

# Caracterización molecular y bioinformática de la proteína CagA de *Helicobacter pylori* a partir de biopsias gástricas de pacientes colombianos

## Molecular and bioinformatic characterization of *Helicobacter pylori* CagA protein using gastric biopsies of Colombian patients

Paula Nicole Acosta Amador,<sup>1</sup> María del Pilar Delgado,<sup>2</sup> María Camila Montealegre Ortiz,<sup>3</sup> María Magdalena Echeverry de Polanco,<sup>4</sup> Carlos Jaramillo Henao.<sup>5</sup>

<sup>1</sup> Microbióloga, Facultad de Ciencias, Universidad de los Andes. Bogotá, Colombia.

<sup>2</sup> Magister en Microbiología, Facultad de Ciencias, Universidad de los Andes. Bogotá, Colombia.

<sup>3</sup> Magister en Microbiología, Facultad de Ciencias Facultad de Ciencias, Universidad de los Andes. Bogotá, Colombia.

<sup>4</sup> Ph.D en Genética, Facultad de Ciencias Facultad de Ciencias, Universidad de los Andes. Bogotá, Colombia.

<sup>5</sup> Magister en Microbiología Universidad de los Andes. Bogotá, Colombia.

### Instituciones

Laboratorio de Diagnóstico Molecular y Bioinformática, Universidad de los Andes. Carrera 1ª N° 18A-10, Bogotá, DC., Colombia. Universidad privada.

Grupo de Citogenética, Filogenia y Evolución de poblaciones. Universidad del Tolima. Santa Helena, Ibagué, Colombia. Universidad pública.

Difusión del trabajo. Este trabajo fue presentado en el XIX Congreso Latinoamericano y VI Ecuatoriano de Microbiología, llevado a cabo del 15 al 18 de octubre del 2008 en Quito, Ecuador, a manera de presentación oral.

Fecha recibido: 14-07-09

Fecha aceptado: 21-10-09

### Resumen

El gen *cagA* de *Helicobacter pylori* codifica para la proteína CagA considerada uno de los factores de virulencia cuya presencia se asocia a un mayor riesgo de padecer enfermedades gástricas severas. El presente estudio planteó como objetivo el diseño de una estrategia molecular y bioinformática útil en la determinación de la presencia de secuencias repetitivas que pueden contener uno o más motivos de fosforilación (EPIYA). Se amplificó y secuenció la región variable de *cagA* en muestras *H. pylori* CagA positivas. Se realizó una búsqueda y selección de herramientas bioinformáticas que permitieran establecer las características de los motivos EPIYA. La presencia de motivos tipo EPIYA-A y EPIYA-B, seguido por una a dos repeticiones de EPIYA-C, similares a los reportados para países de Occidente, fueron encontrados. De las aplicaciones bioinformáticas evaluadas, solo un conjunto de herramientas demostró ser útil en la caracterización de las unidades de repetición en la proteína CagA.

### Palabras clave

*Helicobacter pylori*, proteína CagA, secuenciación, estrategia bioinformática, motivos de fosforilación.

### Abstract

*Helicobacter pylori* CagA protein, the *cagA* gen product, has been considered as a virulence factor associated with a considerable increase risk for develops severe gastric illness. The purpose of this research was to design a molecular and bioinformatics strategy that allowed the establishment of phosphorylation status of the tyrosine residue of the CagA protein. The amplification and sequencing of the variable fragment region of *cagA* in the positive CagA samples were used to do the bioinformatics analysis in order to establish the characteristics of the EPIYA motifs. The presence of the EPIYA-A and EPIYA-B motifs, followed by one or two EPIYA-C repetitions, similar to those reported previously for occidental countries were set up. From the different bioinformatics applications that were employed only one group of tools proved to be useful to characterize the repeated units presents in the CagA protein.

### Key words

*Helicobacter pylori*, CagA protein, sequencing, bioinformatics strategy, phosphorylation motifs.

## INTRODUCCIÓN

*Helicobacter pylori* es una bacteria microaerófila que ha sido implicada en el desarrollo de desórdenes gastrointestinales como gastritis crónica, úlcera péptica y cáncer gástrico (1). Esta última patología ocupa el cuarto lugar entre las neoplasias de diagnóstico más frecuente, con 934.000 nuevos

casos y 700.000 muertes anuales a nivel mundial (2). *H. pylori* ha sido categorizada como carcinógeno de tipo I por la Agencia Internacional para la Investigación del Cáncer, la cual hace parte de la Organización Mundial de la Salud (3). *H. pylori* infecta alrededor del 50% de la población mundial, con alta incidencia en continentes en vía de desarrollo como Asia, África y Sur América. La tasa de preva-

lencia varía entre poblaciones y entre grupos de una misma población (4).

El genoma de *H. pylori* es aproximadamente de 1,7 Mpb, con un contenido de G-C de 40% (5). La adaptabilidad genética de *H. pylori* le facilita su persistencia en el hospedero debido a que posee mecanismos para incrementar su diversidad y así poder maximizar la utilización de recursos en diversos nichos y micronichos. La generación de diversidad se debe a procesos como mutaciones endógenas y recombinación, razón por la cual, se puede considerar que la mayoría de las cepas de esta especie presentan un fenotipo hipermutador, el cual favorece el surgimiento de nuevas variantes como resultado de una presión de selección (6). La gran diversidad de *H. pylori* implica la selección de varios genotipos que se pudieron adaptar mejor al hombre (7).

La isla de patogenicidad *cag* (*cag*-PAI) de *H. pylori* es una región de 40 Kb que contiene 31 genes, entre los cuales se incluyen componentes importantes del sistema de secreción Tipo IV (T4SS) (8). Debido a la diferencia en su contenido de G-C (35%), respecto al del genoma (40%), se cree que el origen de *cag*-PAI es exógeno y que este fue adquirido de otro microorganismo por un evento de transferencia horizontal (9). Naito y colaboradores, en el 2006 (10), propusieron que en los países de Occidente, las personas *H. pylori cag*-PAI positivo, tenían un mayor riesgo de desarrollar gastritis atrófica, adenocarcinoma y úlcera péptica.

Estudios previos han permitido identificar el gen asociado a la citotoxina (*cagA*), como un marcador de la presencia de *cag*-PAI y un factor de riesgo importante en la presentación de úlcera y cáncer gástrico (11-14). El gen se encuentra ubicado corriente abajo del *cluster* de *cag*-PAI, en el extremo 3' del gen de glutamato racemasa, flanqueado por repeticiones de ADN directas de 39 pb y no existen homólogos del mismo en otras especies de *Helicobacter* (4). Este gen contiene una región 5' altamente conservada y una región 3' que presenta variaciones en el número de secuencias repetitivas, lo que se manifiesta en cambios en el tamaño de la proteína, la cual puede variar entre 120 y 140 kD (15, 16). Las diferentes cepas de *H. pylori* poseen un amplio repertorio genético que permite la variación de fenotipos de CagA en respuesta a hospederos particulares, micronichos o cambios ambientales (17).

CagA induce cambios morfológicos en las células epiteliales tales como elongación y proliferación celular (18). La exposición de células epiteliales a cepas de *H. pylori cag*-PAI positivas puede activar los proto-oncogenes *c-fos* y *c-jun*, siendo un paso crucial para el desarrollo de neoplasias relacionadas con la presencia *H. pylori* (4). Varios reportes indican que la mayoría de cepas *cagA* positivas son capaces de producir CagA y retener todos los genes de *cagA*-PAI, a diferencia de cepas negativas (19).

La proteína CagA se clasifica en dos grandes categorías basadas en los polimorfismos de los sitios de fosforilación de tirosina: CagA del Oriente de Asia y CagA de Occidente (20). Se ha descrito que el estatus de fosforilación de estos residuos de tirosina en CagA es importante para la patogenicidad de *H. pylori* (21). Cuando CagA es translocada al citosol de las células epiteliales ocurre una fosforilación de los motivos de tirosina (TPM) que contienen la secuencia EPIYA, presentes en el extremo C-terminal de la región variable de esta proteína (20, 22-25). Conforme a las secuencias de aminoácidos que se encuentren presentes hacia el C-terminal de las secuencias EPIYA, estas se subclasifican en EPIYA-A, -B, -C, y -D (26). En países de Occidente comúnmente circulan cepas de *H. pylori* que codifican CagA que posee sitios EPIYA-A y EPIYA-B, seguidos por una a tres repeticiones EPIYA-C (ABC, ABCC y ABCCC), de los cuales el más frecuente es el tipo ABC. En países Asiáticos, por el contrario la mayoría de las cepas CagA positivas poseen EPIYA-A, -B y -D (Tipo ABD) (8). La fosforilación de los motivos de tirosina de CagA está mediada así: EPIYA-A y-B vía Src kinasa y EPIYA-C o -D, por la acción de una proteína homóloga a la Src2 fosfatasa (20, 22-25). Yamaoka y colaboradores (27) desarrollaron una metodología para la determinación del número de repeticiones EPIYA y FPLKRHDKVDDLSKV, a partir de la amplificación de dos regiones (FR y WSR) presentes en el extremo 3' del gen *cagA* (27) (figura 1). Algunas cepas clínicas CagA positivas, con un alto número de motivos de fosforilación EPIYA, que han sido aisladas en poblaciones del Oriente de Asia, se han podido asociar con la presencia de gastritis crónica y atrófica. Estudios usando cepas de Occidente han mostrado que un incremento en el número de motivos EPIYA puede estar relacionado con un aumento en la secreción de interleuquina-8 y una pronunciada elongación celular (23). Así mismo se ha encontrado que CagA interactúa con varias rutas importantes de transducción de señales (17).

En Colombia, no hay muchos estudios respecto a la caracterización de los TPM y su posible relación con la patogénesis generada por *H. pylori*. La infección por *H. pylori* en Colombia, es común en adultos y la mayoría de las infecciones son causadas por cepas CagA positivas (28, 29). Asimismo el cáncer gástrico en Colombia representa la primera causa de mortalidad por cáncer ocasionando alrededor de 6.000 muertes anuales, que conllevan 54.700 años de vida saludable perdidos (30). Por lo anterior, la determinación del número y tipo de motivos EPIYA puede ser un marcador importante para la determinación del grado de patogenicidad de *H. pylori* y su posible asociación con el desarrollo de cáncer gástrico y (21). El presente estudio tuvo como objetivo la caracterización de un fragmento del

extremo 3' del gen *cagA* comprendido por las regiones FR y WSR, y el diseño de una metodología molecular y bioinformática que permitiera establecer las características de los motivos EPIYA y FPLKRHDKVDDLKSKV presentes en muestras de *H. pylori* aisladas a partir de biopsias gástricas.

## MATERIALES Y MÉTODOS

### Extracción de ADN

ADN genómico de *H. pylori* fue extraído a partir de cultivo puro para la cepa de referencia NCTC 11637 y directamente de biopsias gástricas, en el caso de las muestras incluidas en el estudio. Las extracciones fueron realizadas utilizando el kit AquaPure Genomic DNA isolation de BIO-RAD siguiendo las instrucciones del fabricante.

### Muestras de ADN

Se evaluaron 10 muestras de ADN genómico de *H. pylori* provenientes de un estudio previo, que habían sido extraídas de biopsias gástricas de pacientes residentes en Ibagué, Tolima, Colombia, con diferentes patologías gástricas e identificadas como *H. pylori* positivas, mediante la amplificación de un fragmento del gen 16S ADNr especie-específico de esta bacteria (31).

### Amplificación de las regiones *cagA*, FR y WSR

La reacción en cadena de la polimerasa (PCR) para la amplificación de las regiones *cagA*, FR y WSR fue realizada por separado para cada región en un volumen final de 50  $\mu$ L que contenía 50 mM de KCL, 20 mM de Tris HCL, pH (8.4), 1,75 mM MgCL<sub>2</sub>, 0,2 mM de cada dNTP, 1pmol/ $\mu$ L de cada iniciador usado en el estudio del gen *cagA* (27) (tabla 1); 1,25 U de Taq ADN polimerasa recombinante (Invitrogen, Carlsbad, California, USA) y 4  $\mu$ L de ADN. En el caso del control positivo se agregó esta cantidad de la cepa de referencia NCTC 11637.

El perfil térmico usado en las PCRs fue el siguiente: denaturación inicial a 92°C por 5 min, seguido por 35 ciclos con una temperatura de denaturación a 92°C por 1 min, anillaje por 1 min a la temperatura indicada en la tabla 1 para cada pareja de iniciadores diseñados por Yamaoka y colaboradores (27) y una elongación a 72°C por 1 min y, finalmente, una extensión a 72°C por 7 min. El tamaño esperado para cada región depende del número de unidades de repetición en aminoácidos EPIYA y FPLKRHDKVDDLKSKV presentes en el extremo 3' del gen *cagA*. Para cada una de las regiones se evaluó la sensibilidad del método utilizando ADN

genómico extraído a partir de una biopsia gástrica y ADN genómico extraído de cultivo bacteriano puro de la cepa de referencia *H. pylori* NCTC 11637. Para el ADN extraído de biopsia gástrica, se utilizó la muestra M49 en las siguientes cantidades en ng: 0,005; 0,01; 0,05; 0,1; 0,5; 1,0; 3,0; 5,0 y 10,0. Para el ADN extraído del cultivo puro se utilizaron las siguientes cantidades de ADN: 0,0039 ng y así, aumentando por un factor de 2 hasta 4 ng.

Los productos de la PCR correspondientes a las regiones *cagA*, FR y WSR fueron separados mediante electroforesis en geles de poliacrilamida al 8%. Las electroforesis se corrieron en refrigeración a 10°C en Buffer TAE 1x por 2 horas a 100V. Para su tinción se utilizó bromuro de etidio (10 mg/mL). Para las pruebas de sensibilidad realizadas y la confirmación de los productos de purificación, se utilizaron geles de agarosa al 1% y las condiciones de corrido fueron 1 hora a 100V. Un transiluminador BIO-RAD®, el software ChemiDoc System XRS y el programa *Quantity one* fueron usados para llevar a cabo el procesamiento y análisis de las imágenes.

Los productos de la PCR fueron purificados utilizando el kit Wizard SV Gel and PCR Clean-Up System (Promega, Madison-Wisconsin, USA), siguiendo las recomendaciones del fabricante, realizando la elusión en un volumen final de 25  $\mu$ L. Los productos fueron cuantificados y se evaluó su pureza (relación de longitud de onda 260/280) mediante espectrofotometría, utilizando un NanoDrop® (Thermo Fisher Scientific, Wilmington).

### Secuenciación de las regiones *cagA*, FR y WSR

Una vez purificados los productos de PCR se procedió a realizar la reacción de secuencia por separado para cada región y para cada iniciador (*forward* y *reverse*) utilizado en las PCRs iniciales (tabla 1). Se usó el kit BigDye® Terminator v3,1 Cycle Sequencing (Applied Biosystems, Foster City-California, USA). La reacción se ejecutó en un volumen final de 10  $\mu$ L el cual contenía: 0,5x de Premix, 0,5x de buffer, 0,16  $\mu$ M de cada iniciador y 2  $\mu$ L de ADN purificado. Las condiciones de amplificación fueron: un ciclo a 96°C por 1 min, seguido por 25 ciclos de 10 seg a 96°C, 5 seg a 50°C y 4 min a 60°C. Para la precipitación de los productos de la reacción de secuencia, se utilizó etanol/EDTA. La obtención de las secuencias se llevó a cabo en un secuenciador ABI PRISM® 310 Genetic Analyzer (Applied Biosystems, Foster City-California, USA). Todas las secuencias de las regiones *cagA*, FR y WSR de la cepa de referencia: *H. pylori* NCTC 11637 fueron comparadas con las secuencias ya reportadas en las bases de datos, por lo tanto se utilizó como control positivo de la secuenciación.

**Tabla 1.** Iniciadores usados para la amplificación de las regiones *cagA*, FR y WSR.

Región	Iniciadores (5'- 3')		Temp. anillaje (°C)
	Forward	Reverse	
<i>cagA</i>	CAGTF: ACCCTAGTCGGTAATGGG	CAGTR: GCTTTAGCTTCTGAYACYGC	61
FR	CAGTF: ACCCTAGTCGGTAATGGG	CAGWR: GCCCTACAMCACCSAAACCAC	59
WSR	CAGWF: AAAAATTGACCRACCTCAATC	CAGTR: GCTTTAGCTTCTGAYACYGC	60

### **Análisis de las repeticiones EPIYA y FPLKRHDKVDDLSKV**

Los cromatogramas obtenidos fueron analizados y depurados usando el programa ChromasPro, versión 1.41. La secuencia consenso se obtuvo a partir de un alineamiento múltiple Clustal W en el programa BioEdit Sequence Alignment Editor versión 7.0.9.0 (32). Para el análisis y caracterización de las unidades de repetición (EPIYA y FPLKRHDKVDDLSKV) presentes en el extremo 3' del gen *cagA* se buscaron las repeticiones tomando como base la traducción de las secuencias obtenidas en los 6 marcos abiertos de lectura, mediante la herramienta Sixpack de EMBOSS (33). Para la búsqueda y cuantificación de las unidades de repetición se usó el programa *Fuzzpro* de EMBOSS (33). Por último, se caracterizaron los motivos de fosforilación de tirosina -TPM, que contienen las secuencias EPIYA tal como lo describió Panayotopoulou y colaboradores (2007).

### **Aspectos éticos**

La presente investigación formó parte de un estudio multidisciplinario e interinstitucional, realizado de manera conjunta entre el Laboratorio de Citogenética, Filogenia y Evolución de Poblaciones, de la Universidad del Tolima y el Laboratorio de Diagnóstico Molecular y Bioinformática de la Universidad de los Andes, proyecto que contó con la aprobación del Comité de Ética de las dos instituciones participantes.

## **RESULTADOS**

### **Sensibilidad de la amplificación de cada región**

Se procedió a determinar la sensibilidad de detección de la PCR para las regiones *cagA*, FR y WSR. La sensibilidad a partir de ADN extraído de cultivo puro de la cepa referencia *H. pylori* NCTC11637 fue 0,0078 ng para todas las regiones evaluadas. La sensibilidad a partir de ADN extraído de

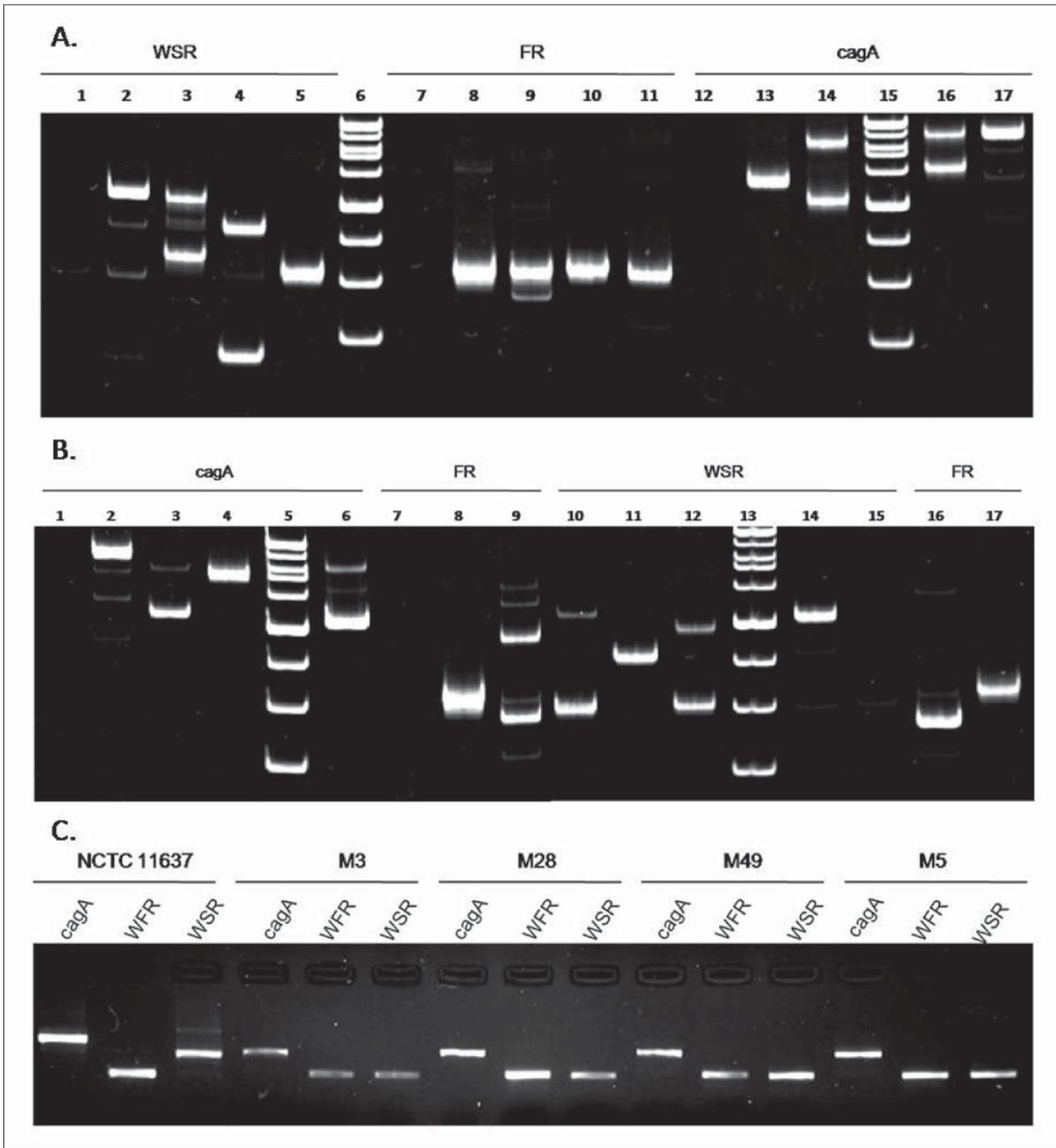
biopsia gástrica para las regiones FR y WSR de la muestra M49 fue 0,05 ng y 0,01 ng respectivamente.

### **Amplificación de las regiones *cagA*, FR y WSR**

Para cada una de las 10 muestras *H. pylori* positivo y para la cepa de referencia *H. pylori* NCTC 11637 [número de acceso GenBank: AF202973.1], se procedió a amplificar las regiones *cagA*, FR y WSR (27) (figura 1). Se determinó el peso molecular de cada una de las regiones analizadas mediante el programa *Quantity one* (ChemiDoc Bio-Rad). El tamaño molecular de los fragmentos resultantes tras la amplificación de la región *cagA* varió entre 484 y 920 pb, para la región FR entre 314 y 342 pb y para la región WSR entre 178 y 541 pb. En 5 de las 10 muestras *CagA* positivo (M004, M010, M012, M23 y M51), al amplificar las regiones *cagA* y WSR se observó la presencia de más de un fragmento de diferente peso molecular en la misma biopsia, lo cual sugiere, en esos casos, una posible coinfección por más de una cepa o bien, este podría ser el resultado de una microevolución de las cepas (figuras 1A y 1B). Se continuó trabajando con las muestras en las cuales solo se evidenció la presencia de un fragmento (M005, M014, M3, M28 y M49), en estas últimas el proceso de purificación fue exitoso (figura 1C) y la concentración de ADN estuvo dentro de los parámetros deseados para secuenciar (concentración mayor a 50 ng/ $\mu$ L) (datos no mostrados).

### **Secuenciación de la región FR**

Se buscaron las repeticiones tanto en las secuencias de nucleótidos (figura 2A) como en la traducción en los seis marcos abiertos de lectura (figura 2B). Para la región FR se determinó la presencia de los TPM EPIYA-A y EPIYA-B a partir de un alineamiento múltiple (figuras 2A y 2B) utilizando el programa BioEdit Sequence Alignment Editor versión 7.0.9.0. Se encontró que el uso de codones que da origen a la repetición EPIYA-B, no fue el mismo respecto al que está presente en la cepa referencia de *H. pylori*, NCTC 11637 (codón TAC, codifica para tirosina (Y) y



**Figura 1.** Visualización de la amplificación de las regiones cagA, FR y WSR, electroforesis en gels de poliacrilamida al 8%. 1A. Carril 1. Control (-), Carril 2. NCTC 11637, Carril 3. M004, Carril 4. M010, Carril 5. M49, Carril 6. MPM, Carril 7. Control (-), Carril 8. NCTC 11637, Carril 9. M004, Carril 10. M010, Carril 11. M49, Carril 12. Control (-), Carril 13. M49, Carril 14. M010, Carril 15. MPM, Carril 16. M004, Carril 17. NCTC 11637. 1B. Carril 1. Control (-), Carril 2. NCTC 11637, Carril 3. M012, Carril 4. M014, Carril 5. MPM, Carril 6. M23, Carril 7. Control (-), Carril 8. NCTC 11637, Carril 9. M012, Carril 10. M023, Carril 11. M014, Carril 12. M012, Carril 13. MPM, Carril 14. NCTC 11637, Carril 15. Control (-), Carril 16. M23 y Carril 17. M014. 1C. Confirmación de los resultados de la purificación de los productos de amplificación para cada región. Los gels fueron teñidos con bromuro de etidio.

codón GCT, codifica para Alanina (A)). En algunas de las muestras evaluadas este uso de codones fue diferente así: el codón TAT codifica para el residuo Y en las muestras M3, M005, M014 y M49; y A está codificado por el codón ACT para las muestras M014 y M49 (figura 2A).

En la región FR se encontraron repeticiones EPIYA del tipo EPIYA-A y -B, caracterizadas por presentar los residuos EPIYAKVNKKK(A/T/V/S)GQ y EPIY(A/T)(Q/K)VAKKVNNAKI respectivamente. Para el tipo EPIYA-A se encontró que los residuos fueron EPIYAKVNKKKTGQ para todas las muestras. Para el tipo EPIYA-B se encontraron los residuos EPIYAQVAKKVNNAKI en la cepa control y en 3 de las 5 muestras evaluadas (11637, M3, M005, M28), mientras que en las otras 2 (M014, M49) los residuos fueron respectivamente, EPIYTQVAKKVTQKI y EPIYTQVAKKVNNAKI (figura 2B).

### Secuenciación de la región WSR

Se amplificó y secuenció la región WSR presente en el extremo 3' del gen *cagA*. Las secuencias evaluadas fueron traducidas en sus 6 marcos abiertos de lectura (ORF). A partir de un alineamiento múltiple de las secuencias de aminoácidos, se buscaron las unidades de repetición EPIYA y FPLKRHDKVDDLSKV, cuyos resultados se muestran en la figura 2C. Así mismo se determinó que el motivo EPIYA se caracteriza por ser del tipo EPIYA-C, encontrándose la secuencia EPIYATIDDLG. Al comparar los alineamientos obtenidos, se observó que las secuencias de aminoácidos provenientes de las diferentes biopsias evaluadas fueron idénticas en cuanto al número de unidades de repetición EPIYA y FPLKRHDKVDDLSKV, menos en la muestra M014, que presentó una repetición de secuencia de 102 bp que codifica para un segmento de 34 aminoácidos que contiene un motivo EPIYA-C más respecto a las otras (tabla 2). El tamaño molecular de la región WSR de esta muestra fue de 411 pb, que es significativamente diferente al de las otras 4 muestras secuenciadas, cuyo promedio fue de 308 pb. Es decir, 4 de las 5 muestras caracterizadas para la repetición EPIYA fueron de tipo ABC y una de tipo ABCC. La cepa *H. pylori* NCTC 11637 se caracteriza por ser de tipo ABCCC. No se encontró ningún motivo EPIYA-D en las biopsias estudiadas.

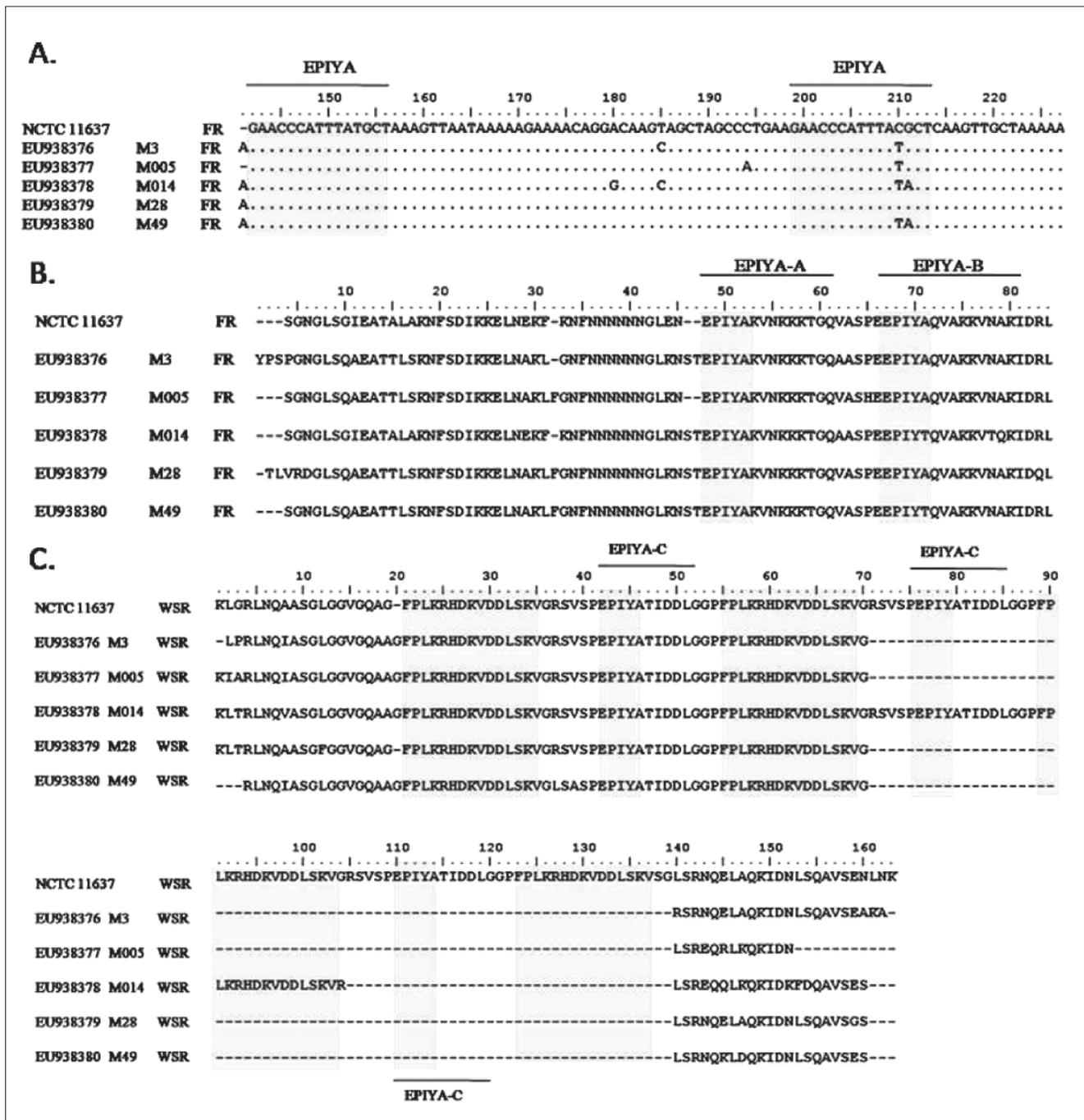
### DISCUSIÓN

Se analizaron 10 muestras de ADN genómico de *H. pylori* proveniente de biopsias gástricas de pacientes de la ciudad de Ibagué, Tolima, Colombia. A estas muestras se amplificó la región *cagA* (compuesta por FR y WSR) descrita anteriormente por Yamaoka (27). La sensibilidad de la PCR a partir del ADN extraído de biopsias fue de 0,05

ng y 0,01 ng para la amplificación de las regiones FR y WSR respectivamente (figura 1B). Estos datos, comparados a los obtenidos a partir de ADN extraído de cultivo puro de la cepa referencia *H. pylori* NCTC 11637, en los cuales, la sensibilidad para todas las regiones evaluadas fue 0,078 ng, es buena, ya que en las biopsias existe una mayor proporción de ADN humano respecto al ADN de *H. pylori*.

El tamaño molecular de los fragmentos obtenidos a partir de las amplificaciones de la región FR para las 10 muestras evaluadas y la cepa de referencia *H. pylori* NCTC11637 fue similar. Este hecho se debe a que en la región FR solo se encuentran nucleótidos que codifican para los motivos EPIYA-A y -B, los cuales son constantes en las muestras *cagA* positivas, por tal razón no hubo variación significativa del tamaño de los amplicones para esta región. En cambio, en 5 de las 10 muestras amplificadas, se evidenció la presencia de bandas de diferente peso molecular en las regiones *cagA* y WSR (figuras 1A y 1B), lo cual concuerda con la idea de que la variación del número de motivos EPIYA en países occidentales, específicamente EPIYA-C, está contenido en el producto de amplificación de la región WSR. Lo anterior, a su vez coincide con que la presencia de más de un fragmento cuando se amplifica las regiones WSR o *cagA*, se puede deber, entre otras circunstancias, a la coinfección por diferentes cepas de *H. pylori* con polimorfismos en la región variable del gen *cagA* en un mismo individuo o microevolución de un misma cepa durante un periodo prolongado de infección. La no presencia de doble banda en las amplificaciones correspondientes a la región FR, se debe a que en FR es constante en el número de repeticiones de EPIYA, y así haya coinfección o microevolución, solo se observará una banda del mismo tamaño, aun ante la presencia de dos fragmentos de ADN pertenecientes en la misma biopsia.

Inicialmente se procedió a calcular el número de repeticiones a partir del peso molecular del amplicón para las regiones *cagA*, FR y WSR. Para este fin se tuvo en cuenta el estudio de Yamaoka (27), donde a partir del trabajo con diferentes cepas provenientes de aislamientos de *H. pylori*, propusieron las fórmulas  $[(218 - 227) + 57 * r]$  y  $[(174 - 177) + 102 * m]$  (r: número de regiones FR, m: número de regiones WSR), para determinar el número de unidades de repetición EPIYA y FPLKRHDKVDDLSKV evaluadas en este estudio. Se hizo una comparación entre los datos arrojados por la secuenciación realizada y por el método propuesto por Yamaoka (27) (datos no mostrados). Los datos arrojados por ambas metodologías divergieron considerablemente respecto al número de repeticiones EPIYA y FPLKRHDKVDDLSKV, consolidándose la secuenciación y análisis bioinformático como estrategia más eficiente para la caracterización de patrones de repetición, que se encuentran dispersos dentro de un genoma.



**Figura 2.** Alineamiento de 5 secuencias de cepas de *H. pylori* provenientes de Tolima, Colombia y la cepa de referencia *H. pylori* NCTC 11637. A. Secuencias de nucleótidos de la región FR. B. Secuencias de aminoácidos de la región FR. C. Secuencias de aminoácidos de la región WSR. (.) similitud entre residuos. (-) gap. Los recuadros grises corresponden a los sitios EPIYA y la secuencia de aminoácidos FPLKRHDKVDL SKV. El alineamiento múltiple se realizó con el programa Clustal W en BioEdit Sequence Alignment Editor versión 7.0.9.0.

Se pudo determinar que las secuencias obtenidas para las 5 muestras que presentaron una sola banda (M3, M005, M014, M28, M49) y la cepa *H. pylori* NCTC 11637 (usada en este caso, como referencia) fueron óptimas, ya que al

comparar los alineamientos entre la región FR y WSR para cada muestra, con la región *cagA* amplificada para la misma muestra, se encontró total coincidencia. Adicionalmente, las secuencias obtenidas de las regiones *cagA*, FR y WSR

para la cepa de *H. pylori* NCTC 11637, mostraron total coincidencia y por lo tanto una excelente reproducibilidad en relación con las secuencias reportadas previamente para esta cepa de referencia.

Se tradujeron las secuencias obtenidas en los 6 marcos abiertos de lectura (ORF) a partir del uso del programa Sixpack de EMBOSS para así poder obtener la secuencia de aminoácidos correspondiente para cada muestra, teniendo en cuenta el código genético (EGC) 11, correspondiente al de bacterias, tal como lo describe la clasificación propuesta por NCBI (34). Se determinó el número de unidades de repetición EPIYA y FPLKRHDKVDDLSKV mediante el uso del programa *Fuzzpro* de EMBOSS (33). En la región FR, para todas las muestras secuenciadas, se encontró la presencia de dos repeticiones EPIYA separadas entre sí por 14 residuos de aminoácidos (figura 2B). Se presentaron variaciones en los codones que codifican para los residuos Tirosina y Alanina. La caracterización de los motivos EPIYA está determinada por los residuos que se encuentran hacia el C-terminal de la secuencia de aminoácidos EPIY(A/T) (20). Al realizar la caracterización de los sitios EPIYA se encontró que tanto el primero como el segundo motivo EPIYA (designado s EPIYA-A y EPIYA-B, respectivamente) estuvieron presentes en todas las secuencias evaluadas (tabla 2). En este estudio, se observaron en la región FR las secuencias de aminoácidos EPIYAKVNKKKTGQ y EPIYAQVAKKVNNAKI correspondientes al tipo EPIYA-A y EPIYA-B para todas las muestras evaluadas, excepto las muestras M014 y M49 en donde se observaron en el sitio EPIYA-B las secuencias EPIYTQVAKKVTQKI / EPIYTQVAKKVNNAKI respectivamente, reportándose una nueva variante para la caracterización del sitio EPIYA-B en la muestra M014 respecto a lo reportado por Panayotopoulou y colaboradores en el 2007.

En la región WSR, se encontró una mayor variación en cuanto al número de unidades de repetición EPIYA presentes en el extremo 3' del gen *cagA*, respecto a la encontrada en la región FR. Según Yamaoka (27), la región WSR está constituida por una repetición FPLKRHDKVDDLSKV seguida por 6 aminoácidos, un motivo EPIYA, inmediatamente una sucesión de 8 aminoácidos y por último nuevamente la repetición FPLKRHDKVDDLSKV (figura 2C). Este patrón se encontró en las secuencias de aminoácidos obtenidos de la traducción en todas las muestras evaluadas. El motivo EPIYA, en la región WSR es EPIYA-C (EPIYATIDDLG). El análisis bioinformático realizado respecto a la cepa de referencia *H. pylori* NCTC 11637 permitió demostrar la validez de la metodología realizada, debido a que este control positivo a nivel de las unidades de repetición EPIYA y FPLKRHDKVDDLSKV concordó con lo descrito en la literatura (ABCCC) (20). También se corroboró la naturaleza genética referente a los polimorfismos presentes en la región variable de este gen para la cepa NCTC 11637 (ABCCC), tal como lo reportaron Higashi, et al (26). La cantidad de unidades de repetición EPIYA-C varió entre 1 y 3. El número de repeticiones para el motivo EPIYA-C varía de 0 a 3, por lo tanto los datos obtenidos concuerdan con lo anteriormente reportado en cepas aisladas en Europa, América y Australia (26). Cuatro de las muestras evaluadas fueron de tipo ABC, y esto se asemeja a las estadísticas en países occidentales, donde comúnmente circulan cepas de *H. pylori* CagA positivo, que poseen sitios EPIYA tipo ABC (35). Solo la muestra M014 tuvo un mayor número de unidades de repetición EPIYA-C, con relación a las demás muestras (tabla 2), es decir su genotipo fue ABCC. Es coherente no haber encontrado en las muestras evaluadas el sitio EPIYA-D, ya que este pertenece

**Tabla 2.** Determinación del número de unidades de repetición EPIYA y FPLKRHDKVDDLSKV presentes en muestras de biopsias gástricas procedentes de pacientes residentes en Ibagué, Colombia.

Muestra	Patología	CagA			
		Región FR		Región WSR	
		EPIYA-A	EPIYA-B	EPIYA-C	FPLKRHDKVDDLSKV
M005	GN	1	1	1	2
M014	GA	1	1	2	3
M3	GN	1	1	1	2
M28	GN	1	1	1	2
M49	GN	1	1	1	2
NCTC11637	NA	1	1	3	4

(GN) Gastritis no atrófica, (GA) Gastritis atrófica, (NA) No aplica.

a cepas endémicas de *H. pylori* en países del este de Asia (9, 26).

La variación en la región C-terminal de la proteína CagA ha sido relacionada con la actividad biológica de esta proteína, es decir, la patogenicidad de *H. pylori* *cagA* positiva se ve influenciada con la duplicación del motivo EPIYA-C, ya que es un determinante crítico de CagA para perturbar la señalización celular como factor de virulencia (26) y generar patologías importantes como el cáncer gástrico (36). El número de sitios EPIYA-C está directamente correlacionado con el nivel de fosforilación de SHP-2. El complejo CagA-SHP-2 se detecta comúnmente en la mucosa atrófica, lo cual podría estar relacionado al desarrollo de gastritis atrófica y la transición de atrofia a metaplasia intestinal (37). También es importante el análisis de los sitios EPIYA-A y EPIYA-B, debido a que algunos estudios indican que estos motivos son importantes para la función de la proteína CagA, ya que son responsables de la asociación de CagA a la membrana de las células epiteliales (26).

Los resultados del presente estudio hacen evidente el hecho de que usando ADN obtenido en forma directa, extraído a partir de biopsias gástricas, es posible llevar a cabo una caracterización molecular y bioinformática que permite determinar con precisión el tipo y número de repeticiones EPIYA del extremo 3' del gen *cagA*. Ya que se ha sugerido que el número y tipo de motivos EPIYA podrían ser un marcador importante para la determinación del grado de patogenicidad de *H. pylori*, los estudios aplicados en países latinoamericanos como Colombia, en los que el número de cepas de *H. pylori* CagA positivo reportadas en las bases de datos de GenBank es relativamente alto (38), el establecimiento de este tipo de estrategias para el análisis de las secuencias de nucleótidos y de las secuencias aminoácidas de las proteínas codificadas por genes relacionados con la virulencia podrían conducir a establecer posibles marcadores de susceptibilidad al desarrollo de patologías severas como cáncer gástrico.

## Agradecimientos

A los participantes del proyecto: Uso de técnicas moleculares para la genotipificación de *Helicobacter pylori* y para la detección de polimorfismos del gen de la Interleucina 1 $\beta$  humana y su asociación con la presentación de patologías gastroduodenales en pacientes colombianos. A los miembros del Laboratorio de Diagnóstico Molecular y Bioinformática. Al Instituto Nacional de Cancerología por haber aportado la cepa de referencia que utilizamos en este estudio. A la Facultad de Ciencias de la Universidad de los Andes, Bogotá, Colombia.

## Conflicto de intereses y financiación

Los autores manifiestan que no existe ningún conflicto de interés.

## REFERENCIAS

1. Blaser M, Parsonnet J. Parasitism by the "slow" bacterium *Helicobacter pylori* leads to altered gastric homeostasis and neoplasia. *J Clin Invest* 1994; 94(1): 4.
2. Parkin DM, Bray FI, Devesa SS. Cancer burden in the year 2000. The global picture. *Eur J Cancer* 2001; 37: 4-66.
3. IARC. IARC monographs of the evaluation of carcinogenic risks to humans. IARC 1994; 61: 177-241.
4. Kauser F, Khan AA, Hussain MA, Carroll IM, Ahmad N, Tiwari S, et al. The *cag* Pathogenicity Island of *Helicobacter pylori* is Disrupted in the Majority of Patient Isolates from Different Human Populations. *J Clin Microbiol* 2004; 42(11): 5302-5.
5. Zamboni CF, Navaglia F, Basso D, Rugge M, Plebani M. *Helicobacter pylori* *babA2*, *cagA*, and *s1 vacA* genes work synergistically in causing intestinal metaplasia. *Br Med J* 2003; 56(4): 287.
6. Blaser MJ, Atherton JC. *Helicobacter pylori* persistence: biology and disease. *The J Clin Invest* 2004; 113(3): 321-33.
7. Kersulyte D, Mukhopadhyay AK, Velapatino B, Su WW, Pan ZJ, Garcia C, et al. Differences in Genotypes of *Helicobacter pylori* from Different Human Populations. *J Bacteriol* 2000; 182(11): 3210.
8. Handa O, Naito Y, Yoshikawa T. CagA protein of *Helicobacter pylori*: A hijacker of gastric epithelial cell signaling. *Biochem Pharmacol* 2007; 73(11): 1697-702.
9. Azuma T, Yamakawa A, Yamazaki S, Ohtani M, Ito Y, Muramatsu A, et al. Distinct Diversity of the *cag* Pathogenicity Island among *Helicobacter pylori* Strains in Japan. *J Clin Microbiol* 2004; 42(6): 2508-17.
10. Naito M, Yamazaki T, Tsutsumi R, Higashi H, Onoe K, Yamazaki S, et al. Influence of EPIYA-Repeat Polymorphism on the Phosphorylation-Dependent Biological Activity of *Helicobacter pylori* CagA. *Gastroenterology* 2006; 130(4): 1181-90.
11. Blaser MJ. Infection with *Helicobacter pylori* strains possessing *cagA* is associated with an increased risk of developing adenocarcinoma of the stomach. *Cancer Res* 1995; 55(10): 2111-5.
12. Alm RA, Ling LS, Moir DT, King BL, Brown ED, Doig PC, et al. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen. *Helicobacter pylori* 1999. p. 176-80.
13. Kuipers EJ, Perez-Perez GI, Meuwissen SGM, Blaser MJ. *Helicobacter pylori* and atrophic gastritis: importance of the *cagA* status. *J Natl Cancer Inst* 1995; 87(23): 1777-80.
14. Peek Jr RM. *Helicobacter pylori* *cagA*<sup>+</sup> strains and dissociation of gastric epithelial cell proliferation from apoptosis. *J Natl Cancer I* 1997; 89(12): 863-8.

15. Evans Jr DJ, Queiroz DM, Mendes EN, Evans DG. Diversity in the variable region of *Helicobacter pylori* cagA gene involves more than simple repetition of a 102-nucleotide sequence. *Biochem Biophys Res Commun* 1998; 245(3): 780-4.
16. Azuma T, Yamakawa A, Yamazaki S, Fukuta K, Ohtani M, Ito Y, et al. Correlation between variation of the 3' region of the cagA gene in *Helicobacter pylori* and disease outcome in Japan. *J Infect Dis* 2002; 186(11): 1621-30.
17. Blaser MJ, Atherton JC. *Helicobacter pylori* persistence: biology and disease. *Am Soc Clin Investig* 2004; 321-33.
18. van Doorn LJ, Figueiredo C, Rossau R, Jannes G, van Asbroeck M, Sousa JC, et al. Typing of *Helicobacter pylori* vacA Gene and Detection of cagA Gene by PCR and Reverse Hybridization. *J Clin Microbiol* 1998; 36(5): 1271.
19. Maeda S, Yoshida H, Ikenoue T, Ogura K, Kanai F, Kato N, et al. Structure of cag pathogenicity island in Japanese *Helicobacter pylori* isolates. *Br Med J* 1999; 44(3): 336.
20. Argent RH, Zhang Y, Atherton JC. Simple method for determination of the number of *Helicobacter pylori* CagA variable-region EPIYA tyrosine phosphorylation motifs by PCR. *J Clin Microbiol* 2005; 43: 791-5.
21. Mimuro H, Suzuki T, Tanaka J, Asahi M, Haas R, Sasakawa C. Grb2 Is a Key Mediator of *Helicobacter pylori* CagA Protein Activities. *Mol Cell* 2002; 10(4): 745.
22. Odenbreit S, Püls J, Sedlmaier B, Gerland E, Fischer W, Haas R. Translocation of *Helicobacter pylori* CagA into Gastric Epithelial Cells by Type IV Secretion. *Science* 2000; 287(5457): 1497.
23. Panayotopoulou EG, Sgouras DN, Papadakos K, Kalliaropoulos A, Papatheodoridis G, Mentis AF, et al. Strategy To Characterize the Number and Type of Repeating EPIYA Phosphorylation Motifs in the Carboxyl Terminus of CagA Protein in *Helicobacter pylori* Clinical Isolates. *J Clin Microbiol* 2007; 45(2): 488-95.
24. Asahi M, Azuma T, Ito S, Ito Y, Suto H, Nagai Y, et al. *Helicobacter pylori* CagA Protein Can Be Tyrosine Phosphorylated in Gastric Epithelial Cells. *J Exp Med* 2000; 191(4): 593-602.
25. Stein M, Rappuoli R, Covacci A. Tyrosine phosphorylation of the *Helicobacter pylori* CagA antigen after cag-driven host cell translocation. *Proc Natl Acad Sci USA* 2000; 97(3): 1263.
26. Higashi H, Tsutsumi R, Fujita A, Yamazaki S, Asaka M, Azuma T, et al. Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. *Proc Natl Acad Sci USA* 2002; 99(22): 14428.
27. Yamaoka Y, El-Zimaity HMT, Gutiérrez O, Figura N, Kim JK, Kodama T, et al. Relationship between the cagA 3' repeat region of *Helicobacter pylori*, gastric histology, and susceptibility to low pH. *Gastroenterology* 1999; 117(2): 342-9.
28. Breuer T, Malaty HM, Graham DY. The epidemiology of *H. pylori*-associated gastroduodenal diseases. Lippincott-Raven 1997. p. 1-14.
29. Yamaoka Y, Kodama T, Gutiérrez O, Kim JG, Kashima K, Graham DY. Relationship between *Helicobacter pylori* iceA, cagA, and vacA Status and Clinical Outcome: Studies in Four Different Countries. *J Clin Microbiol* 1999; 37(7): 2274.
30. Moros M, Jurado C, Mora H, Wilches G, Escobar R, González G, et al. Estrategia de intervención al cáncer gástrico en el Norte de Santander. *Rev Colomb Gastroenterol* 2004; 19: 9-12.
31. Montealegre M, Jaramillo C, Bohorquez M. *Helicobacter pylori* detection and human Interleukin 1-β genotyping in Colombian patients affected by gastroduodenal diseases. *Rev Col Gastroenterol* 2008; 23(1): 40-4.
32. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 1999; 41: 95-8.
33. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet* 2000; 16(6): 276-7.
34. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Church DM, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2005; 33.
35. Covacci A, Censini S, Bugnoli M, Petracca R, Burroni D, Macchia G, et al. Molecular characterization of the 128-kDa immunodominant antigen of *Helicobacter pylori* associated with cytotoxicity and duodenal ulcer. *Proc Natl Acad Sci USA* 1993; 90(12): 5791-5.
36. Tsutsumi R, Higashi H, Higuchi M, Okada M, Hatakeyama M. Attenuation of *Helicobacter pylori* CagA SHP-2 Signaling by Interaction between CagA and C-terminal Src Kinase. *J Biol Chem* 2003; 278(6): 3664-70.
37. Yamazaki S, Yamakawa A, Ito Y, Ohtani M, Higashi H, Hatakeyama M, et al. The CagA Protein of *Helicobacter pylori* is translocated into Epithelial Cells and Binds to SHP-2 in Human Gastric Mucosa. *J Infect Dis* 2003; 187(2): 334-7.
38. Azuma T, Yamazaki S, Yamakawa A, Ohtani M, Muramatsu A, Suto H, et al. Association between diversity in the Src homology 2 domain-containing tyrosine phosphatase binding site of *Helicobacter pylori* CagA protein and gastric atrophy and cancer. *J Infect Dis* 2004; 189(5): 820-7.