Analysis of \textit{cagA} in \textit{Helicobacter pylori} Strains from Colombian Populations with Contrasting Gastric Cancer Risk Reveals a Biomarker for Disease Severity

John T. Loh\textsuperscript{1}, Carrie L. Shaffer\textsuperscript{3}, M. Blanca Piazuelo\textsuperscript{2}, Luis E. Bravo\textsuperscript{5}, Mark S. McClain\textsuperscript{1}, Pelayo Correa\textsuperscript{2}, and Timothy L. Cover\textsuperscript{1,2,4}

Abstract

\textbf{Background:} \textit{Helicobacter pylori} infection is a risk factor for the development of gastric cancer, and the bacterial oncoprotein CagA contributes to gastric carcinogenesis.

\textbf{Methods:} We analyzed \textit{H. pylori} isolates from persons in Colombia and observed that there was marked variation among strains in levels of CagA expression. To elucidate the basis for this variation, we analyzed sequences upstream from the CagA translational initiation site in each strain.

\textbf{Results:} A DNA motif (AATAAGATA) upstream of the translational initiation site of CagA was associated with high levels of CagA expression. Experimental studies showed that this motif was necessary but not sufficient for high-level CagA expression. \textit{H. pylori} strains from a region of Colombia with high gastric cancer rates expressed higher levels of CagA than did strains from a region with lower gastric cancer rates, and Colombian strains of European phylogeographic origin expressed higher levels of CagA than did strains of African origin. Histopathologic analysis of gastric biopsy specimens revealed that strains expressing high levels of CagA or containing the AATAAGATA motif were associated with more advanced precancerous lesions than those found in persons infected with strains expressing low levels of CagA or lacking the AATAAGATA motif.

\textbf{Conclusions:} CagA expression varies greatly among \textit{H. pylori} strains. The DNA motif identified in this study is associated with high levels of CagA expression, and may be a useful biomarker to predict gastric cancer risk.

\textbf{Impact:} These findings help to explain why some persons infected with \textit{cagA}-positive \textit{H. pylori} develop gastric cancer and others do not. \textit{Cancer Epidemiol Biomarkers Prev}; 20(10); 2237–49. ©2011 AACR.

Introduction

\textit{Helicobacter pylori} is a gram-negative bacterium that persistently colonizes the human gastric mucosa. Most \textit{H. pylori}-infected persons remain asymptomatic, but the presence of \textit{H. pylori} is associated with an increased risk of gastric cancer. The clinical outcome of \textit{H. pylori} infection is determined by a combination of bacterial, host, and environmental factors (1–6).

An important virulence factor produced by some \textit{H. pylori} strains but not others is the CagA protein (7, 8). Epidemiologic studies have shown that infection with \textit{cagA}-positive strains is associated with an increased risk of gastric cancer, compared with infection with \textit{cagA}-negative strains (9–14). CagA is an antigenic protein that is translocated into host cells by a type IV secretion apparatus (15–19). Upon entry into host cells, CagA is tyrosine phosphorylated by Src kinases at conserved EPIYA motifs within the CagA protein. CagA interacts with multiple host cell components and causes numerous alterations in cell signaling and morphology (17, 20–23).

The incidence of gastric cancer varies considerably throughout the world. Within Colombia, 2 regions located less than 200 km apart in the state of Narino have a markedly different incidence of gastric cancer. In a study conducted in 1976, the incidence of gastric cancer in the Andean mountainous region was estimated to be 150 cases per 100,000 inhabitants, whereas in the coastal region, gastric cancer rates were estimated to be 6 cases per 100,000 inhabitants (24, 25). The \textit{H. pylori} infection rates in the high-risk region are very similar in the populations inhabiting these regions (26). The prevalence of \textit{cagA}-positive strains in the low-risk region is slightly (~9%) lower than that found in the high-risk region (27), but this difference is unlikely to account for the 25-fold difference in gastric cancer rates. CagA proteins with a greater number of EPIYA motifs
undergo more tyrosine phosphorylation and potentially have increased intracellular activity compared with proteins with fewer EPIYA motifs (28–30), but the number of CagA EPIYA motifs is not significantly different in strains from the low-risk and high-risk regions of Colombia (31). Recently, it was noted that cagA-positive strains found in the low- and high-risk regions of Colombia differ in phylogeographic origin (32). Multilocus sequence typing (MLST) analysis indicated that all of the tested H. pylori strains from the high-risk region were of European origin, whereas the majority of strains from the low-risk region were of African origin (32).

In this study, we investigated CagA expression in H. pylori strains isolated from Colombian patients, and observed that there is considerable variation in the levels of CagA expressed by different strains. We identified a motif upstream of the CagA ATG initiation site that is present more frequently in strains expressing high levels of CagA than in strains expressing low levels of CagA, and we show experimentally that this motif is a determinant of CagA expression levels. We report that H. pylori strains of European and African phylogeographic origin differ in the levels of CagA expressed. Finally, we report that, in comparison with H. pylori strains expressing low levels of CagA, strains expressing higher levels of CagA are associated with more advanced precancerous lesions in gastric biopsy samples.

Materials and Methods

Bacterial strains

H. pylori strains used in this study were isolated from antral mucosa biopsy samples from subjects in the state of Nariño, Colombia, and H. pylori strain 26695 (33) was used as a reference strain. The strains analyzed in this study were randomly selected from a previously characterized set of Colombian strains (32) using a blinded code. All strains were cagA-positive and contain type s1m1 vacA (32). Several strains were excluded from analysis due to contamination or failure to grow in liquid broth culture. The 36 strains analyzed in this study (17 from subjects in a region of low cancer risk and 19 from subjects in a region of high cancer risk) are shown in Table 1. H. pylori strains were grown in room air supplemented with 5% CO2 at 37°C. For routine growth, H. pylori strains were maintained on trypticase soy agar plates supplemented with sheep blood. In addition, H. pylori strains were grown in modified Brucella broth containing 5% FBS (BB-FBS) or on BB-FBS agar plates. When necessary, BB-FBS agar plates were supplemented with metronidazole (7.5 μg/mL) or chloramphenicol (5 μg/mL). Escherichia coli strains were grown on Luria Bertani medium. When necessary, the E. coli culture medium was supplemented with ampicillin (50 μg/mL) or chloramphenicol (25 μg/mL).

Human subjects

The H. pylori strains used in this study were obtained from 36 male subjects between the ages of 39 and 60 years with dyspeptic symptoms who underwent gastrosmotestinal tract endoscopy in 2 public hospitals in the State of Nariño, Colombia (32). The hospitals are located in 2 cities with contrasting gastric cancer risks: Túquerres in the Andes Mountains where gastric cancer incidence is high, and Tumaco on the coast where gastric cancer incidence is low. The 36 individuals included in this study were part of a larger series of subjects that has been described previously (32).

Gastric mucosa biopsy samples from the 36 subjects were obtained from the antrum, incisura angularis, and corpus, and embedded in paraffin for histology. One additional antral biopsy from each subject was immediately frozen in glycerol and thiglycolate for H. pylori culture, and kept at −80°C. The samples were shipped on dry ice to Vanderbilt University (Nashville, Tennessee) for analysis. All participants provided informed consent and all samples were sent to Vanderbilt University without identifiers. The protocol was approved by the Committees on Ethics of Universidad del Valle and Hospital Departamental in Nariño, Colombia, and by the Institutional Review Board at Vanderbilt University.

Mutagenesis of the cagA promoter region in H. pylori 26695

To facilitate introduction of unmarked mutations into H. pylori, a cat-rdxA cassette was synthesized and cloned into pUC57 vector (Generscript). This cassette confers resistance to chloramphenicol mediated by the chloramphenicol acetyl-transferase (cat) gene from Campylobacter coli, and susceptibility to metronidazole is mediated by an intact rdxA gene (HP0954) from H. pylori 26695 (34). To generate unmarked mutant strains, we used a method that is a variant of counterselection methods used previously in H. pylori (35, 36). As a first step, the rdxA gene, which confers resistance to metronidazole, and approximately 0.5 kb of flanking DNA on each side, was PCR-amplified from H. pylori 26695 and cloned into pGEM670 by using pMM670 as the template for inverse PCR and then ligating the PCR product, we generated a modified plasmid (pMM672) in which the coding region of rdxA was deleted. H. pylori 26695 was transformed with pMM672 (which is unable to replicate in H. pylori) and metronidazole-resistant colonies were selected (34). PCR analysis confirmed that the rdxA locus was deleted from a mutant strain designated H. pylori 26695ΔrdxA.

We next modified H. pylori 26695ΔrdxA to generate a strain (26695 cagA:cat-rdx-9) containing a cat-rdxA cassette immediately downstream of the ATG initiation site of the cagA gene. To generate this mutant, primers 5’TGCCTTGTAGAAAATTGTAGATCTTTAGG3’ and 5’-TATCCCAACAAATCCACCACCACG-3’ were utilized to PCR amplify a DNA fragment from H. pylori 26695 comprising approximately 650 bp upstream and 650 bp downstream of the cagA translational start site. This PCR product was cloned into pGEM-670 (Promega), and the resultant plasmid was used as a template for inverse PCR with primers that introduced a BamHI site immediately

Downloaded from cebp.aacrjournals.org on June 25, 2017. © 2011 American Association for Cancer Research.
following the cagA translation start site. The resultant plasmid was digested with BamHI and a cassette encoding the cat-rdxA cassette (described above) was ligated into the restriction site. The cagA::cat-rdxA plasmid, which is unable to replicate in H. pylori, was transformed into H. pylori 26695, and single colonies resistant to chloramphenicol (5 μg/mL) but sensitive to metronidazole (7.5 μg/mL) were selected. Transformant 26695 cagA:cat-rdxA-9 was selected for further studies. Immuno- blot analysis revealed the loss of CagA expression in this strain, and the insertion of the cat-rdxA cassette into the cagA gene was confirmed by PCR amplification and DNA

<table>
<thead>
<tr>
<th>Strain</th>
<th>CagA expression*</th>
<th>Number of AATAAGATA motifsb</th>
<th>Regionc</th>
<th>MLST typed</th>
<th>Histopathology scoree</th>
</tr>
</thead>
<tbody>
<tr>
<td>5001</td>
<td>1.0</td>
<td>0</td>
<td>Low</td>
<td>hpAfrica1</td>
<td>2.3</td>
</tr>
<tr>
<td>5004</td>
<td>1.4 ± 0.2</td>
<td>0</td>
<td>Low</td>
<td>hpAfrica1</td>
<td>2.3</td>
</tr>
<tr>
<td>5005</td>
<td>0.4 ± 0.2</td>
<td>0</td>
<td>Low</td>
<td>hpAfrica1</td>
<td>1.7</td>
</tr>
<tr>
<td>5007</td>
<td>1.4 ± 0.3</td>
<td>0</td>
<td>Low</td>
<td>hpAfrica1</td>
<td>2.0</td>
</tr>
<tr>
<td>5009</td>
<td>1.4 ± 0.2</td>
<td>1</td>
<td>Low</td>
<td>hpEurope</td>
<td>1.7</td>
</tr>
<tr>
<td>5010</td>
<td>0.5 ± 0.3</td>
<td>0</td>
<td>Low</td>
<td>hpAfrica1</td>
<td>2.0</td>
</tr>
<tr>
<td>5016</td>
<td>1.8 ± 0.2</td>
<td>0</td>
<td>Low</td>
<td>hpAfrica1</td>
<td>3.5</td>
</tr>
<tr>
<td>5017</td>
<td>0.9 ± 0.3</td>
<td>0</td>
<td>Low</td>
<td>hpAfrica1</td>
<td>2.3</td>
</tr>
<tr>
<td>5018</td>
<td>1.6 ± 0.5</td>
<td>1</td>
<td>Low</td>
<td>hpAfrica1</td>
<td>3.5</td>
</tr>
<tr>
<td>5022</td>
<td>0.9 ± 0.3</td>
<td>0</td>
<td>Low</td>
<td>hpEurope</td>
<td>4.5</td>
</tr>
<tr>
<td>5024</td>
<td>1.9 ± 0.1</td>
<td>0</td>
<td>Low</td>
<td>hpAfrica1</td>
<td>2.0</td>
</tr>
<tr>
<td>5028</td>
<td>1.1 ± 0.4</td>
<td>1</td>
<td>Low</td>
<td>hpEurope</td>
<td>3.8</td>
</tr>
<tr>
<td>5032</td>
<td>2.3 ± 0.2</td>
<td>1</td>
<td>Low</td>
<td>hpEurope</td>
<td>3.5</td>
</tr>
<tr>
<td>5038</td>
<td>3.2 ± 0.5</td>
<td>1</td>
<td>Low</td>
<td>hpEurope</td>
<td>3.5</td>
</tr>
<tr>
<td>5043</td>
<td>3.7 ± 0.7</td>
<td>1</td>
<td>Low</td>
<td>hpEurope</td>
<td>5.3</td>
</tr>
<tr>
<td>5051</td>
<td>2.4 ± 0.2</td>
<td>1</td>
<td>Low</td>
<td>hpAfrica1</td>
<td>2.0</td>
</tr>
<tr>
<td>5054</td>
<td>1.6 ± 0.6</td>
<td>0</td>
<td>Low</td>
<td>hpAfrica1</td>
<td>2.3</td>
</tr>
<tr>
<td>5056</td>
<td>4.0 ± 0.8</td>
<td>1</td>
<td>High</td>
<td>hpEurope</td>
<td>4.5</td>
</tr>
<tr>
<td>5057</td>
<td>2.4 ± 0.4</td>
<td>1</td>
<td>High</td>
<td>hpEurope</td>
<td>3.5</td>
</tr>
<tr>
<td>5068</td>
<td>2.4 ± 0.4</td>
<td>1</td>
<td>High</td>
<td>hpEurope</td>
<td>4.3</td>
</tr>
<tr>
<td>5089</td>
<td>3.5 ± 0.6</td>
<td>1</td>
<td>High</td>
<td>hpEurope</td>
<td>4.3</td>
</tr>
<tr>
<td>5072</td>
<td>0.7 ± 0.2</td>
<td>1</td>
<td>High</td>
<td>hpEurope</td>
<td>3.5</td>
</tr>
<tr>
<td>5074</td>
<td>5.1 ± 1.1</td>
<td>1</td>
<td>High</td>
<td>hpEurope</td>
<td>4.7</td>
</tr>
<tr>
<td>5080</td>
<td>0.5 ± 0.3</td>
<td>1</td>
<td>High</td>
<td>hpEurope</td>
<td>3.5</td>
</tr>
<tr>
<td>5082</td>
<td>4.4 ± 1.3</td>
<td>1</td>
<td>High</td>
<td>hpEurope</td>
<td>4.8</td>
</tr>
<tr>
<td>5086</td>
<td>4.8 ± 1.1</td>
<td>1</td>
<td>High</td>
<td>hpEurope</td>
<td>4.6</td>
</tr>
<tr>
<td>5090</td>
<td>1.0 ± 0.2</td>
<td>0</td>
<td>High</td>
<td>hpEurope</td>
<td>2.0</td>
</tr>
<tr>
<td>5091</td>
<td>1.9 ± 0.2</td>
<td>1</td>
<td>High</td>
<td>hpEurope</td>
<td>4.5</td>
</tr>
<tr>
<td>5093</td>
<td>3.5 ± 0.3</td>
<td>1</td>
<td>High</td>
<td>hpEurope</td>
<td>4.5</td>
</tr>
<tr>
<td>5095</td>
<td>2.0 ± 0.2</td>
<td>1</td>
<td>High</td>
<td>hpEurope</td>
<td>2.0</td>
</tr>
<tr>
<td>5096</td>
<td>1.5 ± 0.1</td>
<td>0</td>
<td>High</td>
<td>hpEurope</td>
<td>4.3</td>
</tr>
<tr>
<td>5097</td>
<td>1.0 ± 0.3</td>
<td>1</td>
<td>High</td>
<td>hpEurope</td>
<td>4.4</td>
</tr>
<tr>
<td>5100</td>
<td>2.2 ± 0.1</td>
<td>1</td>
<td>High</td>
<td>hpEurope</td>
<td>2.3</td>
</tr>
<tr>
<td>5104</td>
<td>4.7 ± 0.6</td>
<td>1</td>
<td>High</td>
<td>hpEurope</td>
<td>3.5</td>
</tr>
<tr>
<td>5112</td>
<td>2.2 ± 0.1</td>
<td>1</td>
<td>High</td>
<td>hpEurope</td>
<td>4.3</td>
</tr>
<tr>
<td>5114</td>
<td>0.9 ± 0.3</td>
<td>0</td>
<td>High</td>
<td>hpEurope</td>
<td>4.4</td>
</tr>
</tbody>
</table>

* CagA expression was analyzed by immunoblotting, and relative levels of expression were quantified by comparing the CagA signal intensity of each strain with that of a standard (CagA from H. pylori strain 5001). For each strain, the CagA signal was calculated for 3 independent biological samples, and the average of these values is presented. Values represent mean ± SEM.

b The number of AATAAGATA motifs upstream of the ATG translation initiation site is shown.

c H. pylori strains were isolated from regions of Colombia with either high risk or low risk for the development of gastric cancer.

d MLST analysis (32) was previously used to characterize the ancestral origin of the H. pylori strains. hpAfrica1 denotes isolates classified as either West African (hpSAfrica) or South African (hpSAfrica).

e Histopathology scores describing the severity of lesions observed in gastric tissue from which the H. pylori strains were isolated.
sequencing. To introduce alterations into the cagA promoter region, 26695 cagA:catdx-9 was transformed with various plasmids described below, and transformants resistant to metronidazole were selected.

cagA promoter swapping experiments

DNA fragments encompassing approximately 0.5 kb upstream and 0.7 kb downstream from the cagA transcription start site (0.6 kb upstream and 0.6 kb downstream of the ATG initiation site) were PCR amplified from Colombian strains using primers 5'-GCAAAAAACAACCCCAAGCTGA-3' and 5'-ACACAGTAGGCCCTCCTATTITTTC-3', and the amplified products were cloned into the pGEMT-Easy (Promega) vector. The resultant plasmids, which are unable to replicate in H. pylori, were then used to transform H. pylori strain 26695 cagA:catdx-9 (described above). Transformations were plated onto metronidazole-containing BB-FBS agar plates, and metronidazole-resistant clones were isolated. These metronidazole-resistant colonies represent recombination events resulting in the removal of the cat-rdxA cassette from the cagA ORF of H. pylori strain 26695 cagA:catdx-9. As the removal of the cat-rdxA insertion should result in the restoration of an intact cagA gene, all H. pylori transformants were subsequently screened for CagA expression. To ensure that the 26695 cagA promoter region had been replaced with the corresponding wild-type Colombian sequences, the cagA promoter regions of transformants expressing CagA were sequenced.

Targeted mutagenesis of cagA promoter region

To generate mutations in the promoter region of cagA, a 1.2 kb fragment encompassing 0.5 kb of DNA upstream and 0.7 kb downstream of the cagA transcriptional start site was PCR amplified from H. pylori strain 5056 with primers 5'-GCAAAAAACAACCCCAAGCTGA-3' and 5'-ACACAGTAGGCCCTCCTATTITTTC-3' and cloned into the plasmid vector pGEMT-Easy. Targeted mutagenesis was carried out using the Quick-Change mutagenesis kit (Agilent Technologies). The introduction of mutations into plasmids was confirmed by DNA sequencing. These plasmids were then used to transform H. pylori strain 26695 cagA:catdx-9 and transformants were selected as described above.

Western blot analysis

To compare the levels of CagA expression in H. pylori Colombian strains, the bacterial strains were inoculated into BB-FBS broth and grown overnight. Cultures were then inoculated into fresh BB-FBS to an initial OD600 of 0.1. Broth cultures were grown for 15 hours and harvested; the OD600 of the H. pylori cultures at the time of harvest was approximately 0.5 to 0.6, corresponding to cultures in logarithmic phase. H. pylori cells were lysed using NP-40 lysis buffer (37, 38), and each sample (5 μg of protein) was subjected to SDS-PAGE and Western blot analysis. CagA expression was analyzed using a 1:10,000 dilution of a CagA-specific polyclonal antibody generated against amino acids 1-300 (Santa Cruz Biotechnology), or a 1:30,000 dilution of an independent polyclonal anti-CagA antiserum generated against a larger segment of CagA (amino acids 1-880). Goat anti-rabbit horseradish peroxidase (1:6,000) was used as a secondary antibody and immunoreactive bands were visualized using a chemiluminescence kit (GE healthcare). Following CagA analysis, blots were treated with Restore Western Blot Stripping Buffer (Thermo), and reprobed with antiserum generated against soluble proteins of H. pylori (anti-HP, 1:10,000; ref. 39). To quantify levels of CagA expression, we compared the CagA signal in each sample with the CagA signal of a reference strain (strain 5001). Densitometry was done using ImageJ software (NIH), and background signal levels were subtracted from measurements of CagA-specific signal intensities. For each strain, 3 independent samples were analyzed, and the average of these values was calculated.

Real-time PCR

Total RNA was isolated from H. pylori using Trizol Reagent (Gibco) according to manufacturer’s instructions. The RNA was digested with RQ1 RNase-free DNase (Promega), and the RNA samples were then subjected to a clean-up step using RNAeasy columns (Qiagen). cDNA synthesis was done on 100 ng of purified RNA using the first-strand cDNA synthesis kit (Biorad). As controls, first-strand cDNA reactions were carried out in parallel without reverse transcriptase. The cDNA and control reactions were diluted 1:20 and used in real-time PCR reactions. Real-time PCR was done using an ABI Real-Time PCR machine, with SYBR green as the fluorochrome. Abundance of transcript was calculated using the ΔΔCt method, with each transcript signal normalized to the abundance of the 16S rRNA internal control. The normalized transcript signal for each sample was then compared with similarly normalized values obtained with the 26695 wild-type strain. The primers used for real-time analysis are as follows:

16S rRNA: 5'-GGAGTACCGTGCAAGATTAAA-3' and 5'-CTACGGGATTCCTCCAATGTCAA-3'; cagA: 5'-GAGTCATAATGGCATAGAACCTGAA-3' and 5'-TGGTGCAAGAAATTTCCATGAAA-3'; HP1588 (encoding a hypothetical protein): 5'-GCTGTCAATGTTGGCAGAATGC-3' and 5'-TGCAATATCACATCGCTCCATA-3'; ureA (encoding urease subunit A): 5'-GAAGACATCATATTCAACAGAGCAA-3' and 5'-GTTACGGCCAAATGCAATCAA-3';

Analysis of histopathology

For each of the 36 subjects from whom H. pylori strains were isolated, a global gastric histologic diagnosis was determined independently by 2 pathologists (P.C. and M.B.P.). This diagnosis was based on analysis of all gastric biopsies from antrum, incisura angularis, and corpus, according to the updated Sydney system for the classification of gastritis (40) and to the Padova International...
Classification for dysplasia (41). Both pathologists were blinded to the geographic origins of the subjects and data pertaining to \(H. pylori\) isolates. Global diagnoses were assigned on an ordinal scale from (1) to (6), as follows: (1) mild to moderate nonatrophic gastritis; (2) severe nonatrophic gastritis; (3) multifocal atrophic gastritis without intestinal metaplasia; (4) intestinal metaplasia (IM); (5) dysplasia; and (6) adenocarcinoma. A previously validated histopathology scoring system taking into account the extent and degree of atrophic, metaplastic, and dysplastic changes was also used to quantify differences in morphologic variables within each global diagnosis category. The assignment of these values was accomplished as described previously (32).

Results

Expression of CagA in Colombian \(H. pylori\) strains

This study began with an analysis of CagA expression in \(H. pylori\) strains isolated from subjects in Colombia. We detected considerable variation in the levels of CagA expressed by these strains. For example, immunoblotting studies revealed that CagA expression in strains 5001, 5004, 5005, and 5007 was lower than CagA expression in strains 5043, 5056, 5074, and 5093 (Fig. 1A), and this difference was recapitulated when the samples were immunoblotted with a different CagA antiserum (data not shown). The relative level of CagA expression was quantified by comparing the CagA signal of each \(H. pylori\) strain to the CagA signal of a strain selected as a reference (strain 5001). As shown in Table 1, the relative expression levels of CagA expression in the panel of Colombian strains ranged from a minimum of 0.4 to a maximum of 5.1.

Strain-specific sequences upstream from cagA influence levels of CagA expression

We next sought to determine whether the observed heterogeneity in levels of CagA expression among Colombian strains was attributable to variations in sequences upstream of the cagA ATG translation initiation site. As a first approach, we amplified 1.2 kb sequences (0.5 kb upstream of the cagA transcriptional start site) from the 8 Colombian strains described above, and introduced these sequences into \(H. pylori\) strain 26695 cagA::catrdx-9 (Fig. 1B). These experiments resulted in replacement of the 26695 sequence with a minimum of 146 bp upstream of the cagA transcriptional start site (or 250 bp upstream of the cagA ATG initiation site), derived from Colombian strains. The regions of the \(H. pylori\) 26695 genome replaced with sequences derived from Colombian strains are shown in Figure 1C. As shown in Figure 1B, \(H. pylori\) 26695 cagA::catrdx-9 transformants containing DNA derived from strains 5001, 5004, 5005, and 5007 expressed relatively low levels of CagA compared with transformants containing DNA derived from strains 5043, 5056, 5074, and 5093. Thus, the levels of CagA expression observed in transformants correlated with the relative levels of CagA expression in the corresponding wild-type strains that were the sources of transformed DNA (Fig. 1A and B). These results indicate that variation among \(H. pylori\) strains in levels of CagA expression is attributable to variations in the region of DNA incorporated into \(H. pylori\) 26695 cagA::catrdx-9. The DNA incorporated into \(H. pylori\) 26695 cagA::catrdx-9 from Colombian strains contains the transcriptional initiation site of cagA, the cagA promoter, an AT-rich inverted repeat important for cagA transcription (42), and the ATG translation initiation site (Fig. 1D).
Comparative analysis of sequences upstream from \textit{cagA}

Next, we analyzed the nucleotide sequences upstream from \textit{cagA} in strains expressing high levels of CagA and strains expressing low levels of CagA, in an effort to identify sequence differences that might account for the differences in CagA expression. Specifically, we sequenced and analyzed an approximately 1.2 kb region (0.5 kb upstream and 0.7 kb downstream of the transcriptional start site) from the 8 strains described above (4 strains expressing relatively high levels of CagA and 4 strains expressing relatively low levels of CagA). Several differences among strains were identified upstream of the \textit{cagA} transcriptional start site (Fig. 2A). These differences include variations in the number of copies of a TGGATC motif (1 or 2 copies) located 344 bp upstream of the transcriptional start site, and variation in the number of copies of a TAATGA motif (0, 1, 2 copies) located within an AT-rich region of the \textit{cagA} promoter that contains an inverted repeat (Fig. 2A). The predicted $-10$ promoter sequences also differed among the strains; the observed $-10$ sequences included TATAATGA, TATAGTGA, TATTATGA, and TATTATAA (Fig. 2B). In addition, there was variation among strains in the presence of aTgGn motif.

![Figure 2. Analysis of an AATAAGATA motif.](image)

A, regions upstream from \textit{cagA} that exhibit substantial sequence heterogeneity among Colombian \textit{H. pylori} strains are illustrated. The Colombian strains vary in the number of copies of TGGATC, TAATGA, and AATAAGATA motifs. The TAATGA motifs are located within an AT-rich region of the \textit{cagA} promoter that contains an inverted repeat. Variations in the \textit{cagA} $-10$ sequence are also observed among the Colombian strains. The location of each motif relative to the \textit{cagA} transcriptional start site is shown. TS denotes the position of the transcriptional start site. B, sequences immediately upstream of the \textit{cagA} ATG initiation site (indicated by a large star) in 8 \textit{H. pylori} strains were aligned. Four strains express relatively low levels of CagA (5001, 5004, 5005, 5007) and 4 express relatively high levels of CagA (5043, 5056, 5074, 5093). The AATAAGATA motif is underlined and is present in 4 strains that express high levels of CagA. The $-10$ sequences and TS are also shown. C, CagA expression in each strain was analyzed by immunoblotting, and levels of CagA expression were quantified by comparing the CagA signal intensity of each strain to that of a standard (CagA in strain 5001). The relationship between number of AATAAGATA motifs and level of CagA expression is shown. The $P$ value was determined by $t$ test.

$P = 0.0013$
(43, 44) located immediately upstream from the -10 promoter sequence. Finally, an examination of a region located between the transcriptional start site and the ATG initiation site revealed a region of nucleotide sequence heterogeneity that correlated with CagA expression levels. Specifically, an AATAAGATA motif located approximately 56 nucleotides downstream of the transcription start site (40 nucleotides upstream of the cagA ATG initiation codon; Fig. 2A) was present in the strains expressing high levels of CagA (5043, 5056, 5074, and 5093), but was absent in the strains expressing low levels of CagA (5001, 5004, 5005, and 5007; Fig. 2B).

We sequenced the same regions from 28 additional H. pylori strains, scoring the strains for the presence or absence of the AATAAGATA motif. Of the 36 sequenced regions, 23 were found to contain the AATAAGATA motif, while the remaining 13 did not contain the motif (Table 1). A comparison of the level of CagA expression with the presence or absence of the AATAAGATA motif is shown in Figure 2C and Table 1. The presence of the AATAAGATA motif is associated with higher levels of CagA expression, and the absence of the motif is associated with lower levels of CagA expression (Fig. 2C, $P = 0.0013$, t test). Among 23 strains possessing this motif, 16 expressed CagA at a level that was at least 2-fold greater than the level of CagA expressed by strain 5001 (Table 1). In contrast, all 13 strains lacking this motif expressed relatively low levels of CagA (less than 2-fold greater than the level expressed by strain 5001; $P = 0.0006$, Fisher’s exact test). No significant association was observed between levels of CagA expression and other sites of heterogeneity illustrated in Figure 2A, including the type of -10 sequence ($P = 0.15$, ANOVA), the presence of a TGn motif corresponding to an extended -10 promoter sequence ($P = 0.16$, t test), the number of copies of TGGATC motifs ($P = 0.68$, t test), or the number of copies of TAATGTA motifs ($P = 0.96$, t test; 0, 1, vs. 2).

**Targeted mutagenesis of the AATAATAGA motif**

Next, we sought to experimentally test whether the AATAAGATA motif influenced the level of CagA expression. To do this, we started with a cloned 1.2 kb DNA segment (0.5 kb upstream and 0.7 kb downstream of the cagA transcriptional start site) derived from strain 5056, which, when introduced into strain 26695 cagA::catrdx-9, resulted in higher levels of CagA expression than what was observed in wild-type H. pylori 26695 (Fig. 1B). Two separate site-specific mutations were introduced into the AATAAGATA motif of the cloned 5056 DNA. These nucleotide changes involved changing the AATAAGATA motif to either AATAATAGA (designated Mut1) or AAATTTGATA (designated Mut2; Fig. 3A). The 5056 sequences (wild-type and mutant derivatives) were introduced into H. pylori strain 26695 cagA::catrdx-9 by natural transformation. A minimum of 355 bp upstream of the transcriptional start site (459 bp upstream of the ATG initiation site) was introduced into H. pylori 26695 cagA::catrdx-9, resulting in the introduction of the 5056 cagA promoter plus wild-type or mutant variants of the AATAAGATA motif (Fig. 3B). Consistent with previous results (Fig. 1), the introduction of wild-type 5056 DNA into H. pylori strain 26695 cagA::catrdx-9 resulted in a marked increase in CagA expression compared with what was observed in wild-type H. pylori 26695 (Fig. 3C and D). In contrast, transformants harboring 5056 DNA containing mutations Mut1 and Mut2 expressed significantly lower levels of CagA (Fig. 3C). This provided evidence that the AATAAGATA motif influences levels of CagA expression.

To verify that alterations in the AATAAGATA motif resulted in alterations of cagA transcription, real-time PCR was done to measure cagA transcript levels in the aforementioned H. pylori 26695 transformants. Consistent with results obtained at the protein level, H. pylori strains transformed with DNA from wild-type strain 5056 showed a significant increase in cagA mRNA levels compared with that of wild-type strain 26695 (Fig. 3E). In contrast, little or no increase in cagA transcript levels was observed in transformants harboring 5056 sequences that contain mutations to either the TAA or GATA portions of the AATAAGATA motif. This provided further evidence that the AATAAGATA motif influences levels of CagA transcription.

**Targeted mutagenesis of additional nucleotides surrounding the AATAAGATA motif**

H. pylori strain 26695 consistently expresses relatively low levels of CagA (Fig. 1B), but an analysis of nucleotide sequences upstream of the 26695 cagA ATG initiation site revealed the presence of the AATAAGATA motif that is associated with high CagA expression (Fig. 3A). Similarly, we found that some Colombian strains failed to produce high levels of CagA, despite harboring the AATAAGATA motif (Table 1). We sought to elucidate the reason for this apparent discrepancy by further studying strain 26695. Careful inspection of the 26695 sequence revealed that even though it contained the AATAAGATA motif, it differed from strains expressing high levels of CagA in nucleotides just downstream of the AATAAGATA motif. For example, a nucleotide sequence CCGAATAGGATT immediately follows the AATAAGATA motif in strain 5056. In contrast, in H. pylori 26695, the AATAAGATA motif is followed by the nucleotide sequence CCGTAGTAT (Fig. 3A, compare 5056 wild type with 26695 wild type). We evaluated the importance of these underlined nucleotides by altering the CCGAATAGGATTAT nucleotide sequence of 5056 to CCGTAGTAT (Mut3), thus mimicking the nucleotide sequence surrounding the AATAAGATA motif in strain 26695. As shown in Figure 3D, introduction of the altered sequence (Mut3) into strain 26695 cagA::catrdx-9 resulted in a lower level of CagA expression than was observed following introduction of the wild-type sequence from strain 5056 into strain 26695 cagA::catrdx-9. Consistent with the results of immunoblotting studies, cagA transcript levels were significantly decreased in H. pylori transformants harboring the Mut3 mutation (Fig. 3E). These results indicate the importance
of not only the AATAAGATA motif, but also additional nucleotides located downstream of this motif in determining the levels of CagA expressed.

Analysis of CagA expression and the AATAAGATA motif in strains from 2 regions of Colombia

The *H. pylori* strains used in these studies originated from 2 regions in Colombia (Túquerres and Tumaco), characterized by a high risk or low risk for the development of gastric cancer, respectively. Therefore, we sought to determine whether there were any detectable differences in levels of CagA expression in strains from the 2 regions. As shown in Figure 4A, strains isolated from patients in the high-risk region expressed higher levels of CagA than did strains isolated from the low-risk region (*P* = 0.028). We also detected differences in the presence of the AATAAGATA motif in strains from the 2 regions (Fig. 4B). The AATAAGATA motif was detected more commonly in *H. pylori* strains from the high-risk region (where strains typically express high levels of CagA), than in strains from the low-risk region. Among 19 strains from the high-risk region, 16 possessed this motif. In contrast, fewer than half of the strains from the low-risk region (7 of 17) possessed the AATAAGATA motif (*P* = 0.014; Table 1, Fig. 4B).

Analysis of CagA expression and the AATAAGATA motif in strains of European or African phylogeographic origins

MLST analysis was previously used to determine the phylogeographic origin of Colombian *H. pylori* strains, and all of the strains analyzed were determined to be
either European or African in origin (32). As shown in Figure 4C, strains expressing the highest levels of CagA were European in origin, and the strains of African origin expressed relatively low levels of CagA ($P = 0.0187$). The AATAAGATA motif was present more commonly in European strains (21 of 25 strains) than in African strains (2 of 11 strains) ($P = 0.0003$; Fig. 4D). When comparing African strains (all from the low risk region) with European strains from the same low-risk region, the AATAA-GATA motif was again found to be present more commonly in European strains (5 of 6 strains), compared with African strains (2 of 11 strains; $P = 0.03$). Thus, in comparison with strains of African origin, strains of European origin are more likely to contain the AATAAGATA cagA motif and express high levels of CagA.

**Correlation of CagA expression and the AATAAGATA motifs with histopathology scores**

All of the Colombian $H. pylori$ strains analyzed in the current study were isolated from gastric biopsies, and the severity of gastric pathology (focusing on premalignant lesions) was scored as described in the Methods. To determine whether levels of CagA expression are associated with severity of gastric disease, we compared the histopathology score of each patient with the levels of CagA expressed by the corresponding $H. pylori$ strain. As shown in Figure 5A, there was a significant correlation between these 2 parameters ($r_s = 0.453$, $P = 0.0057$). $H. pylori$ strains expressing high levels of CagA were consistently associated with high histology scores, and the lowest histology scores were associated with strains expressing low levels of CagA (Fig. 5A). Figure 5B analyzes levels of CagA expression in strains from patients with gastric histologic diagnoses that are considered either precancerous (histology score $\geq 3$, indicative of atrophic gastritis with and without intestinal metaplasia or dysplasia) or nonprecancerous (histology score $< 3$, indicative of nonatrophic gastritis). The level of CagA expression in $H. pylori$ strains from patients with precancerous lesions was significantly higher than the level of CagA expression in $H. pylori$ strains from patients with nonprecancerous lesions ($P = 0.02$, Fig. 5B). Representative images illustrating gastric lesions of varying severity are shown in Figure 5C.

We next analyzed whether the presence of the AATAAGATA motif was associated with histopathology scores (Fig. 6). Strains containing the cagA AATAAGATA motif were associated with higher histopathology scores, compared with strains in which the motif is absent ($P = 0.025$, Fig. 6A). We also analyzed whether there was
any correlation between histology scores and other sites of heterogeneity upstream from \textit{cagA}, including predicted -10 sequences and the previously identified TGGATC and TAATGA motifs (Fig. 2A). As shown in Figure 6, there was no significant correlation between histopathology scores and the number of TGGATC motifs (Fig. 6B), number of TAATGA motifs (Fig. 6C), or type of -10 sequence (Fig. 6D). Therefore, an increased severity of preneoplastic lesions was correlated with both high levels of CagA expression (Fig. 5) and the presence of the AATAAGATA motif (Fig. 6A).

Discussion

Gastric cancer develops in a small subset of \textit{H. pylori}-infected persons, and the remainder remains asymptomatic. The risk of gastric cancer is determined by bacterial factors, host factors, and environmental factors (1–6). Numerous studies have shown that \textit{cagA}-positive \textit{H. pylori} strains are associated with an increased risk of gastric cancer compared with strains that lack \textit{cagA} (9–14). The risk of gastric cancer is also determined by several related bacterial features, including sequence variation within CagA (such as the number and type of EPIYA motifs; refs. 28–30), and the presence or absence of a functional \textit{cag} type IV secretion system (which translocates CagA into host cells; refs. 15–19). In this study, we show that there is considerable variation among \textit{H. pylori} strains in expression of CagA, and we provide evidence that the level of CagA expression may be an additional factor that influences gastric cancer risk.

By comparing DNA sequences upstream of the \textit{cagA} ATG initiation site, we identified an AATAAGATA DNA motif that was present significantly more frequently in...
strains expressing high levels of CagA than in strains expressing low levels of CagA. All of the strains expressing high levels of CagA contained this motif, whereas only 35% of strains expressing low levels of CagA contained this motif. Experimental studies revealed that the AATAAGATA motif is required for high levels of CagA expression, and sequences immediately downstream from this motif are also required for high-level CagA expression. Collectively, these results indicate that the AATAAGATA motif is necessary but not sufficient for high levels of CagA expression. In future studies, it will be important to precisely map the boundaries of the region that is required for high-level CagA expression. Nucleotide sequences located near promoters are often sites for binding of regulatory proteins, which can positively or negatively regulate gene transcription. In this case, since this AATAAGATA motif is located downstream of the cagA transcriptional start site within the 5′ untranslated region (5′-UTR; ref. 42), at a site somewhat distant from the promoter, it seems relatively unlikely that it serves as a binding site for a regulatory protein. Instead, we speculate that presence of the AATAAGATA motif may confer increased transcript stability.

One possible mechanism by which RNA stability is regulated is through base pairing with small regulatory RNA. The existence of numerous H. pylori sRNAs potentially involved in gene regulation has recently been described (45). In most cases, binding of sRNA to the target mRNA, typically in the 5′-UTR, results in reduced stability of the target mRNA through occlusion of the ribosome binding site and degradation of the mRNA transcript (46). However, at least one example of increased transcript stability mediated by sRNAs has been described (47). In this case, it is thought that in the absence of the sRNAs, a hairpin structure in the 5′ UTR forms, and that this hairpin structure forms a recognition site for RNA degradation. Base pairing of the sRNAs to binding sites in the 5′ UTR results prevents the stem loop structure, leading to increased stability of the mRNA transcript. Given the location of the AATAAGATA motif on the 5′ UTR of CagA, we speculate that a similar phenomenon might be operative with cagA.

The H. pylori strains used in this study were isolated from patients in 2 regions of Colombia characterized by either a high risk or low risk for the development of gastric cancer. It was recently showed by MLST analysis that cagA-positive H. pylori strains from the high-risk region are predominantly European in origin, and strains from the low-risk region are either African or European in origin (32). In this study, we observe that H. pylori isolates from the region with a high risk of gastric cancer express higher levels of CagA than the do H. pylori isolates from the region with a low cancer risk, and we find that H. pylori isolates of European origin express CagA at higher levels than do H. pylori strains of African origin. Consistent with these observations, the AATAAGATA motif was found predominantly in H. pylori strains from the high-risk region and in strains of European origin. Thus far, there
are relatively few robust genetic markers that allow differentiation of African and European strains (48). It will be important in future studies to test additional H. pylori isolates, including H. pylori strains derived from native European or African subjects, and evaluate whether presence of the AATAAGATA consistently correlates with geographic origin of strains and whether this motif provides a robust marker for CagA expression levels.

In a previous study, it was noted that Colombian H. pylori strains of European origin were associated with increased severity of preneoplastic lesions and increased DNA damage in gastric biopsy samples, compared with Colombian strains of African origin (32). In the current study, we observed that H. pylori strains expressing the highest levels of CagA were consistently associated with precancerous lesions; in contrast, only about half of the strains expressing lower levels of CagA were associated with precancerous lesions, and the remainder had nonatrophic gastritis. We propose that the differences in gastric cancer rates between the low- and high-risk regions may be attributed, at least in part, to differences in levels of CagA produced by the corresponding H. pylori strains. The differences in the gastric cancer rates in these 2 regions of Colombian are also influenced by many other microbial, host, and environmental factors (49–51). For example, persons in the low-risk region typically consume higher amounts of fruit, fresh vegetables, and seafood compared with persons in the high-risk region. Vitamin C, found in fruits and vegetables, and selenium, found in fish, may provide protection against the development of gastric cancer (52–54). As another example, intestinal parasites are more common in the low-risk region than in the high-risk region, and may also influence gastric cancer rates by diminishing the inflammatory response to H. pylori (51, 55).

In summary, many previous studies have shown the importance of CagA in the development of gastric diseases, and specific polymorphisms in CagA (such as the number and type of EPTYA motifs) have been associated with the severity of disease. In this study, we provide evidence that variation among strains in levels of CagA expression is also a factor that correlates with the severity of disease.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Grant Support

This work was supported by NIH R01 AI068009, P01 CA116087, P01 CA028842, PD0380058404 (Vanderbilt Digestive Disease Research Center), and the Department of Veterans Affairs.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 13, 2011; revised July 25, 2011; accepted August 14, 2011; published OnlineFirst August 22, 2011.

References

...remaining at risk for gastric cancer. Helicobacter 2004; 9:262−70.


Analysis of cagA in *Helicobacter pylori* Strains from Colombian Populations with Contrasting Gastric Cancer Risk Reveals a Biomarker for Disease Severity

John T. Loh, Carrie L. Shaffer, M. Blanca Piazuelo, et al.


**Updated version**
Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-11-0548

**Cited articles**
This article cites 54 articles, 22 of which you can access for free at:
http://cebp.aacrjournals.org/content/20/10/2237.full#ref-list-1

**Citing articles**
This article has been cited by 14 HighWire-hosted articles. Access the articles at:
http://cebp.aacrjournals.org/content/20/10/2237.full#related-urls

**E-mail alerts**
Sign up to receive free email-alerts related to this article or journal.

**Reprints and Subscriptions**
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

**Permissions**
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.