cagA-Positive Helicobacter pylori Populations in China and The Netherlands Are Distinct

ARIE VAN DER ENDE,†* ZHI-JUN PAN,†,‡ ALDERT BART,‡ RENÉ W. M. VAN DER HULST,‡ MONIQUE FELLER,‡ SHU-DONG XIAO,‡ GUIDO N. J. TYTGAT,‡ and JACOB DANKERT†

Departments of Medical Microbiology† and Gastroenterology,‡ Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands, and Shanghai Institute of Digestive Disease, Shanghai Second Medical University, Shanghai, People’s Republic of China

Received 26 November 1997/Returned for modification 23 January 1998/Accepted 11 February 1998

The aim of this research was to study whether and to what extent Chinese cagA-positive Helicobacter pylori isolates differ from those in The Netherlands. Analysis of random amplified polymorphic DNA (RAPD)-PCR-assessed DNA fingerprints of chromosomal DNA of 24 cagA-positive H. pylori isolates from Dutch (n = 12) and Chinese (n = 10) patients yielded the absence of clustering. Based on comparison of the sequence of a 243-nucleotide part of cagA, the Dutch (group I) and Chinese (group II) H. pylori isolates formed two separate branches with high confidence limits in the phylogenetic tree. These two clusters were not observed when the sequence of a 240-bp part of glmM was used in the comparison. The number of nonsynonymous substitutions was much higher in cagA than in glmM, indicating positive selection. The average levels of divergence of cagA at the nucleotide and protein levels between group I and II isolates were found to be high, 13.3 and 17.9%, respectively. Possibly, the pathogenicity island (PAI) that has been integrated into the chromosome of the ancestor of H. pylori now circulating in China contains a different cagA than the PAI that has been integrated into the chromosome of the ancestor of H. pylori now circulating in The Netherlands. We conclude that in China and The Netherlands, two distinct cagA-positive H. pylori populations are circulating.

Helicobacter pylori infection in humans is one of the most widespread infections today, and its cure prevents peptic ulcer recurrence (26, 35). Besides asymptomatic gastritis and peptic ulcer disease (PUD), H. pylori infection is strongly associated with gastric cancer, gastric mucosa-associated lymphoid tissue (MALT), and adenocarcinoma of the stomach (3, 9, 24).

The heterogeneity of the clinical outcome of H. pylori infection may be related either to differences among the hosts or to differences in virulence among H. pylori strains. The latter assumption is supported by the finding that the product of the cagA gene is also related to an increased risk to develop atrophic gastritis, intestinal metaplasia (16, 35), or gastric cancer (25).

Recently, the complete genome sequence of H. pylori has become available (30). A 40-kb region of the H. pylori chromosome containing cagA was sequenced earlier by Censini et al. (4). This locus, comprising at least 40 genes, has a GC content different from that of the rest of the chromosome, forms a so-called pathogenicity island (PAI), and is assumed to have been integrated into the H. pylori chromosome only recently (4, 6). The proteins encoded by the PAI genes possess features similar to those of bacterial type II, type III, and most notably type IV secretion systems. It was hypothesized that such proteins may function to export macromolecules that may be involved in the H. pylori-host cell interaction (6).

China is one of the countries with a high prevalence of H. pylori infection and a high incidence of gastroduodenal diseases (39). The prevalence of H. pylori infection increases with age to about 70% of the people over 30 years old (22, 33, 39). The prevalence of cagA-positive H. pylori populations in Chinese patients with PUD and FD is almost universally high (21). Data obtained from this recent study further suggested that H. pylori genotypes distinct from those present in Western Europe may circulate in China.

The aim of this study is to investigate this hypothesis by comparison of the random amplified polymorphic DNA (RAPD)-PCR-assessed genotype of 24 randomly collected cagA-positive H. pylori isolates from 12 Dutch (14 isolates) and 10 Chinese patients. We used four different primers in each of four amplifications of H. pylori genomic DNA. In addition, part of cagA and glmM of the H. pylori isolates was sequenced. Sequences were analyzed for similarity by a computer-based program by using the neighbor-joining algorithm of Saitou and Nei (27).
silent mutations, i.e., without amino acid substitutions; the estimated phylogenies. The proportions of synonymous substitutions (or resampling analyses (1,000 replicates) were performed to assign confidence limits. The neighbor-joining algorithm of Saitou and Nei (27), with the Kimura two-parameter distance measures (15) as implemented in the MEGA program. Bootstrap analysis (1,000 replicates) demonstrated a high confidence (that is, identical branch points occurred in all bootstrap replicates) of the difference between the two main groups comprising the 12 H. pylori isolates from Dutch patients and the group of H. pylori isolates from 10 Chinese patients (Table 1). Clustering analysis revealed two main groups comprising the H. pylori strains from all Dutch patients (group I) and the H. pylori strains from all Chinese patients (group II) (Fig. 2).

RESULTS

RAPD-PCR of H. pylori isolates from Dutch and Chinese patients. Assessment by RAPD-PCR of chromosomal DNA of 22 cagA-positive H. pylori isolates, 12 from 12 Dutch patients and 10 from 10 Chinese patients, showed that each isolate had a unique RAPD pattern. The initial isolate 79A and isolate 79J cultured from sequential biopsy specimens taken from the same patient were identical. Likewise, the initial isolate 161A was identical to isolate 161L. Clustering analysis did not reveal any clusters of isolates on the basis of either clinical manifestations, or origin of geographic area.

Comparison of cagA sequences of H. pylori isolates from Dutch and Chinese patients. Comparison of a 243-bp part of the cagA gene sequence region between nucleotides 1537 and 1780 (notation according to Covacci et al. [5]) from the 24 clinical H. pylori isolates showed 21 alleles, with mutations at 67 possible positions (Fig. 1). Both sequentially recovered H. pylori isolates from two Dutch patients (strains 161A and 161L; strains 79A and 79J) and two H. pylori isolates from two Chinese patients (strain R27 and R30) had identical cagA sequences. In Fig. 2, the polymorphic site in the cagA region between nucleotides 1537 and 1780 of cagA is shown. The total number of 67 nucleotide substitutions resulted in 22 possible amino acid substitutions. The \( \delta_d \) and \( \delta_n \) values were similar in the 12 H. pylori isolates from 12 Dutch patients and the group of H. pylori isolates from 10 Chinese patients (Table 1).

Clustering analysis revealed two main groups comprising the H. pylori strains from all Dutch patients (group I) and the H. pylori strains from all Chinese patients (group II) (Fig. 2). Bootstrap analysis (1,000 replicates) demonstrated a high confidence (that is, identical branch points occurred in all bootstrap replicates) of the difference between the two main groups comprising the H. pylori isolates from Dutch and Chinese patients. The cagA sequence of group I strains (excluding strains 79J and 161L) showed 3.9% average divergence at the nucleotide level and 6.2% average divergence at the amino acid level.
level. The levels of average divergence of the \textit{cagA} sequence among the group II strains were similar, 4.8 and 5.8% at the nucleotide and amino acid levels, respectively. Evidently, the difference in the \textit{cagA} sequence was more extensive (two to three times larger) when the strains of the two groups were compared with each other (Table 2).

Comparison of \textit{glmM} sequences of \textit{H. pylori} isolates from Dutch and Chinese patients.

To compare sequence heterogeneity of \textit{cagA}, located on the PAI, with that of a gene outside the PAI, part of \textit{glmM} (formerly called \textit{ureC} [18]) was sequenced. Of the 24 \textit{H. pylori} isolates, the same 240-bp part of \textit{glmM} was sequenced as described by Kansau et al. (14). Twenty-two alleles with mutations at 32 possible positions were found (Fig. 3). The two sequentially recovered \textit{H. pylori} isolates from each of the two Dutch patients (strains 161A and 161L; strains 79A and 79J) were identical. The total number of 32 nucleotide substitutions resulted in only 3 possible amino acid substitutions. The $d_S/d_N$ ratio ($d_S/d_N = 0.1103/0.0068 = 16.2$) was much higher in \textit{glmM} than in \textit{cagA}. In contrast to the \textit{cagA} sequence, clustering analysis of \textit{glmM} did not result in any robust cluster formation.

DISCUSSION

Data obtained from a recent report suggested that \textit{H. pylori} genotypes circulating in China are distinct from those in Western Europe due to allelic variation in \textit{cagA} (21). The aim of our study was to provide evidence that Chinese patients and Dutch patients are colonized with distinct \textit{cagA}-positive \textit{H. pylori} strains.

RAPD-PCR analysis of 14 \textit{H. pylori} isolates from 12 Dutch patients and 10 from 10 Chinese patients demonstrated a high level of genetic diversity among the 24 strains. In previous studies using this technique, it was shown that \textit{H. pylori} comprises a genetically highly heterogeneous group, with patient-to-patient variation (1). In addition, patients can harbor a heterogeneous \textit{H. pylori} population (12, 31, 33, 37). On the basis of the RAPD-PCR patterns, the 24 \textit{H. pylori} strains could be clustered according to neither the various clinical entities nor the geographic origin of the patient. Results obtained with multilocus enzyme electrophoresis suggested clustering of 23 \textit{H. pylori} isolates into four clusters (11). The authors concluded that the genetic diversity in \textit{H. pylori} may be sufficient to classify \textit{H. pylori} strains into four or more cryptic species.

**TABLE 1.** Proportion (Jukes-Cantor corrected) of synonymous and nonsynonymous substitutions per site among the 243-nucleotide sequenced part of \textit{cagA} between nucleotides 1573 and 1780 (notation according to Covacci et al. [5]) of 25 \textit{H. pylori} isolates

<table>
<thead>
<tr>
<th>Source of isolates</th>
<th>No. of isolates</th>
<th>Proportion of substitutions (mean ± SE)</th>
<th>$d_S$</th>
<th>$d_N$</th>
<th>$d_S/d_N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch patients</td>
<td>12</td>
<td>0.102 ± 0.071 0.027 ± 0.021</td>
<td>3.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese patients</td>
<td>10</td>
<td>0.140 ± 0.053 0.025 ± 0.010</td>
<td>5.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ The \textit{cagA} sequences of \textit{H. pylori} isolate 79J (identical to 79A but isolated from the same patient 6 years later) and 161L (identical to 161A but isolated from the same patient 4 years later) were not taken into account.

**TABLE 2.** Sequence diversity among a part of 243 nucleotides of the \textit{cagA} region between 1573 and 1780 (notation according to Covacci et al. [5]) of \textit{H. pylori} isolates from 12 Dutch and 10 Chinese patients

<table>
<thead>
<tr>
<th>Source of isolates</th>
<th>No. of isolates</th>
<th>% Differences (mean ± SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Nucleotide</td>
</tr>
<tr>
<td>Group I</td>
<td>12$^a$</td>
<td>3.9 ± 2.5</td>
</tr>
<tr>
<td>Group II</td>
<td>10</td>
<td>4.8 ± 1.6</td>
</tr>
<tr>
<td>Groups I and II</td>
<td>22</td>
<td>13.3 ± 1.4</td>
</tr>
</tbody>
</table>

$^a$ Group I, Dutch patients; group II, Chinese patients.

$^a$ The \textit{cagA} sequences of \textit{H. pylori} isolate 79J (identical to 79A but isolated from the same patient 6 years later) and 161L (identical to 161A but isolated from the same patient 4 years later) were not taken into account.
However, the similarity of strains within a cluster was rather low and varied between 30 and 70%. In addition, a similar analysis revealed that no clustering among 74 H. pylori isolates occurred, and a very high mean genetic diversity was found (10).

The phylogenetic tree based on the cagA sequences showed a robust division between H. pylori isolates from Dutch patients (group I) and H. pylori isolates from Chinese patients (group II). In addition, the cagA sequences previously published by Covacci et al. (5) and Tummuru et al. (31) fit into the branch comprising the Dutch H. pylori strains, while the cagA sequence of the H. pylori isolate from a Japanese patient fit into the branch comprising the Chinese H. pylori strains, without altering the robust division between the Western and Chinese H. pylori isolates in the tree (data not shown). The percentage difference between cagA of groups I and II is larger than the differences between cagA of H. pylori isolates within its appropriate group (Table 2). In contrast both phylogenetic trees based on RAPD patterns and on glmM sequences showed the overall genetic variation without any robust clustering. Therefore, we assume that in the past, the PAI that has been integrated into the genome of the ancestor of H. pylori now circulating in China contained a different cagA than the PAI that has been integrated into the genome of the ancestor of H. pylori now circulating in The Netherlands. Most likely, the PAI that has been integrated into the chromosome of the ancestor of H. pylori now circulating in China contained a different cagA than the PAI that has been integrated into the chromosome of the ancestor of H. pylori now circulating in The Netherlands.

ACKNOWLEDGMENTS

This work was supported by grants from the Dutch Ministry of Education and Science, the Royal Dutch Academy of Science, and the Chinese Ministry of Public Health (1994).

REFERENCES


17. Kumar, S., K. Tamura, and M. Nei. 1993. MEGA, Molecular Evolutionary Genetics Analysis, v 1.01. The Pennsylvania State University, University Park. (Distributed by the authors.)

Editor: J. T. Barbieri
cagA-Positive *Helicobacter pylori* Populations in China and The Netherlands Are Distinct

ARIE VAN DER ENDE, ZHI-JUN PAN, ALDERT BART, RENÉ W. M. VAN DER HULST, MONIQUE FELLER, SHU-DONG XIAO, GUIDO N. J. TYTGAT, AND JACOB DANKERT

Departments of Medical Microbiology and Gastroenterology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands, and Shanghai Institute of Digestive Disease, Shanghai Second Medical University, Shanghai, People’s Republic of China

Volume 66, no. 5, p. 1822–1826, 1998. Page 1823, column 2: The following paragraph should be inserted at the end of Materials and Methods:

**Nucleotide sequence accession numbers.** The nucleotide sequences of *cagA* and *glmM* have been deposited in the GenBank database under accession no. AJ252963 to AJ252986 and AJ252987 to AJ253010, respectively.

---

Genomic Analysis Reveals Variation between *Mycobacterium tuberculosis* H37Rv and the Attenuated *M. tuberculosis* H37Ra Strain

ROLAND BROSCH, WOLFGANG J. PHILLIPP, EVANGELOS STAVROPOULOS, M. JOSEPH COLSTON, STEWART T. COLE, AND STEPHEN V. GORDON

Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 75724 Paris Cedex 15, France, and Division of Mycobacterial Research, National Institute for Medical Research, London NW7 1AA, United Kingdom

Volume 67, no. 11, p. 5768–5774, 1999. Page 5773, column 2, “Acknowledgments”: Because of an administrative error, the second paragraph was incomplete and should read as follows:

“Financial support for this work was provided by the Wellcome Trust, the Biomed Program of the European Community (grant BMH4/CT97/2277), the Institut Pasteur, and l’Association Française Raoul Follereau. S. V. Gordon was the recipient of a Wellcome Trust International Travelling fellowship.”