

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

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

¹ Microbiologist, Science Department, Universidad de los Andes (Los Andes University). Bogotá, Colombia.

² Master in Microbiology. Science Department, Los Andes University Bogotá, Colombia.

³ Master in Microbiology. Science Department, Los Andes University, Bogotá, Colombia.

⁴ PhD in Genetics, Science Department, Los Andes University, Bogotá, Colombia.

⁵ Master in Microbiology, Los Andes University, Bogotá, Colombia.

Institutions

Laboratorio de Diagnóstico Molecular y Bioinformática (Molecular diagnosis Laboratory and Bioinformatics Laboratory), Los Andes University, Carrera 1ª N° 18A-10, Bogotá, DC., Colombia. Private University.

Grupo de Citogenética, Filogenia y Evolución de Poblaciones (Cytogenetics, Phylogeny and Evolution of Populations group). Universidad del Tolima (Tolima University). Santa Helena Public University, Ibagué, Colombia.

Received: 14-07-09

Accepted: 21-10-09

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.

Work diffusion. This work was presented as an oral presentation in the XIX Congreso Latinoamericano y VI Ecuatoriano de Microbiología (19th Latin-American and 6th Ecuatorian Congress of Microbiology) between October 15 and 18 of 2008 in Quito, Ecuador.

INTRODUCTION

Helicobacter pylori is a microaerophilic bacterium which is involved in the development of several gastrointestinal disorders including chronic gastritis, peptic ulcer disease and gastric cancer (1). Among the most frequently diagnosed neoplasias, gastric cancer ranks 4th, with 934,000 new cases and 700.000 annual deaths worldwide (2). The International Agency for Research on Cancer, a World Health Organization agency, has categorized *H. pylori* as a type I carcinogenic agent (3). *H. pylori* infects around 50% of the world's population and has a high impact on the developing continents of Asia, Africa, and South America. The prevalence rate varies among populations and groups within the same population (4).

The *H. pylori* genome has an average size of 1.7 Mpb, with 40% guanine-cytosine content (GC-content) (5). *H. pylori*'s genetic adaptability facilitates persistence in the host because it has mechanisms which increase its diversity allowing it to maximize use of resources in diverse niches and microniches. Its generation of diversity is due to processes such as endogenous mutations and recombination. Consequently, most strains have a hypermutator phenotype favoring development of new variants in response to selection pressure (6). The great diversity of *H. pylori* implies a selection of several genotypes that can adapt better to humans (7).

The *cag* Pathogenicity Island (*cag*-PAI) of *H. pylori* is a 40Kb region containing 31 genes. They include important components of the type IV secretion system (T4SS) (8).

Because its G-C content (35%) is different than that of the complete genome (40%), the origin of the *cag*-PAI is probably exogenous. It was probably acquired from another microorganism through horizontal gene transfer events (9). In 2006, Naito and colleagues (10) proposed that people in Western countries who are infected with *cag* PAI-positive *H. pylori* strains have a higher risk of developing atrophic gastritis, gastric adenocarcinoma and peptic ulcers.

Previous studies have identified the gene associated with *cagA* cytotoxin as a marker for the presence of *cag*-PAI presence and as an important risk factor for gastric ulcers and gastric cancer (11-14). The gene is located downstream from the *cag*-PAI cluster in the 3' end of the glutamate racemase gene. It is flanked by direct repeats of 39 base pairs (bp) of DNA. There are no homologues for other *Helicobacter* species (4). This gene contains a highly conserved 5' region and a 3' region that presents variations in the repetitive sequence number. This feature represents protein size changes of between 120 and 140 kD (15, 16). Different *H. pylori* strains have wide genetic repertoires that allow phenotypical variation of CagA protein in response to particular hosts, micro-niches and environmental changes (17).

CagA induces morphological changes in epithelial cells. These include elongation and cell proliferation (18). The exposure of epithelial cells to *cag*-PAI positive *H. pylori* strains can activate *c-fos* and *c-jun* proto-oncogenes. This is a crucial step in the development of neoplasia related to *H. pylori* strains (4). Several reports indicate that most *cagA* positive strains, unlike negative strains, are able to produce CagA protein and retain all *cagA*-PAI genes (19).

CagA protein is classified into two large categories based on polymorphisms of tyrosine phosphorylation sites: East Asian CagA protein, and Western CagA protein (20). The phosphorylation status of the tyrosine residue of the CagA protein is thought to be important for *H. pylori* pathogenicity (21). When CagA protein is translocated to the cytosol of epithelial cells in the C-terminal variable region of the protein, tyrosine phosphorylation motifs (TPM) containing EPIYA sequences occur (20, 22-25). These are sub classified as EPIYA-A, B, -C and -D according to the amino acid sequences located in the C-terminal of EPIYA (26). In Western countries, commonly circulating *H. pylori* strains that code for CagA protein have EPIYA-A and EPIYA-B sites, followed by one or three EPIYA-C repeats (ABC, ABCC y ABCCC). ABC is the most common repeat. In contrast, in Asian countries most of the CagA positive strains have EPIYA-A,-B and -D sites (ABD type) (8). Mediation of phosphorylation of EPIYA-A and -B occurs through Src kinase activity, while mediation of EPIYA-C or -D occurs through the action of a homologue Src2 phosphatase protein (22,22-25). Yamaoka and associates (27) have developed a method to determine the number of EPIYA and FPLKRHDKVDDLSKV repeats

from the amplification of two regions present in the 3' end of *cagA* gene (27, Figure 1). Some CagA positive clinical strains with high numbers of EPIYA phosphorylation motifs, isolated from East Asian populations, have been associated with chronic and atrophic gastritis. Some studies using Western strains have demonstrated increases in EPIYA number motifs. This can be related to an increase of interleukin-8 secretion and pronounced cell elongation (23). Similarly, it has been found that CagA protein interacts with several important signal transduction routes (17).

In Colombia, few studies characterizing TPM and their possible relationships with *H. pylori* pathogenicity have been undertaken. *H. pylori* infections are common in adults in Colombia. Most are caused by CagA positive strains (28, 29). In addition, gastric cancer causes 6,000 deaths each year in Colombia, more than any other single type of cancer. This represents the equivalent of 54,700 years of healthy lives lost (30). Because of this, the number and type of EPIYA motifs could become important markers for the determination of the degree of *H. pylori* pathogenicity and its possible associations with the development of gastric cancer (21). This study has two objectives. The first is to characterize one fragment in the 3' region of the *cagA* gene which comprises the FR and WSR regions. In addition, this study is aimed at the design of a molecular and bioinformatics methodology to establish the characteristics of EPIYA and FPLKRHDKVDDLSKV motifs in *H. pylori* samples which have been isolated from gastric biopsies.

MATERIALS AND METHODS

DNA Extraction

H. pylori genomic DNA was extracted from a pure culture of reference strain NCTC 11637 and directly from gastric biopsy samples included in the study. The extractions were performed with the AquaPure Genomic isolation DNA BIO-RAD kit, following the manufacturer's instructions.

DNA samples

10 samples of *H. pylori* genomic DNA were evaluated. The samples were originally taken for a previous study from gastric biopsies of patients from Ibagué, Tolima Department, Colombia. Samples had different gastric pathologies which had been identified as *H. pylori* positive through amplification of species specific 16S rDNA fragments (31).

Amplification of *cagA*, FR and WSR regions

Separate polymerase chain reactions (PCR) were performed to amplify *cagA* FR and *cagA* WSR regions. The 50 µL

final volume of each reaction contained 50 mM of KCL; 20 mM of Tris HCL, pH (8.4); 1.75 mM MgCL₂; 0.2 mM of each dNTP; 1pmol/ μL of each primer used in the *cagA* gene study (27) (Table 1); 1.25 U of Taq DNA recombinant Polymerase (Invitrogen, Carlsbad, California, USA) and 4 μL of DNA. Reference strain NCTC 11637 was used as a positive control.

The thermal profile used in the PCR's were as follows: 92°C initial denaturation step for 5 minutes; followed by 35 one minute denaturation cycles at 92°C; followed by a one minute annealing step (designed by Yamaoka and associates) at the temperatures listed in Table 1 for each pair of primers (27); followed by a one minute elongation step at 72°C; and finally followed by a seven minute extension step at 72°C. The size expected for each region depends on the number of repeat units in EPIYA and FPLKRHDKVDDLSKV amino acids present in the 3' end region of the *cagA* gene. For each region, the sensitivity of the method was evaluated using genomic DNA extracted from a gastric biopsy and genomic DNA extracted from a pure bacterial culture of *H. pylori* reference strain NCTC 11637. For the DNA extracted from a gastric biopsy, an M49 sample was used in the following quantities: 0.005 ng, 0.01 ng, 0.05 ng, 0.1 ng, 0.5 ng, 1.0 ng, 3.0 ng, 5.0ng and 10.0ng. For the DNA extracted from the pure culture a base quantity of 0.0039 ng was used, followed by quantities increasing by a factor of 2 to a maximum of 4.0ng.

The PCR products corresponding to *cagA* FR and WSR regions were separated through electrophoresis using an 8% polyacrylamide gel. Electrophoresis was run under refrigeration at 10°C in 1x TAE Buffer for 2 hours at 100V. 10 mg/Ml of ethidium bromide stain was used. 1% agarose gels were used for 1 hour at 100V for sensitivity tests and product purification confirmation. BIO-RAD® transilluminator, ChemiDoc XRS System software and *Quantity one* software were used for processing and image analysis.

PCR products were purified using Wizard SV Gel Kit and PCR Clean-Up System (Promega, Madison-Wisconsin, USA), following manufacturer's instructions. A 25 μL final volume elution was performed. Products were quan-

tified and their purities were evaluated by spectrophotometry, using NanoDrop® (Thermo Fisher Scientific, Wilmington).

Sequencing of *cagA*, FR and WSR regions

Once the PCR products were purified a separate chain reaction was performed for each region and for each primer (forward and reverse) used in the first PCR's (Table 1). The BigDye® Terminator v3.1 Cycle Sequencing kit was used (Applied Biosystems. Foster City, California USA). The final 10μL volume of the reaction contained 0.5x of Premix, 0.5x of buffer, 0.16 μM of each primer and 2 μL of purified DNA. Amplification conditions were as follows: an initial one minute cycle at 96°C, followed by 25 ten second cycles seconds at 96°C, followed by five seconds at 50°C and four minutes at 60°C. Ethanol/EDTA was used for precipitation of the chain reaction products. Sequences were obtained with ABI PRISM® 310 Genetic Analyzer sequencer (Applied Biosystems. Foster City, California USA). All *cag*, FR and WSR region sequences of the *H. pylori* NCTC 11637 reference strain were compared with reported sequences in databases so that it could be used as a positive control for sequencing.

Analysis of EPIYA y FPLKRHDKVDDLSKV regions

The chromatograms obtained were analyzed and refined using ChromaspRO version 1.41. The consensus sequence was obtained from Clustal W multiple alignment performed in BioEdit Sequence Alignment Editor version 7.0.9.0 (32). Repetitions found in translations of sequences obtained from 6 open reading frames were taken as the bases for analysis of EPIYA and FPLKRHDKVDDLSKV repeat units present in the 3' end region of *cagA* genes. EMBOSS *Sixpack* was used for this process (33) while EMBOSS *Fuzzpro* was used for searching and quantification. Finally, the tyrosine phosphorylation motifs which contain the EPIYA sequences were characterized as described by Panayotopoulou and colleagues (2007).

Table 1. Primers for amplification of *cagA* FR and WST regions.

Region	Primers (5'- 3')		Anneal-ing tempera-ture. (°C)
	Forward	Reverse	
<i>cagA</i>	CAGTF: ACCCTAGTCGGTAATGGG	CAGTR: GCTTTAGCTTCTGAYACYGC	61
FR	CAGTF: ACCCTAGTCGGTAATGGG	CAGWR: GCCCTACAMCACCSAAACCAC	59
WSR	CAGWF: AAAAATTGACCRACCTCAATC	CAGTR: GCTTTAGCTTCTGAYACYGC	60

Ethical aspects

This investigation was part of a multidisciplinary and inter-institutional study, conducted by the Cytogenetics, Phylogeny and of Population Evolution laboratory of Tolima University and the Molecular Diagnosis and Bioinformatics Laboratories of Los Andes University. This Project was approved by the Ethics Committees of both universities.

RESULTS

Sensitivity of amplification of each region

Determination of the sensitivity of PCR detection for *cagA* FR and WSR regions showed that the sensitivity of the DNA extracted from pure culture of the *H. pylori* NCTC 11637 reference strain was 0.0078 ng.

The sensitivity of the DNA extracted from gastric biopsies for the FR region was 0.05 ng, while it was 0.01 ng for the WSR region extracted from the M49 sample.

Amplification of *cagA* FR and WSR regions

CagA FR and WSR regions were amplified for the 10 *H. pylori* positive samples and for the NCTC 11637 reference strain [Genebank access number: AF202973.1] (27, Figure 1). For each region, the molecular weight was determined using *Quantity one* (ChemiDoc Bio-Rad). The molecular size for the resulting fragments for *cagA* regions varied between 484 and 920 bp, while it varied between 314 and 342 bp for FR regions and between 178 and 541 bp. for WSR regions. More than one fragment in *cagA* and WSR regions, each with different molecular weights, was observed in each of 5 of the 10 *CagA* positive samples (M004, M010, M012, M23 y M51). This result could indicate co-infection for more than one strain, or it could indicate microevolution of strains (Figure 1A and Figure 1B). The purification process was successfully conducted on the samples in which only one fragment was found (M005, M014, M3, M28 y M49). In these samples (Figure 1C) the DNA concentrations were within the established parameters for sequencing (concentration greater than 50 ng/ μ L) (data not shown).

Sequencing of the FR region

We searched for repeats within the nucleotide sequences (Figure 2A) and in the translation sequences obtained from 6 open reading frames (Figure 2B). For the FR region we determined EPIYA-A and EPIYA-B TPM's from a multiple alignment (Figure 2A, B) using BioEdit Sequence Alignment Editor version 7.0.9.0. We found that the use of

codons from which EPIYA-B repeats arises is different from that in the NCTC 11637 *H. pylori* reference strain (The TAC codon codes for tyrosine (Y) and the GCT codon codes for Alanine (A)). In some of the evaluated samples use of the codons was also different. For example, TAT codon codes for Y residue in M3, M005, M014 and M49 samples, while A is coded by the ACT codon for samples M014 and M49 (Figure 2A).

In the FR region we found the EPIYA-A were characterized by presenting EPIYAKVNKKK(A/T/V/S)GQ residues, while the EPIYA-B repeats were characterized by presenting EPIY(A/T)(Q/K)VAKKVNAKI residues. EPIYAKVNKKKTGQ residues were found for all samples of the EPIYA-A type. EPIYAQVAKKVNAKI residues were found in the control strain and in 3 of the 5 evaluated samples (11637, M3, M005, and M28) of EPIYA-B. In the other two samples (M014, M49) the residues were EPIYTQVAKKVTQKI and EPIYTQVAKKVNAKI respectively (Figure 2B).

Sequencing of the WSR region

The WSR regions in the 3' end regions of *cagA* gene were amplified and sequenced. The evaluated sequences were translated into their 6 open reading frames (ORF). We searched for the EPIYA and FPLKRHDKVDDLKSKV unit repeats using a multiple alignment of sequences amino acids. The results are shown in the figure 2C. It was determined that the EPIYA motif is characterized by being EPIYA-C type found in the EPIYATIDDLG sequence. When the alignments were compared, it was observed that the number of EPIYA and FPLKRHDKVDDLKSKV repeat units for the amino acid sequences from the biopsies were identical with the exception of the M014 sample. It presented a sequence repeat of 102 bp which codes for a segment of 34 amino acids which contains one more EPIYA-C motif than the other segments do (Table 2). In this sample, the molecular size of the WSR region was 411 bp, a result significantly different from the other 4 sequenced samples, the average size of which was 308 bp. Those 4 samples had ABC type EPIYA repeats, while the other had ABCCC type. EPIYA-D motifs were not found in the biopsies studied.

DISCUSSION

10 gastric biopsy samples of *H. pylori* genomic DNA from patients in Ibagu e, Tolima, Colombia were analyzed. These samples' *cagA* regions, composed of FR and WSR, were amplified. This region had been previously described by Yamaoka (27). The PCR sensitivity for the amplification of the FR of the DNA extracted from biopsy samples was 0.05 ng, while for the amplification of the WSR regions it was

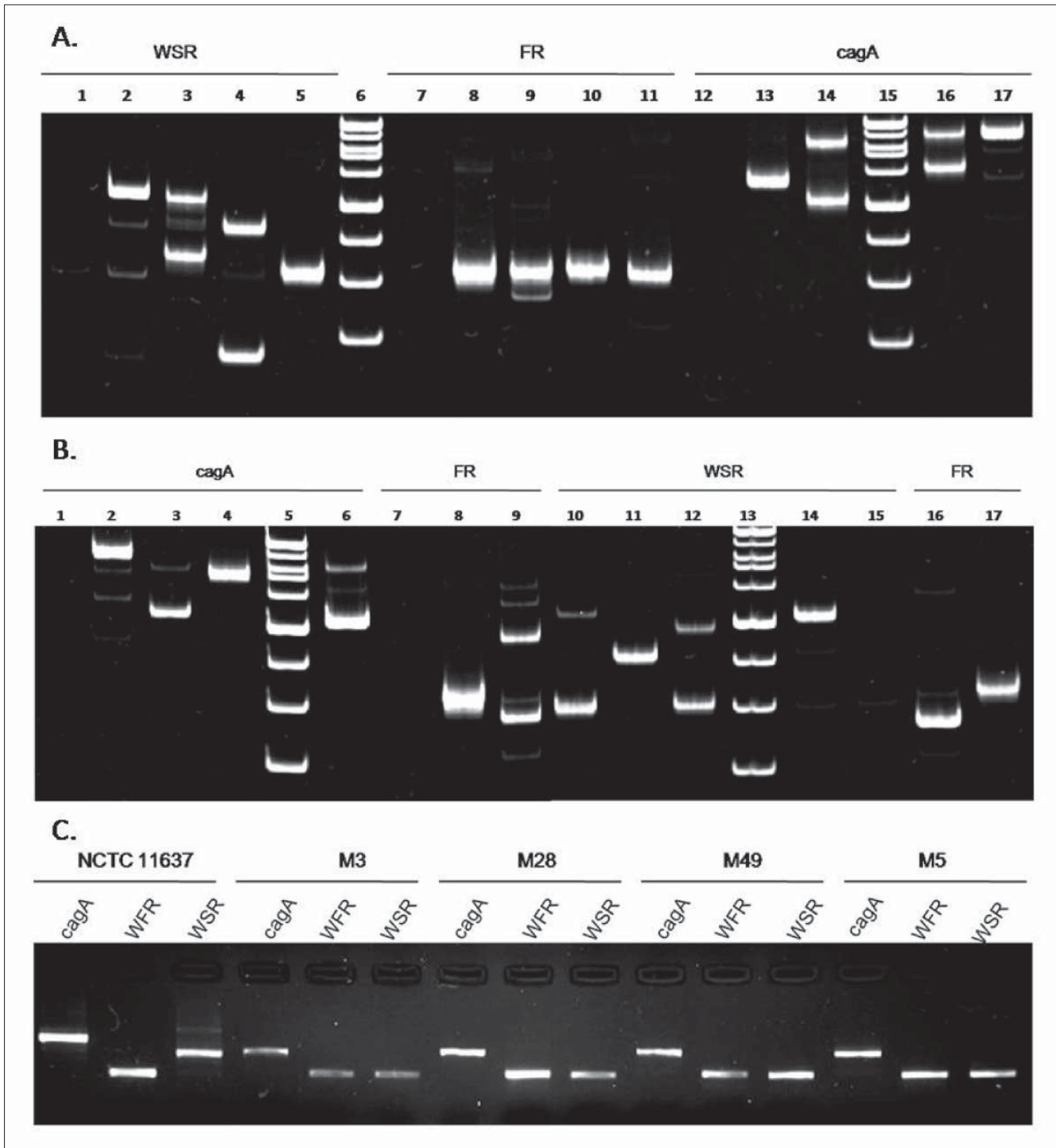


Figure 1. Amplification of *cagA* FR and WSR regions. 8% Polyacrylamide gel electrophoresis 1A) Negative control (1), NCTC 11637 (2), M004 (3), M010 (4), M49 (5), MPM (6), Negative control (7), NCTC 11637 (8), M004 (9), M010 (10), M49 (11), negative control (12), M49 (13), M010 (14), MPM (15), M004 (16), NCTC 11637 (17). 1B) Negative Control (1), NCTC 11637(2), M012 (3), M014 (4), MPM (5), M23 (6), negative Control (7), NCTC 11637 (8), M012 (9), M023 (19), M014 (11), M012 (12), MPM (13), NCTC 11637 (14), negative control (15), M23 (16) and M014 (17). 1C) Confirmation of the results of purification of the amplification products for each region. Gels were dyed with ethidium bromide.

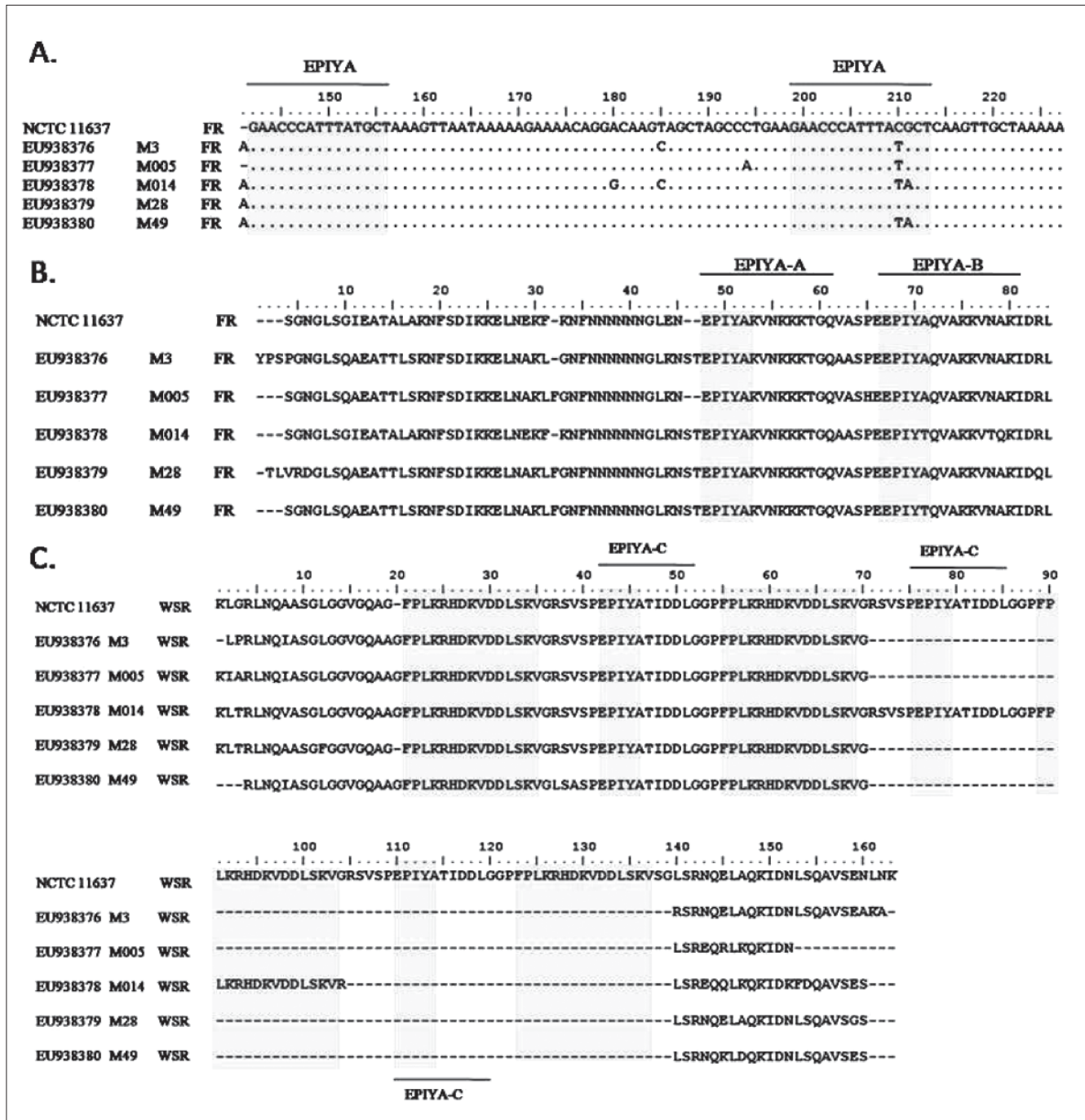


Figure 2. Alignment of 5 sequences of *H. pylori* strains from Tolima, Colombia and *H. pylori* NCTC 11637 reference strain. A) Nucleotide sequences of FR region. B) Amino acid sequence of FR region. C) Amino acid sequence of WSR region. (·) residue similarity (-) gap. The gray blocks correspond to EPIYA motifs and FPLKRHDKVDDL SKV amino acid sequences. Multiple alignment was performed using Clustal W in BioEdit Sequence Alignment Editor version 7.0.9.0.

0.01 ng (Figure 1B). These data are a good match with the results of the DNA extracted from the pure culture of *H. pylori* NCTC 11637 reference strain. The sensitivity of the reference strain was 0.078 ng for all evaluated regions, but since there is a greater proportion of human DNA than *H. pylori* DNA in the biopsies, these are good results for PCR sensitivity.

The molecular sizes of the fragments obtained from the amplification of the FR regions were similar for the samples and for the *H. pylori* NCTC11637 reference strain. This occurs as the result of the fact that in the FR region there are only nucleotides that code for EPIYA-A and -B motifs, which are constant in *cagA* positive samples. For this reason there was no significant variation of the amplicon size for this region. In contrast, in 5 of the 10 amplified samples there were two bands with different molecular weights in *cagA* and WSR region (Figures 1A and 1B). This result is consistent with the idea that in Western countries the variation of the number of EPIYA motifs, specifically EPIYA-C motifs, is contained in the amplification product of WSR region. It is also consistent with explaining the presence of more than one fragment in the WSR and *cagA* region amplifications as the product of co-infection by different *H. pylori* strains with polymorphisms in the variable region of the *cagA* gene in the same individual or by the microevolution of the same strain during a long infection period. The absence of a double band in the FR region amplifications is caused by the constant number of EPIYA repeats in this region. Even if there is co-infection or microevolution, and even if two DNA fragments from the same biopsy are present, only one band of the same size will be observed.

We first calculated the number of repeats from the molecular weights of the amplicons for *cagA* FR and WSR regions. We also calculated the number of repeats using formulas developed by Yamaoka in one of his studies with different strains of *H. pylori* isolations (27). Those formulas for EPIYA and FPLKRHDKVDDLSKV repeats are as follows: $[(218 - 227) + 57 * r] y [(174 - 177) + 102 * m]$ (r: number of FR regions, m: number of WSR region). We then compared our sequencing data with the results derived from the method proposed by Yamaoka (27) (data not shown). The number of EPIYA and FPLKRHDKVDDLSKV repeats diverged considerably between the two sets, further supporting sequencing and bioinformatics as the most efficient strategy for characterization of repeated patterns dispersed within the genome. It was determined that the sequences obtained for the 5 samples which presented single bands (M3, M005, M014, M28, M49) and for the *H. pylori* NCTC 11637 reference strain were optimal. Comparing the alignments for FR and WSR regions for each sample, with the *cagA* region amplified, we found complete coincidence. In addition the sequences obtained for *cagA* FR

and WSR regions in *H. pylori* NCTC 11637 also showed total coincidence, and therefore excellent reproducibility in relation to the previously reported sequences for this reference strain.

The sequences obtained for the 6 open reading frames (ORF) were translated for each sample with *Sixpack* by EMBOSS allowing us to obtain the amino acid sequence of each sample while taking into consideration the bacterial genetic code (NCBI translation table 11) (34). The numbers of EPIYA and FPLKRHDKVDDLSKV repeats were determined using *Fuzzpro* by EMBOSS (33). In each FR region of every sample sequenced, two EPIYA repeats separated by 14 amino acid residues were found. (Figure 2B).

Some variations were found in the codons that code for tyrosine and Alanine. EPIYA motif characterization is determined by the residues towards the C-terminal region in EPIYA (A/T) amino acid sequences (20). Once the characterization for EPIYA sites had been performed, it was found that both motifs (EPIYA-A and EPIYA-B) were in all of the sequences evaluated (Table 2). In this study, EPIYAKVNKKKTGQ and EPIYAQVAKKVNNAKI amino acid sequences corresponding to EPIYA-A AND EPIYA-B types were observed in the FR region in all evaluated samples with the exceptions of M014 and M49. We observed EPIYTQVAKKVTKI in the EPIYA-B site of M014 and EPIYTQVAKKVNNAKI in the EPIYA-B site of M49. Our observation of M014 is a new variant for the EPIYA-B site compared to that reported by Panayotopoulou and colleagues in 2007.

A higher variation in the number of EPIYA repeats for the 3' region of the *cagA* gene was found in the WSR region than in the FR region. According to Yamaoka (27), the WSR region is formed from an FPLKRHDKVDDLSKV repeat followed by 6 amino acids, then an EPIYA motif, followed immediately by a succession of 8 amino acids followed finally by another FPLKRHDKVDDLSKV repeat (Figure 2C). This pattern was found in the amino acid sequences obtained from the translations of all samples. The EPIYA motif in the WSR region is EPIYA-C (EPIYATIDDLG). The bioinformatic analysis performed with *H. pylori* NCTC 11637 reference strain allowed us to demonstrate the validity of this methodology since this positive control for EPIYA and FPLKRHDKVDDLSKV repeats has also been established in the literature (ABCCC) (20). In addition, this experiment confirmed the genetic nature of the polymorphisms in the variable region of this gene for NCTC 11637 (ABCCC), as Higashi, *et. al.* (26) have reported. The quantity of repeat units for EPIYA-C motifs varies from 1 to 3. The number of repetitions for the EPIYA-C motif varies from 0 to 3, therefore the data obtained agrees with those previously reported in strains isolated in Europe, America and Australia (26). Four of the samples evaluated were

Table 2. Determination of the number of EPIYA and FPLKRHDKVDDL SKV repeats presented in gastric biopsy samples from patients in Ibagué, Colombia.

Sample	Pathology	CagA			
		FR Region		WSR Region	
		EPIYA-A	EPIYA-B	EPIYA-C	FPLKRHDKVDDL SKV
M005	NG	1	1	1	2
M014	AG	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

(NG) Non-atrophic gastritis, (AG) Atrophic gastritis, (NA) Not applicable.

ABC type. This result resembled the statistics in Western countries where *H. pylori* CagA positive strains with EPIYA type ABC motifs commonly circulate (35). Only the M014 sample had a higher number of repeats in the EPIYA-D motif than did the other samples (Table 2), thus its genotype was ABCC. The fact that we did not find the EPIYA-D motif in the samples evaluated is consistent with findings that this motif belongs only to *H. pylori* strains endemic in East Asia (9, 26).

The variation in the C-end region of the CagA protein has been related to the biological activity of this protein. *H. pylori* cagA's positive pathogenicity is influenced by the EPIYA-C duplication because it is a critical determinant of the CagA protein. This protein, which is an important virulence factor, is able to disturb cell signaling, (26) and is also able to generate important pathologies such as gastric cancer (36). The number of EPIYA-C motifs is directly correlated with phosphorylation levels of SHP-2. The CagA SHP-2 complex is commonly detected in the atrophic mucosa which could be related to the development of atrophic gastritis and the transition of the atrophy to intestinal metaplasia (37). In this study the analysis of the EPIYA-A and EPIYA-B motifs is also important because other studies have indicated that these motifs are important for the functioning of the CagA protein. They are responsible for the association of the CagA protein to epithelial cell membranes (26).

This study's results make evident that with the use of DNA extracted directly from gastric biopsies it is possible to perform molecular and bioinformatic characterization that accurately determines the type and number of EPIYA repeats in the 3' end region of the cagA gene. It has been suggested that the number and type of EPIYA motifs could be an important marker for determination of the degree of *H. pylori* pathogenicity. Since studies undertaken in Latin

American countries including Colombia have reported relatively high numbers of *H. pylori* positive strains to the Genebank data base (38), the establishment of this kind of strategy for analysis of nucleotide and amino acid sequences of proteins coded by genes related to this virulence could lead to the establishment of markers for susceptibility to the development of severe pathologies such as gastric cancer.

Acknowledgements

To the project's participants for the use of molecular techniques for *H. pylori* genotyping and for the detection of human interleukin-1 β gene and its association with gastroduodenal pathologies in Colombian patients. To the members of the Molecular diagnosis and Bioinformatics Laboratory. To the National Institute of Cancerology for contributing the reference strain used in the study. To the Science Department of Los Andes University

Financial conflicts of interest

The authors affirm that there are no conflicts of interest.

REFERENCES

- 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.
- Parkin DM, Bray FI, Devesa SS. Cancer burden in the year 2000. The global picture. *Eur J Cancer* 2001; 37: 4-66.
- IARC. IARC monographs of the evaluation of carcinogenic risks to humans. IARC 1994; 61: 177-241.
- 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. Zambon 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+ 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, Papadacos 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.