Heterogeneity of cag Genotypes in Helicobacter pylori Isolates from Human Biopsy Specimens

Maria Luisa Tomasini,1 Stefania Zanussi,1 Michele Sozzi,2* Rosamaria Tedeschi,1 Giancarlo Basaglia,1 and Paolo De Paoli1

Microbiology, Immunology, and Virology1 and Gastroenterology2 Units, Centro di Riferimento Oncologico, IRCCS, Aviano, Italy

Received 14 August 2002/Returned for modification 16 September 2002/Accepted 9 December 2002

The Helicobacter pylori chromosomal cluster of genes known as the cytotoxin-associated gene (cag) island may have different compositions in infecting strains. In this study, we analyzed 150 single colonies obtained from gastric biopsy specimens from 10 patients infected with cagA-positive H. pylori strains and sweep isolates (isolates harvested with sweep in different points of the plate) from 6 patients infected with cagA-negative strains. Three loci in the cag island (cagA, cagE, and virB11) and the conserved gene glmM (ureC) were investigated by PCR. The levels of anti-H. pylori and anti-CagA antibodies in patient sera were also measured. For subjects infected with cagA-negative strains, all sweep isolates were also negative for cagE and virB11, suggesting the complete absence of the cag island. For subjects infected with cagA-positive strains, most of the isolates were positive for all three genes studied, whereas 24.7% of the isolates had a partial or total deletion of the cag island. cagA, cagE, and virB11 were, respectively, present in 87.3, 77.3, and 90% of the colonies. The deletion of virB11 was always associated with the deletion of cagA and/or cagE. H. pylori colonies with different cag genotypes were isolated within a single gastric biopsy specimen from 3 of the 10 patients and were further characterized by random amplified polymorphic DNA (RAPD) analysis and by sequencing of an arbitrarily selected gene segment. Although the colonies had different cag genotypes, their RAPD profiles were highly similar within each patient, and the nucleotide sequences of the selected gene segment were identical. All of the patients had detectable antibodies against H. pylori, and 9 of 10 had anti-CagA antibodies. In conclusion, we show that a single infecting H. pylori strain may include variable proportions of colony subtypes with different cag genotypes. The extension of our analysis to patients with well-characterized gastric diseases may provide significant information on the relationship between cag genotypes and clinical outcomes of H. pylori infections.

Helicobacter pylori is a gram-negative spiral bacterium that colonizes the human stomach. Infection with H. pylori is associated with chronic gastritis, peptic ulcer, gastric adenocarcinoma, and gastric mucosa-associated lymphoid tissue lymphoma (25). Although the pathogenesis of H. pylori infection is not completely understood, several putative virulence factors may contribute to mucosal damage (5, 26). The cytotoxin-associated gene (cag) island, an approximately 40-kb cluster of genes in the H. pylori chromosome, probably was inherited by horizontal transfer from an unknown microorganism. Many of the cag genes have homology to those of other bacteria encoding virulence factors and a type IV secretion system. Therefore, they may play an important role in microbial virulence and pathogenicity (6, 10). Among H. pylori isolates, the presence of insertion elements causes differences in the compositions of the cag island; the consequent cag instability may produce differences in pathogenicity and host adaptability within a bacterial strain. The cag island can be present as a single uninterrupted unit or can be divided into two segments (cagI and cagII) by an interposed insertion element (IS605); moreover, it can be partially or totally deleted (8, 24, 28, 33). Although deletions of the cag island have been reported (17, 18, 28, 31) and cag instability has been shown for mice (34), detailed analyses of the cag island in human isolates are still insufficient.

The aims of our study were to evaluate, for H. pylori colonies isolated from human gastric biopsy specimens, heterogeneity in cag genotypes by assessing the presence of some representative genes located in different segments of the cag island. From among several genes included in the cag island, we considered the following: the cagA gene (located in the right half of the cag island), encoding the highly immunodominant protein CagA, which affects host cell physiology after being delivered to gastric epithelial cells (3, 10, 29); the cagE (picB) gene (located upstream of cagA in the right half of the cag island), encoding a protein involved in the process of interleukin 8 expression in gastric epithelial cells (4, 8, 38); and the virB11 gene (located in the left half of the cag island), encoding, together with other genes, a type IV secretion system which allows the delivery of the CagA protein to gastric epithelial cells (20, 21). To confirm the identification of the bacteria as H. pylori, PCR was also performed for glmM, a conserved gene formerly known as ureC (22, 23, 24).

The serological response against the CagA protein is an important epidemiological tool for the study of H. pylori-associated diseases (35). Therefore, we also assessed the relationship between the integrity of the cag island in H. pylori isolates and the serological presence of anti-CagA antibodies.

* Corresponding author. Mailing address: Division of Gastroenterology and Digestive Endoscopy, Centro di Riferimento Oncologico, Via Pedemontana Occidentale 12, 33081 Aviano (Pordenone), Italy. Phone: 39/0434/659362. Fax: 39/0434/659402. E-mail: msozzi@cro.it.
MATERIALS AND METHODS

Study population and endoscopy. Ten subjects infected with *H. pylori* strains with positive PCR results for *cagA* and 6 subjects infected with *cagA*-negative strains were selected from a population of dyspeptic patients who had undergone an upper gastrointestinal tract endoscopy in the gastroenterological unit of our institution.

None of the subjects had been previously treated with anti-*H. pylori* eradication therapy, proton pump inhibitors, or nonsteroidal anti-inflammatory drugs. For 6 of the 10 patients infected with *cagA*-positive strains and for 3 of the 6 patients harboring *cagA*-negative strains, multiple biopsy specimens were obtained for culturing and histological analyses were endoscopically taken from both the antrum and the corpus. The specimens were obtained by using sterilized biopsy forceps containing only from one gastric region (the antrum).

Corpus. For the remaining seven patients, multiple biopsy specimens were obtained by using sterilized biopsy forceps Pylori Selective Medium (bio-Mérieux, Rome, Italy). Cultures were incubated at 37°C in a microaerophilic environment (Campygen Oxoid, Ltd., Basingstoke, UK) for 3 to 4 days. The cultured bacteria were identified as *H. pylori* based on gram-negative staining, curved or spiral shape, and positivity for catalase, oxidase, and urease production. Identification was further confirmed by PCR. Several sweeps of colonies (obtained by harvesting isolates with sweep in different points of the plate) and (only for patients infected with *cagA*-negative strains) 10 single colonies from the primary cultures for each biopsy specimen were cultured in C in a microaerophilic environment (Campygen Oxoid, Ltd., Basingstoke, UK). Bacteria from both the 10 single colonies and the sweep suspensions were considered representative of each *H. pylori* strain, and sweep suspensions were considered representative of the whole *H. pylori* population.

Genomic DNA extraction. Bacteria from both the 10 single colonies and the sweeps were resuspended in 1.2 mL of 0.9% NaCl. Bacterial pellets were obtained by centrifugation at 5,233 × g for 5 min, and genomic DNA was extracted by using a DNeasy tissue kit (Qiagen). The amount of DNA was calculated by spectrophotometry.

**PCRs for *glmM, cagA, virB11*, and *cagE***. Primers for PCR amplifications were designed based on published sequences for *glmM, cagA, virB11*, and *cagE* in *H. pylori* strain 26695 (34) (Table 1).

Each PCR was performed with a volume of 50 μL containing 50 mM KCl, 10 mM Tris- HCl (pH 8.3), 1.5 mM MgCl₂ (Applied Biosystems, Monza, Italy), 200 μM (each) deoxynucleotide (Pharmacia Biotech, Milan, Italy), 1.5 U of *Taq* DNA polymerase (Applied Biosystems), 0.5 μM (each) primer, and 10 μL of DNA at a suitable concentration. Each reaction mixture for *glmM* was amplified as follows: denaturation at 95°C for 3 min and then 35 cycles of denaturation at 94°C for 1 min, annealing at 52°C for 1 min, and extension at 72°C for 1 min. For *cagA, cagE*, and *virB11*, the incubation conditions were as follows: 3 min at 95°C and then 50 cycles of 94°C for 1 min, 48°C (cagA), 53°C (cagE), or 49°C (virB11) for 45 s, and 72°C for 45 s (34). Cycle numbers were chosen on the basis of sensitivity tests performed on dilutions ranging from 10⁻³ to 10⁻⁷ ng of DNA extracted from five different colonies. We detected an amplification signal at up to 10⁻⁴ ng of DNA for the four genes, making the sensitivity of the PCRs for *cagA, cagE*, and *virB11* comparable to the sensitivity of the *glmM* PCR with the same standardized amount of DNA. Amplifications were performed in duplicate with a DNA thermocycler (GeneAmp PCR system 9600; Perkin-Elmer, Ueberlingen, Germany; GeneAmp PCR system 2400; Perkin-Elmer, Langen, Germany). PCR products were loaded on 8% polyacrylamide gels, electrophoresed, stained with ethidium bromide, and visualized by photography under UV transillumination.

**RAPD-PCR and sequencing of the *glmM* PCR product**. Colonies from samples 1, 4, and 8 were analyzed by random amplified polymorphic DNA (RAPD)-PCR to verify strain identity despite the diversity in cag genotypes. Two patients (patients 4 and 8) were chosen as test patients on the basis of their mixed cag genotypes, while patient 1, who had colonies with identical cag genotypes, was chosen as a control (Table 2). More information is given in Results.

**TABLE 1. Primers used for PCR and RAPD-PCR**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Positions</th>
<th>Sequence (5’–3’-50 bp)</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>glmM</em></td>
<td>F</td>
<td>780–805</td>
<td>GGATAACGCTTTTGAAGGTATGAGGG</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>1052–1075</td>
<td>GCTTGCTTTCAACTCAACGGCC</td>
</tr>
<tr>
<td><em>cagA</em></td>
<td>F</td>
<td>1232–1259</td>
<td>ATATACTGAAATTAGCACCTTGA</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>1355–1395</td>
<td>AGAACCAAAACGGCAATGACATCC</td>
</tr>
<tr>
<td><em>cagE</em></td>
<td>F</td>
<td>914–940</td>
<td>TTGAACACTCCAAAGTGAGGATACC</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>1398–1422</td>
<td>GCCCTAGGTTATATACCCATTACC</td>
</tr>
<tr>
<td><em>virB11</em></td>
<td>F</td>
<td>73–95</td>
<td>TTAATACCTCTAAGGGCATGTCAT</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>542–564</td>
<td>GATATACTGGTTTACCCGCTT</td>
</tr>
<tr>
<td>1254</td>
<td></td>
<td></td>
<td>CCGCAGGCAAC</td>
</tr>
<tr>
<td>1281</td>
<td></td>
<td></td>
<td>AAGCGGCGAAC</td>
</tr>
<tr>
<td>1290</td>
<td></td>
<td></td>
<td>GTGGATGGCGA</td>
</tr>
</tbody>
</table>

*a* F, forward; *R*, reverse.

*b* Nucleotides from the translation initiation codon of the respective gene in *H. pylori* strain 26695 (32).

**TABLE 2. *glmM, cagA, cagE*, and *virB11* PCR products obtained from *cagA*-positive *H. pylori* strains isolated from 10 patients and respective serological responses**

<table>
<thead>
<tr>
<th>Patient</th>
<th>No. of colonies positive for the given gene/no. tested</th>
<th>Presence (+) or absence (−) of:</th>
<th>Anti-<em>H. pylori</em> IgG</th>
<th>Anti-<em>CagA</em> IgG</th>
</tr>
</thead>
<tbody>
<tr>
<td>(source of biopsy specimen)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (A and C)</td>
<td>20/20 20/20 20/20 20/20</td>
<td>+ − +</td>
<td>+ − +</td>
<td></td>
</tr>
<tr>
<td>2 (A and C)</td>
<td>20/20 20/20 20/20 20/20</td>
<td>+ + +</td>
<td>+ + +</td>
<td></td>
</tr>
<tr>
<td>3 (C)</td>
<td>10/10 10/10 10/10 10/10</td>
<td>− + +</td>
<td>+ + +</td>
<td></td>
</tr>
<tr>
<td>4 (A and C)</td>
<td>20/20 17/20 0/20 17/20</td>
<td>+ − +</td>
<td>+ + +</td>
<td></td>
</tr>
<tr>
<td>5 (A and C)</td>
<td>20/20 20/20 20/20 20/20</td>
<td>+ + +</td>
<td>+ + +</td>
<td></td>
</tr>
<tr>
<td>6 (A and C)</td>
<td>20/20 20/20 20/20 20/20</td>
<td>+ + +</td>
<td>+ + +</td>
<td></td>
</tr>
<tr>
<td>7 (A)</td>
<td>10/10 10/10 10/10 10/10</td>
<td>+ + +</td>
<td>+ + +</td>
<td></td>
</tr>
<tr>
<td>8 (A)</td>
<td>10/10 2/10 4/10 6/10</td>
<td>+ + +</td>
<td>+ + +</td>
<td></td>
</tr>
<tr>
<td>9 (A)</td>
<td>10/10 10/10 10/10 10/10</td>
<td>+ + +</td>
<td>+ + +</td>
<td></td>
</tr>
<tr>
<td>10 (A)</td>
<td>10/10 2/10 2/10 2/10</td>
<td>ND ND</td>
<td>ND ND</td>
<td></td>
</tr>
</tbody>
</table>

*a* A, antrum; C, corpus.

Data shown are for single colonies. For *glmM* and *cagA*, all swipes were positive.

*ND*, not done.
RESULTS

PCR for \textit{H. pylori} sweeps. For all 16 patients, sweep bacterial suspensions were PCR positive for \textit{glmM}, confirming the presence of \textit{H. pylori} DNA. For 10 patients, sweep bacterial suspensions were PCR positive for \textit{cagA}, showing that all gastric specimens contained \textit{cagA}-positive \textit{H. pylori} subtypes (Table 2). For the remaining six patients, who harbored \textit{cagA}-negative strains, sweep bacterial suspensions were PCR negative for \textit{cagE} and \textit{virB11}.

Amplification of \textit{cagA}, \textit{cagE}, and \textit{virB11} in single colonies. A total of 150 single colonies were isolated from the 15 biopsy specimens from the 10 patients infected with \textit{cagA}-positive strains. For seven patients, PCR of all of the single colonies (10 or 20, depending on the presence of one or two gastric samples) revealed full signals for \textit{cagA}, \textit{cagE}, and \textit{virB11}, while for three patients (patients 4, 8, and 10), the colony genotypes were mixed (Table 2).

In detail, \textit{cagA} was amplified in 131 of 150 colonies (87.3%), \textit{cagE} was amplified in 116 of 150 colonies (77.3%), and \textit{virB11} was amplified in 135 of 150 colonies (90%). Most of the colonies (75.3%) were PCR positive for all three genes studied, 14.7% failed to show amplification in the right half of the \textit{cag} island (absence of \textit{cagA} and/or \textit{cagE}), and 10% showed absent PCR signals both in the right half and in the left half (absence of \textit{cagA} and/or \textit{cagE} and \textit{virB11}, respectively). Interestingly, a lack of \textit{virB11} amplification was always associated with an absent signal for at least one of the genes in the right half of the \textit{cag} island.

On the basis of the genetic composition of the \textit{cag} island, the \textit{H. pylori} strains analyzed in our study were divided into seven groups (Table 3). In particular, all colonies from patients 1, 2, 3, 5, 6, 7, and 9 were genotype I; colonies from patient 4 were genotype II from the antrum and genotypes II, IV, and V from the corpus. The colonies from patient 8 were the most heterogeneous, being genotypes I, II, III, IV, VI, and VII. Most of the colonies from patient 10 were genotype III, while two were genotype I.

RAPD-PCR and sequencing of the \textit{glmM} PCR product. RAPD fingerprints were compared on the basis of criteria adopted by other authors (12, 32, 36). We considered fingerprints that were different in only one band to be highly similar; variations in band intensity were not taken into account.

On the basis of these criteria, the RAPD patterns were found to be highly similar in all of the colonies from each patient studied (patients 1, 4, and 8). Figure 1 shows the RAPD pattern of the 10 colonies isolated from patient 8. In this patient, strain identity was confirmed by sequencing and comparing the nucleotide sequences of the \textit{glmM} PCR products, which were identical in all 10 single colonies.

\textbf{Anti-\textit{H. pylori} and anti-CagA antibodies and their association with \textit{cag} genotypes.} Of the 10 patients infected with \textit{cagA}-positive strains, 5 (patients 2, 5, 7, 8, and 9) were positive for both anti-\textit{H. pylori} IgG and IgA, 3 (patients 1, 4, and 6) were positive for IgG but negative for IgA, and 1 (patient 3) was negative for IgG but positive for IgA (Table 2). ELISA and Western blot analyses for the detection of anti-CagA IgG gave a negative result for patient 8 and positive results for the remaining nine patients (Table 2). The six patients infected with \textit{cagA}-negative strains were positive for IgG.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
\textbf{Group} & \textbf{No. of colonies} & \textbf{Genotype} & \textbf{Patient(s) with the indicated pattern (source of specimen)} \\
\hline
I & 113 & + & + & + & 1, 2, 3, 5, 6, 7, 8, 9, 10 \\
II & 15 & + & + & + & 4 (A), 8 \\
III & 10 & + & + & + & 4, 10 \\
IV & 6 & + & + & + & 4 (C), 8 \\
V & 3 & + & + & + & 4 (C) \\
VI & 2 & + & + & + & 8 \\
VII & 1 & + & + & + & 8 \\
\hline
\end{tabular}
\caption{Classification of single colonies isolated from patients infected with \textit{cagA}-positive \textit{H. pylori} strains into seven genotypic groups}
\end{table}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig1.png}
\caption{\textit{H. pylori} colonies with different \textit{cag} genotypes (obtained from patient 8) show highly similar patterns on RAPD-PCR analysis (primer 1290). Lane M, molecular weight marker (Gene Ruler 100-bp DNA Ladder Plus).}
\end{figure}
and/or IgA anti-*H. pylori* antibodies and negative for anti-CagA antibodies.

**DISCUSSION**

In the present study, we investigated the heterogeneity of the *cag* island in sweeps and individual colonies isolated from *H. pylori*-infected patients. Our purpose was to assess the proportions of different *H. pylori* *cag* genotypes in bacterial populations obtained from individual human gastric biopsy specimens.

Three representative genes (*cagA*, *cagE*, and *virB11*) located in different segments of the *cag* island were studied. As expected, for patients harboring *cagA*-negative strains, the other two genes also were lacking in sweep isolates, suggesting the complete absence of the *cag* island. For patients infected with *cagA*-positive *H. pylori* strains (i.e., results of *cagA* PCR were positive for sweeps of colonies from all gastric biopsy specimens), at least one of these genes was lacking in 30% of cases. Of the 150 single colonies isolated from these patients, most (75%) were PCR positive for all three genes, whereas 25% showed deletions either in the right half of the *cag* island (*cagA* and/or *cagE*) or in both the right half and the left half but never in the left half alone (*virB11*).

Although the coexistence of *cagA*-positive and *cagA*-negative *H. pylori* strains has been described for a single patient and even for a single biopsy specimen (15, 39), our analysis showed that the proportions of *cagA*-negative colonies may be quite variable in gastric biopsy specimens obtained from different subjects and that *cagA* negativity is almost always associated with *cagE* and/or *virB11* negativity. Collectively, these data suggest that future studies aimed at finding correlations between *H. pylori* *cag* genotypes and different clinical outcomes need to take into account the fact that each biopsy specimen encompasses a complex mixture of *H. pylori* subtypes.

The serological response to the CagA protein has been used to identify subjects carrying *cagA*-positive strains and to investigate the relationship between such strains and disease risk (9, 11, 14, 31, 40). In our study, all patients carrying only *cagA*-positive, *cagE*-positive, and *virB11*-positive *H. pylori* colonies also had detectable antibodies against the CagA protein; however, of the three patients (patients 4, 8, and 10) carrying mixed colonies, one was anti-CagA antibody negative (patient 8). As expected, all patients infected with *cagA*-negative strains showed no serological response to the CagA protein.

These data suggest that in some patients, low proportions of *cagA*-positive colonies may produce amounts of CagA protein that are insufficient to trigger a detectable antibody response. The simultaneous deletion of *cagE*, a gene that is involved in the secretion of interleukin 8, may also result in the inability of the immune system to produce antibodies against CagA (4, 8). In contrast, the presence of a detectable anti-CagA antibody response in other patients carrying low proportions of *cagA*-positive and/or *cagE*-positive colonies suggests that additional factors, such as specific host genetic traits, play a significant role in controlling the immune response against CagA, as has been seen with other microbial proteins (13, 27). Collectively, our data show that patients negative for anti-CagA antibodies may harbor, although rarely, *cagA*-positive strains, suggesting that an epidemiological analysis based on antibody detection may not be accurate.

In order to understand whether the mixed *cag* genotypes that we observed for some of our patients were the result of multistrain *H. pylori* infections instead of the expression of true genomic modifications within the same strain (e.g., deletions, mutations, and recombinations), which are commonly described for this species (2, 16, 18, 36), we performed RAPD-PCR analysis of single colonies.

This technique showed that each patient carrying mixed colonies was infected with a single *H. pylori* strain. The data also indicated that small deletions, such as the ones that we observed, do not usually affect DNA fragment migration; the result is similar RAPD patterns (1). Identity among strains was confirmed by sequencing of the internal 294-bp fragment of the *glmM* gene, which has been shown to be a highly polymorphic region in all *H. pylori* strains (19). The association between the presence of single genes of the *cag* island and both serological expression and clinical expression of infection with *cagA*-positive *H. pylori* strains has been studied (34, 30). Our data suggest that humans harbor variable proportions of bacterial subtypes within a single infecting *H. pylori* strain and that the qualitative detection of a single gene as a marker for virulence probably does not completely reflect the in vivo situation. Therefore, a detailed genetic analysis may be required to fully elucidate the pathogenic role of *cagA*-positive *H. pylori* strains. In addition, an analysis of the *cag* island in a given patient at different times may clarify whether any observed changes in *cag* genotypes within an *H. pylori* strain during chronic infection are associated with specific histopathological modifications and different trends in disease progression.

In conclusion, our PCR-based study shows that highly heterogeneous *cag* genotypes can be detected in *cagA*-positive *H. pylori* strains isolated from single gastric biopsy specimens. The extension of such an analysis to patients with well-characterized gastric diseases may provide significant information on the relationship between *cag* genotypes and clinical outcomes of *H. pylori* infections.

**ACKNOWLEDGMENTS**

Special acknowledgment goes to M. J. Blaser, Department of Medicine, New York University, whose contribution to the genesis of this article was noteworthy. The article itself was based on experimental work carried out by M. Sozzi under the supervision of M. J. Blaser in his previous laboratory in Nashville, Tenn.

**REFERENCES**


