Composition and Gene Expression of the cag Pathogenicity Island in Helicobacter pylori Strains Isolated from Gastric Carcinoma and Gastritis Patients in Costa Rica

ALESSANDRA OCCHIALINI, ARMELLE MARAIS, MARIA URDACL, RAFAELA SIERRA, NUBIA MUÑOZ, ANTONELLO COVACCI AND FRANCIS MÉGRAUD

Laboratoire de Bactériologie, Université Victor Segalen Bordeaux 2 et Hôpital Pellegrin, 33076 Bordeaux, Laboratoire de Microbiologie, Ecole Nationale d’Ingénieurs des Travaux Agricoles, 33170 Grudignan, and Centre International de Résearch sur le Cancer, 69372 Lyon, France; University of Costa Rica, San José, Costa Rica; and IRIS, 53100 Siena, Italy

Received 7 August 2000/Returned for modification 25 September 2000/Accepted 20 November 2000

The composition and in vitro expression of the cag pathogenicity island genes in a group of Helicobacter pylori strains obtained from patients suffering from chronic gastritis-associated dyspepsia (n = 26) or gastric carcinoma (n = 17) were analyzed. No significant difference in the distribution of the 10 studied regions was found between the cases and the controls. Nine strains did not harbor any of the selected regions: eight (36.4%) isolated from patients with gastritis only and one (5.9%) from a patient with gastric carcinoma. No association was found between the number of repeated sequences at the 3' end of the cagA gene or the presence of tyrosine phosphorylation motifs and the clinical origin of the strains. The virB10 homolog gene was the sole gene studied to be significantly expressed more often in cancer strains than in gastritis strains (P = 0.03).

Since the first culture of Helicobacter pylori in 1982, chronic infection with this bacterium has been identified as the etiological agent of gastritis and duodenal ulcer (18, 21). Evidence for a causal relationship between chronic H. pylori infection and stomach cancer came first from epidemiological studies (10, 19), and, more recently, from animal models of carcinogenesis (12, 32). In 1994, the International Agency for Research on Cancer-World Health Organization (11) defined H. pylori infection as a group I carcinogen (definite carcinogen).

Despite the very high rate of H. pylori infection in some populations, the rate of gastric cancer is relatively low. Beside environmental factors and hereditary predisposition, as recently illustrated with interleukin-1β polymorphism (8), a possible explanation may be found in the different patterns of pathogenicity of H. pylori strains. The cag pathogenicity island (PAI) is one of the major H. pylori virulence factors found more frequently in patients suffering from severe gastroduodenal diseases, including peptic ulcers (5, 13, 16) and gastric adenocarcinomas (16, 26).

Recent studies have shown that the CagA protein, encoded by the cag PAI, is translocated into host cell by the cag PAI type IV secretion system, by which it becomes phosphorylated by an unknown cell kinase (4, 25, 27, 31), inducing an active reorganization of actin (31). The sites of possible phosphorylation of CagA have been identified (25). The CagA protein varies in size from 128 to ca. 140 kDa, and this variation is related to the presence of a variable number of repeated sequences in the 5' region of the gene (9, 35, 36).

The aim of this study was to compare the presence, the composition, and the expression of several cag PAI genes, including cagA, in Costa Rican strains obtained from patients suffering from gastric carcinoma (n = 17) and from chronic gastritis-associated dyspepsia (n = 26), as well as to determine the terminal sequence of the cagA gene and the putative tyrosine phosphorylation motifs and the clinical origin of the strains. The virB10 homolog gene was the sole gene studied to be significantly expressed more often in cancer strains than in gastritis strains (P = 0.03).

Molecular composition of the cag PAI among H. pylori strains. After culture of the 43 H. pylori isolates on selective medium (20), the DNA was extracted using the standard phenol-chloroform procedure (24). Four reference strains were included in the study: J99 (3), 26695 (33), NCTC 11638 (2, 5, and SS1 (15). The presence of cag PAI was analyzed by detecting 10 different regions (the 5' and 3' ends of cagA, cagC, virB10, cagFl, cagII, cagIl, cagIII, cagIV, cagV, cagVI, cagVII, cagVIII, and cagIV) by PCR amplification and dot blot hybridization as previously described (24). The presence of IS605 (tphA and tphB) was also evaluated (Table 1). An example of cagFl, cagII, cagIII, and IS605 detection by PCR is presented in Fig. 1A. As expected, a 1,803-bp fragment was revealed for the strains NCTC 11638, J99, and 26695, and no amplification was obtained in the control when DNA was replaced by water (lane NC). Among the 22 clinical strains presented in this figure, 15 yielded the expected fragment. In the case of IS605 amplification, a 1,486-bp amplicon was obtained in controls (26695 and NCTC 11638 strains), as well as in six strains among the 22 presented. As expected from the genome sequence, the IS605 fragment was not amplified from DNA of strain J99.

Negative results by PCR may be due either to an absence of the tested gene (true negative) or to a lack of primer annealing due to interstrain variation in the sequences targeted by the primers. For this reason, dot blot analyses were performed to confirm the absence of the selected genes by hybridization conducted under stringent conditions and as described elsewhere (24). Moreover, the specificity of the amplicons was also confirmed by a combination of Southern blot hybridization and PCR.
confirmed by this technique. Dot blot analysis were then performed on all strains with all of the probes. An example of detection of the cagL/cagM/cagN region and IS605 by hybridization is illustrated in Fig. 1B with the same strains as used for PCR detection (Fig. 1A). The same results were obtained for all of these strains except for strain SS1, which gave positive results by dot blot and negative results by PCR. With the IS605 probe, two contradictory results were obtained with DNA from strains in the lines 5 and 14. In summary, the correlation between the results obtained by PCR and those obtained by
dot blot were 100% for cagC, virB4/cagF, cagS/virB7, virB9, virB10, virB11, and virD4; 97.9% for cagA D008-R008, cagA Ys-A6, and cagL/cagM/cagN; and 95.7% for cagP/cagR and IS605. A strain was defined as lacking a given region if the results obtained both by PCR and by hybridization were negative. In the case of a given region detected by hybridization and not by PCR, the strain was considered positive for this region. The reverse case (positive PCR and negative hybridization) was never observed.

According to the rule established above, the results were analyzed first globally and then according to the associated disease. Comparison of open reading frame (ORF) patterns obtained from the different strains was performed using the unweighted-pair-group method with averages (30). The distribution of the selected cag PAI regions among strains is presented in Fig. 2. Each strain was assigned to a pattern based on the presence or absence of the selected region. The strains could be classified into four groups. The largest group included
32 strains (74.4%) which were positive for all of the cag PAI regions studied. One strain (2.3%) exhibited a deletion of the cagA gene while the other regions were present, and one cagA-positive strain (2.3%) lacked all other regions, indicating that an association between cagA and cag PAI, while very strong, is not absolute. The fourth group included none of the cag PAI regions studied. The SS1 reference strain was found to be positive for all of the regions studied (type 1 strain).

In order to confirm the complete absence of cag PAI in the nine reference strains, an amplification was performed using two primers (Table 1) located in the two genes flanking the cag PAI in the reference strains (HP519 and HP549; Fig. 2). As expected from the literature (1, 14), a single 324-bp amplicon was obtained from each of the nine cag PAI-negative strains. The nucleotide sequences of the so-called "empty site" were determined and deposited in GenBank under the accession numbers AF289390 to AF289398. The nine strains considered to be cag PAI negative based on PCR and dot blot results were, therefore, confirmed to be truly cag PAI negative.

The results are also presented according to the associated disease (Fig. 2). The proportion of each expression profile was compared between strains isolated from patients with gastric carcinoma and patients with chronic gastritis-associated dyspepsia using the chi-square test. Strains with no cag PAI regions were found more frequently among gastritis strains than among carcinoma strains, but the difference was not statistically significant (30.8 versus 5.9%, respectively, \( P = 0.11 \)). At the same time, strains containing all of the tested cag PAI regions were found more frequently among the gastric carcinoma strains (82.3%) than among the gastritis strains (69.2%). However, this difference was still not significant (\( P = 0.28 \)). Profiles II and III were not considered in this analysis because they did not contain enough strains (\( n = 1 \) in each group).

Concerning the results obtained for IS603, 11 strains were found to be positive (25.6%): 8 in the gastritis group (30.8%) and 3 in the carcinoma group (17.6%). Although IS603 seemed to be more frequently found in strains isolated from patients with gastritis than in those isolated from patients with gastric cancer, this difference was not significant (\( P = 0.54 \)).

Analysis of the cagA terminal sequences. The sequence of two fragments of the cagA gene (at the 5' and 3' ends) was analyzed for the 33 cagA-positive strains. The 5' end amplicons were identical in size (394 bp), and their nucleotide sequences exhibited an overall similarity of 96.8% (accession numbers AF289399 to AF289413 for the gastric carcinoma strains and accession numbers AF289432 to AF289431 for the gastritis strains). In contrast, an interstrain size variation of the Ys-A6 fragment of the cagA gene (3' end) was observed (from 570 to 680 bp). Each amplicon was therefore submitted to sequencing (accession numbers AF289432 to AF289446 for the carcinoma strains and accession numbers AF289447 to AF289464 for the gastritis strains). The deduced amino acid sequences obtained were analyzed according to the method of Evans et al. (9). Eleven different sequence types were obtained, indicating a large variability. Regarding the number of repeated sequences, 23 strains contained only one repeat (69.7%) and 10 strains contained two repeats (30.3%). No example of three repeats was found. The number of repeats was analyzed according to the origin of the strains. Although two repeated sequences were found more frequently in strains from patients with gastric cancer (40 versus 22.2%), this difference was not significant (\( P = 0.46 \)).

Analysis of TPMs of the CagA protein. To determine whether the CagA protein of the clinical strains was phosphorylated, putative CagA TPMs were sought in their sequences. Three sites of phosphorylation have been recently identified by Odenbreit et al. (25): site "A" (KFGDQRY) at amino acid (aa) 122, site "B" (KNSTPEY) at aa 899, and site "C" (KLKD STKY) at aa 1029. Thus, the sequences of the two fragments of the cagA gene which may contain these three sites (TPM A in the A1-A2 fragment and TPMs B and C in the Ys-A6 fragment) were analyzed. The cagA sequences of all 33 Costa Rican strains have the TPM KFGDQRY in position A. Nineteen strains (12 of 15 patients with gastritis and 7 of 18 patients with gastric carcinoma) presented the TPM KNST/GEPY in position B, the other strains having a variable sequence which did not correspond to a phosphorylation motif (KNEPY, 12 strains; ENSAEPY, two strains). No strain had a TPM in position C. Thus, globally, the majority of the strains (57.6%)
was similar to the 26695 strain which contains TPMs A and B. The remaining strains contained only the TPM A site, as did the strain NCTC 11638. In conclusion, the CagA protein from all of the Costa Rican strains tested contained at least one TPM; however, no association was found between the number of TPMs and the type of disease from which the strain originated. Indeed, 66.7% of the strains from patients with chronic gastritis only presented two TPMs (A and B) in the CagA protein versus 46.7% of the gastric carcinoma strains.

**Expression of the cag PAI genes in H. pylori strains.** To investigate whether the ORFs present in the cag PAI regions were expressed in vitro, transcription of 15 genes of cagPAI (cagA, cagC, virB4, cagF, cagL, cagN, cagM, cagP, cagQ, cagS, virB7, virB9, virB10, virB11, and virD4) was investigated by reverse transcription using a random primer mixture followed by a PCR (Enhanced Avian RT-PCR Kit; Sigma Aldrich, St. Quentin Fallavier, France) with specific primers (Table 1). RNA from an in vitro H. pylori culture was isolated according to the previously described method (24). Among the 15 genes, only one was found to be expressed in all of the strains under the conditions tested (cagP). The expression of five genes (cagC, cagP, cagQ, virB11, and virD4) was never detected. The other genes (n = 9) were differentially transcribed among the cag PAI-positive strains: cagL in only 1 strain (3%), cagS and virB4 in 4 strains (12.1%), cagN and virB9 in 5 strains (15.1%), virB7 in 7 strains (21.2%), cagM in 11 strains (33.3%), virB10 in 21 strains (63.6%), and cagD in 29 strains (87.9%). No correlation could be drawn from this analysis. However, two pairs of genes seemed to be linked: cagP and cagQ on the one hand and virB11 and virD4 on the other hand. Indeed, when the cagP gene was not expressed, cagQ was not expressed either. The same observation was noted for the virB11 and virD4 genes. This fact, together with their relative locations (Fig. 2), suggested that these pairs of genes belong to the same transcriptional unit and may be coregulated. The expression status of genes was analyzed according to the disease of the patients from whom the strains were isolated (Table 2). The only cag PAI gene for which a statistically significant difference was found (P = 0.03) was the virB10 gene, whose transcript was detected in 21 strains (48.8%), among which 9 were isolated from gastritis patients (34.6%) and 12 were isolated from carcinoma patients (70.6%). A trend, while not statistically significant, was also observed for cagN (P = 0.07) and cagM (P = 0.08) expression.

The first question posed in this study was whether the development of gastric carcinoma in Costa Rica, known to display a high incidence of this pathology (29), is associated with the presence of cag PAI in infecting H. pylori strains. The results of our study showed that, in this country, most of the H. pylori strains are positive for the cag PAI (74.4%). Nevertheless, while no statistically significant difference was found, eight of the nine strains with no cag PAI were isolated from patients with gastritis but only one was isolated from a gastric carcinoma patient, indicating a tendency (P = 0.11). There are controversial reports concerning the association of the cag PAI-positive H. pylori and more severe gastric diseases such as peptic ulcer disease and gastric carcinoma. Most of the reports from Western countries (Europe and the United States) (5, 13) found a positive association, while no association was found in the Far East (23). This fact can be explained by the high background level of cag PAI-positive H. pylori strains in far eastern countries which prohibits making conclusions based on cross-sectional or case-control studies. It is also possible that, while all H. pylori strains lead to cancer if the patient lives long enough, the process may develop more quickly in certain individuals due to environmental and individual genetic factors. However, concerning the absence of the cag PAI in strains isolated from patients with more severe disease (one case in the present study), it cannot be excluded that the strains studied, which were present at the time of carcinoma diagnosis, were not necessarily the carcinogenic strains, since gastric carcinoma evolves over several decades. Indeed, the possibility of harboring several strains successively, especially in a country with a high prevalence of infection, does exist (22, 23, 28). Despite the fact that cag PAI is present in most of the strains, it may be that the composition of the cag PAI genes or the regulation of their expression varies and influences the associated pathogenicity.

Therefore, in the present study, the composition of the cag PAI was analyzed in depth. The regions sought were chosen based on their location on the island as well as on the putative role of the ORF's encoded (2, 4, 5, 25, 27, 31, 34). Our results did not show an association between a particular region and the disease status: in 32 H. pylori clinical strains, 10 different regions of the cag PAI corresponding to 16 ORFs were all present, and therefore the strains could be categorized as type I according to Censini et al. (5). Interestingly, in one strain, the cag4 gene itself was absent, whereas the other regions of the cag PAI were present, indicating that the use of this gene as the sole marker for the detection of the cag PAI may sometimes lead to erroneous results. Furthermore, in another strain, isolated from a gastric carcinoma patient, the cag4 gene was the only part of the cag PAI present, leading us to the question of what are the pathogenic properties of this strain because the

**TABLE 2. Distribution of expressed genes of the cag PAI between gastritis and gastric carcinoma strains**

<table>
<thead>
<tr>
<th>Gene</th>
<th>No. of genes</th>
<th>% Total</th>
<th>Gastritis Patient isolates</th>
<th>Gastric carcinoma Patient isolates</th>
<th>P*</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>cagA</em></td>
<td>29</td>
<td>67.4</td>
<td>16</td>
<td>61.5</td>
<td>13</td>
</tr>
<tr>
<td><em>cagC</em></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>virB4</em></td>
<td>4</td>
<td>9.3</td>
<td>3</td>
<td>11.5</td>
<td>1</td>
</tr>
<tr>
<td><em>cagF</em></td>
<td>34</td>
<td>79.1</td>
<td>18</td>
<td>69.2</td>
<td>12</td>
</tr>
<tr>
<td><em>cagL</em></td>
<td>1</td>
<td>2.3</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><em>cagM</em></td>
<td>6</td>
<td>11.6</td>
<td>1</td>
<td>3.8</td>
<td>4</td>
</tr>
<tr>
<td><em>cagN</em></td>
<td>11</td>
<td>25.6</td>
<td>4</td>
<td>16.4</td>
<td>7</td>
</tr>
<tr>
<td><em>cagP</em></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>cagQ</em></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>virB7</em></td>
<td>4</td>
<td>9.3</td>
<td>4</td>
<td>15.4</td>
<td>0</td>
</tr>
<tr>
<td><em>virB9</em></td>
<td>7</td>
<td>16.3</td>
<td>4</td>
<td>15.4</td>
<td>3</td>
</tr>
<tr>
<td><em>virB10</em></td>
<td>5</td>
<td>11.6</td>
<td>5</td>
<td>18.2</td>
<td>0</td>
</tr>
<tr>
<td><em>virB11</em></td>
<td>21</td>
<td>48.8</td>
<td>9</td>
<td>34.6</td>
<td>12</td>
</tr>
<tr>
<td><em>virD4</em></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>cagD</em></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

* a, Genes expressed in the J99 and 26695 strains; t, genes not expressed in the J99 and 26695 strains; g, gene expressed in the J99 strain but not in the 26695 strain.

* Data obtained by using Fisher's exact test.
type IV secretion system which delivers the CagA protein into the epithelial cells is encoded by genes which are absent.

The variation of the CagA protein size has been related to the presence of a variable number of repeated sequences in the 3' end of the gene (9). The presence of these multiple repeats in the COOH terminus of the protein was found to be associated with the development of gastric carcinoma in Japanese (36), U.S., and Colombian *H. pylori* strains, but not in Korean strains (35). In our study on Costa Rican strains, it was not possible to confirm this association.

In a recent study, Odenbreit et al. (25) studied the tyrosine phosphorylation of the CagA protein in several *H. pylori* reference strains and established that the translocation of the CagA protein, as well as the presence of at least one of three TPMs in the CagA protein, is necessary for CagA phosphorylation; therefore, this characteristic may be relevant in terms of disease association. Thus, such motifs were sought in the deduced CagA protein sequence in the strains included in our study. The results show that half of them contained only one phosphorylation site (at position 122), such as strain NCTC 11638, whose CagA protein is weakly phosphorylated (25). The variation of the CagA protein size has been related to the presence of any of the genes of the PAI, which are highly complementary in their function since they have been transmitted as a whole throughout evolutionary events subsequent to their incorporation.

Comparison of the genome of strains J99 and 26695 (3, 7, 17, 33) showed the presence of other regions whose role in pathogenicity remains to be explored. It is likely that virulence results from a combination of various factors, such as the presence of the cag PAI together with that of such regions as those encoding, cytotoxin production, and adhesin expression, rather than the presence of a unique virulence factor (6). Moreover, a study of the transcriptome under more physiological conditions and of the proteome may, in the future, provide information on additional elements for what remains the most important pathogenic determinant of *H. pylori*.

We gratefully acknowledge the financial support of the Association pour la Recherche sur le Cancer (ARC), Villejuif (no. 9313), France, and the Conseil Regional d’Aquitaine, Bordeaux, France. A. Ochiali is a fellowship recipient from IRMAD, Paris, France.

We gratefully acknowledge Fernando Garcia (San José, Costa Rica) for providing gastric biopsy samples. We also thank Kathryn Mayo (Laboratoire de Bacteriologie-Université Bordeaux 2) for English corrections.

REFERENCES


10. Huang, J. Q., S. Sridhar, Y. Chen, and P. H. Hunt. 1998. Meta-analysis of a study of the transcriptome under more physiological conditions and of the proteome may, in the future, provide information on additional elements for what remains the most important pathogenic determinant of *H. pylori*.


