmids: a putative source for horizontal gene transfer and gene shuffling. *J Clin Microbiol*, 184(10):2755–2766, 2002.
- [35] S. Jaenicke, C. Ander, T. Bekel, y *et al.* Comparative and joint analysis of two metagenomic datasets from a biogas fermenter obtained by 454-pyrosequencing. *PLoS One*, 6(1):e14519, 2011.
- [36] J. Joo, J. Song, S. Baik, y *et al.* Genetic organization and conjugal plasmid DNA transfer of pHP69, a plasmid from a korean isolate of *Helicobacter pylori*. *J Microbiol*, 50(6):955–961, 2012.
- [37] F. Kauser, A. A. Khan, M. A. Hussain, y *et al.* The *cag* pathogenicity island of *Helicobacter pylori* is disrupted in the majority of patient isolates from different human populations. *J Clin Microbiol*, 42(11):5302–5308, 2004.
- [38] M. Kawai, Y. Furuta, K. Yahara, y *et al.* Evolution in an oncogenic bacterial species with extreme genome plasticity: *Helicobacter pylori* east asian genomes. *BMC Microbiol*, 11:104, 2011.

- [39] M. Khalifa, R. Sharaf, y R. Aziz. *Helicobacter pylori*: a poor man's gut pathogen? *Gut*, 2(1):2, 2010.
- [40] I.J. Kim y S. R. Blanke. Remodeling the host environment: modulation of the gastric epithelium by the *Helicobacter pylori* vacuolating toxin VacA. *Front Cell Infect Microbiol*, 2:37, 2012.
- [41] S. Kumar, A. Kumar, y V.K. Dixit. Diversity in the *cag* pathogenicity island of *Helicobacter pylori* isolates in populations from north and south india. *J Med Microbiol*, 59(Pt 1):32–40, 2010.
- [42] S. Kurtz, A. Phillippy, A.L. Delcher, y *et al.* Versatile and open software for comparing large genomes. *Genome biology*, 5(2):R12, 2004.
- [43] J.G. Kusters, A.H.M. van Vliet, y F. *et al* Kuipers. Pathogenesis of *Helicobacter pylori* infection. *Clin Microbiol Rev*, 19(3):449–490, 2006.
- [44] E.E. Lara-Ramírez, A. Segura-Cabrera, X. Guo, y *et al.* New implications on genomic adaptation derived from the *Helicobacter pylori* genome comparison. *PLoS ONE*, 6(2):e17300, 2011.
- [45] M.A. Larkin, G. Blackshields, N.P. Brown, y *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21):2947–2948, 2007.
- [46] P. Lehours, F.F. Vale, M.K. Bjursell, y *et al.* Genome sequencing reveals a phage in *Helicobacter pylori*. *MBio*, 2(6), 2011.
- [47] L. Li, C.J. Stoeckert, y D.S. Roos. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome res*, 13(9):2178–2189, 2003.
- [48] B. Linz, F. Balloux, Y. Moodley, y *et al.* An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature*, 445(7130):915–918, 2007.
- [49] Y. López-Vidal, S. Ponce-de León, G. Castillo-Rojas, y *et al.* High diversity of *vacA* and *cagA* *Helicobacter pylori* genotypes in patients with and without gastric cancer. *PLoS One*, 3(12):e3849, 2008.

-
- [50] C. Luo, P. Chiou, C. Yang, y *et al.* Genome, integration, and transduction of a novel temperate phage of *Helicobacter pylori*. *J Virol*, 86(16):8781–8792, 2012.
- [51] M. Mazari-Hiriart, Y. López-Vidal, y J.J. Calva. *Helicobacter pylori* in water systems for human use in Mexico City. *Water Sci Technol*, 43(12):93–8, 2001.
- [52] I. Milne, G. Stephen, M. Bayer, y *et al.* Using Tablet for visual exploration of second-generation sequencing data. *Brief Bioinform*, 14(2):193–202, 2013.
- [53] H.M. Mitchell. The epidemiology of *H. pylori* infection. *J Gastroenterol Hepatol*, 15(12):H23, 2000.
- [54] Y. Moodley, B. Linz, Y. Yamaoka, y *et al.* The peopling of the Pacific from a bacterial perspective. *Science*, 323(5913):527–530, 2009.
- [55] L.T. Nguyen, T. Uchida, K. Murakami, y *et al.* *Helicobacter pylori* virulence and the diversity of gastric cancer in Asia. *J Med Microbiol*, 57(Pt 12):1445–53, 2008.
- [56] R.G. Nitharwal, V Verma, S. Dasgupta, y *et al.* *Helicobacter pylori* chromosomal DNA replication: current status and future perspectives. *FEBS Lett*, 585(1):7–17, 2011.
- [57] H. Ogiwara, D.Y. Graham, y Y. Yamaoka. *vac i*-region subtyping. *Gastroenterology*, 134(4):1267, 2008.
- [58] T.D. Otto, G.P. Dillon, W.S. Degrave, y *et al.* RATT: Rapid annotation transfer tool. *Nucl Acids Res*, 39(9):e57, 2011.
- [59] E. Paradis, J. Claude, y K. Strimmer. APE: analyses of phylogenetics and evolution in r language. *Bioinformatics*, 20(2):289–290, 2004.
- [60] R. Patra, S. Chattopadhyay, R. De, y *et al.* Intact *cag* pathogenicity island of *Helicobacter pylori* without disease association in Kolkata, India. *Int J Med Microbiol*, 301(4):293–302, 2011.

- [61] T.T. Perkins, C.Y. Tay, F. Thirriot, y *et al.* Choosing a benchtop sequencing machine to characterise *Helicobacter pylori* genomes. *PloS one*, 8(6):e67539, 2013.
- [62] A.J. Pinto y L. Raskin. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PLoS One*, 7(8):e43093, 2012.
- [63] H Pongstigl. Smalt (version 0.7.4). *Genome Research Limited*, 2013. URL <http://smalt.sourceforge.net/>.
- [64] M. Quiñones, J.E. Knesek, y S.A. McIntire. Sequence and gene expression analyses of plasmid pHPM8 from *Helicobacter pylori* reveal the presence of two operons with putative roles in plasmid replication and antibiotic activity. *Plasmid*, 46(3):223–8, 2001.
- [65] J.L. Rhead, D.P. Letley, M. Mohammadi, y *et al.* A new *Helicobacter pylori* vacuolating cytotoxin determinant, the intermediate region, is associated with gastric cancer. *Gastroenterology*, 133(3):926–936, 2007.
- [66] S. Rohrer, L. Holsten, E. Weiss, y *et al.* Multiple pathways of plasmid DNA transfer in *Helicobacter pylori*. *PloS one*, 7(9):e45623, 2012.
- [67] C. Romo-González, N.R. Salama, J. Burge-So-Ferreira, y *et al.* Differences in genome content among *Helicobacter pylori* isolates from patients with gastritis, duodenal ulcer, or gastric cancer reveal novel disease-associated genes. *Infect Immun*, 77(5):2201–2211, 2009.
- [68] K. Rutherford, J. Parkhill, J. Crook, y *et al.* Artemis: sequence visualization and annotation. *Bioinformatics*, 16(10):944–945, 2000.
- [69] S. Schwarz, G. Morelli, B. Kusecek, y *et al.* Horizontal versus familial transmission of *Helicobacter pylori*. *PLoS Pathog.*, 4(10):e1000180, 2008.
- [70] A. Sheh, R. Chaturvedi, D.S. Merrell, y *et al.* Phylogeographic origin of *Helicobacter pylori* determines host-adaptive responses upon coculture with gastric epithelial cells. *Infect immun*, 81(7):2468–2477, 2013.

- [71] J.V. Solnick, F. Franceschi, D. Roccarina, y *et al.* Extragastric manifestations of *Helicobacter pylori* infection and other *Helicobacter Species*. *Helicobacter*, 11:46–51, 2006.
- [72] R. Suzuki, S. Shiota, y Y. Yamaoka. Molecular epidemiology, population genetics, and pathogenic role of *Helicobacter pylori*. *Infect Genet Evol*, 12(2):203–13, 2012.
- [73] H. Takeuchi y T. Nakazawa. Chromosomal replication, plasmid replication, and cell division. En Hazell SL Mobley HLT, Mendz GL, ed., *Helicobacter pylori: Physiology and Genetics*, pág. chapter3. Springer, 2001.
- [74] N. Tegtmeyer, S. Wessler, y S. Backert. Role of the *cag*-pathogenicity island encoded type iv secretion system in *Helicobacter pylori* pathogenesis. *FEBS J*, 278(8):1190–1202, 2011.
- [75] *Helicobacter* y Cancer Collaborative Group. Gastric cancer and *Helicobacter pylori*: a combined analysis of 12 case control studies nested within prospective cohorts. *Gut*, 49(3):347–353, 2001.
- [76] J.M. Thiberge, C. Boursaux-Eude, P. Lehours, y *et al.* From array-based hybridization of *Helicobacter pylori* isolates to the complete genome sequence of an isolate associated with MALT lymphoma. *BMC Genomics*, 1:368, 2010.
- [77] J.F. Tomb, O. White, A.R. Kerlavage, y *et al.* The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature*, 388(6642):539–47, 1997.
- [78] J. Torres, L. Lopez, y E et al Lazcano. Trends in *Helicobacter pylori* infection and gastric cancer in Mexico. *Cancer Epidemiol Biomarkers Prev*, 14:1874–1877, 2005.
- [79] J. Uchiyama, H. Takeuchi, S. Kato, y *et al.* Complete genome sequences of two *Helicobacter pylori* bacteriophages isolated from Japanese patients. *J virol*, 86(20):11400–11401, 2012.

-
- [80] F.F. Vale y J.M.B. Vitor. Transmission pathway of *Helicobacter pylori*: Does food play a role in rural and urban areas? *Int J Food Microbiol*, 138(1–2):1–12, 2010.
- [81] X. Wan, D. Tang, A. Liu, y *et al.* Isolation of a wild-type virulent phage of *Helicobacter pylori* and its simulated treatments of gastrointestinal Hp *in vitro*. *Nan fang yi ke da xue xue bao = Journal of Southern Medical University*, 31(2):304, 2011.
- [82] S. Wen y S.F. Moss. *Helicobacter pylori* virulence factors in gastric carcinogenesis. *Cancer Lett*, 282(1):1–8, 2009.
- [83] K. Wilson. Preparation of genomic DNA from bacteria. *Curr Protoc Mol Biol*, 0:2.4.1–2.4.5.
- [84] D.R. Zerbino. Using the Velvet *de novo* assembler for short-read sequencing technologies. *Curr Protoc Bioinform*, 31:11.5.1–11.5.12.
- [85] D.R. Zerbino y E. Birney. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res*, 18(5):821–829, 2008.
- [86] C. Zhang, N. Yamada, Y. Wu, y *et al.* *Helicobacter pylori* infection, glandular atrophy and intestinal metaplasia in superficial gastritis, gastric erosion, erosive gastritis, gastric ulcer and early gastric cancer. *World J Gastroenterol*, 11(6):791–6, 2005.
- [87] Y. Zhou, Y. Liang, K.H. Lynch, y *et al.* PHAST: a fast phage search tool. *Nucl Acids Res*, 39(suppl 2):W347–W352, 2011.