The cag Pathogenicity Island of Helicobacter pylori Is Disrupted in the Majority of Patient Isolates from Different Human Populations

Farhana Kauser,† Aleem A. Khan,‡ M. Abid Hussain,† Ian M. Carroll,§ Naheed Ahmad,¶ Santosh Tiwari,¶ Yogesh Shouche,¶ Bimal Das,¶ Mahfooz Alam,¶ S. Mahaboob Ali,¶ C. M. Habibullah,¶ Rafaela Sierra,¶ Francis Megraud,¶ Leonardo A. Sechi,¶ and Niyaz Ahmed†∗

Pathogen Evolution Group, Centre for DNA Fingerprinting and Diagnostics,† and Centre for Liver Research and Diagnostics, Deccan College of Medical Sciences and Allied Hospitals;‡ Hyderabad, Department of Botany, Patna University, Patna,¶ National Centre for Cell Science, Pune,§ and Department of Microbiology, All India Institute of Medical Sciences, New Delhi,¶ India; The Moyne Institute of Preventive Medicine, Trinity College, Dublin, Ireland;¶ Instituto de Investigaciones en Salud, Universidad de Costa Rica, San José, Costa Rica;§ Laboratoire de Bacteriologie—C.H.U. Pellegrin, Bordeaux, France; and Department of Biomedical Sciences, University of Sassari, Sassari, Italy

Received 7 May 2004/Returned for modification 1 June 2004/Accepted 25 June 2004

The cag pathogenicity island (cag-PAI) is one of the major virulence determinants of Helicobacter pylori. The chromosomal integrity of this island or the lack thereof is speculated to play an important role in the progress of the gastroduodenal pathology caused by H. pylori. We determined the integrity of the cag-PAI by using specific flanking and internally anchored PCR primers to know the biogeographical distribution of strains carrying fully integral cag-PAI with proinflammatory behavior in vivo. Genotypes based on eight selected loci were studied in 335 isolates obtained from eight different geographic regions. The cag-PAI appeared to be disrupted in the majority of patient isolates throughout the world. Conservation of cag-PAI was highest in Japanese isolates (57.1%). However, only 18.6% of the Peruvian and 12% of the Indian isolates carried an intact cag-PAI. The integrity of cag-PAI in European and African strains was minimal. All 10 strains from Costa Rica had rearrangements. Overall, a majority of the strains of East Asian ancestry were found to have intact cag-PAI compared to strains of other descent. We also found that the cagE and cagT genes were less often rearranged (18%) than the cagA gene (27%). We attempted to relate cag-PAI rearrangement patterns to disease outcome. Deletion frequencies of cagA, cagE, and cagT genes were higher in benign cases than in isolates from severe ulcers and gastric cancer. Conversely, the cagA promoter and the left end of the cag-PAI were frequently rearranged or deleted in isolates linked to severe pathology. Analysis of the cag-PAI genotypes with a different biogeoclimatic history will contribute to our understanding of the pathogen-host interaction in health and disease.

Infection with Helicobacter pylori has plausible associations with a variety of clinical outcomes. This includes chronic gastritis, duodenal ulcer, gastric cancer, lymphomas and seemingly paradoxical relationships with gastroesophageal reflux disease and adenocarcinoma of the gastric cardia (12, 13, 15, 29, 38). H. pylori inhabits at least 50% of the world's population, with the highest incidence recorded in industrially underdeveloped areas, including Asia (70 to 80%) and Africa (70 to 90%), and a waning occurrence in developed areas such as North America (30 to 40%), South America (80 to 90%), and Europe (30 to 70%) (33). Prevalence rates vary between populations and between groups within the same population (49).

Variation in the clinical outcome of H. pylori induced pathology is multifactorial, involving a complex interplay between the host immune responses, pathogen virulence factors, and niche characteristics. Many putative virulence factors have been identified in H. pylori that contribute to its pathogenesis. Most important among them are the 128-kDa cytotoxin-associated gene-encoded antigen gene (cagA) (18, 45) and the vacuolating cytotoxin antigen gene (vacA) (4, 43). The cagA gene is present downstream of an ~40-kb cluster of virulence genes known as the cag pathogenicity island (cag-PAI) located at the 3' end of the glutamate racemase gene. In some strains, the presence of an insertion element IS605 may disrupt an otherwise-uninterrupted unit into two regions, cag I and cag II, consisting of at least 14 and 16 open reading frames (ORFs), respectively (11). These genes encode a type IV secretion system that forms a syringe-like structure to translocate the immunodominant cagA protein into the gastric epithelial cells. cag-PAI has also been implicated in the induction of interleukin-8 (IL-8) in cultured gastric cells. As shown by the systematic mutagenesis studies of individual genes encoding the cag-PAI, there are at least 17 of 27 genes that are found to be essential for translocation of cagA into host cells (syringe-like function), and 14 were necessary for H. pylori to fully induce
transcription of IL-8 (17). This property contributes to the proinflammatory power of the strains and thus also to its virulence capability. Nevertheless, the induction of IL-8 is not exclusively linked to cag-PAI. Exposure of gastric epithelial cells to cag-PAI-positive H. pylori can activate the proto-oncogenes c-fos and c-jun, a crucial step in the development of H. pylori-related neoplasia (30). An intact cag-PAI has therefore been thought to contribute toward full proinflammatory power of H. pylori. However, the island has proved to be prone to disruption due to various genetic rearrangements occurring within and outside the constituent genes. Further still, the impact of intactness or rearrangement of the island or the constituent genes on the progression of gastroduodenal pathology has been debated. It has been argued that abolition of cagE gene results in a considerable reduction of IL-8 production (2, 11, 14, 36, 38, 51), whereas strains capable of eliciting an IL-8 response irrespective of an intact cag-PAI have been described (23, 28). Some strains causing nonulcer dyspepsia have regions of the promoter of cagA gene deleted (28), although this gene has long been regarded as a marker for the functionality of the cag-PAI. However, due to rearrangements often inhibiting cagA, it does not seem to be a reliable indicator of the virulence spectrum of strains (7). Other members of the cag-PAI have also been scrutinized for their involvement in virulence characteristics. The cagE gene served as a better marker of an intact cag-PAI in Japanese strains (22, 28) and French isolates (5).

The cagG gene was shown to be a better indicator for the presence of an intact cag-PAI than the cagE gene in French isolates (41). Deletion of cagE, cagT, cagA, cagG, and cagM genes was reported in all cases of chronic gastritis, gastric ulcer and gastric cancer, indicating that the pathogenicity of H. pylori may not be determined by cag-PAI genes alone in such cases (24). In a series of patients from Taiwan, the presence of cagA, cagC, cagE, cagF, cagN, and cagT genes in the PAI showed no relationship to the type of disease and/or the histological features present in the patients (46). The involvement of these genes and others in eliciting a strong immune response has also been contested (45). Rearrangements in cag-PAI in H. pylori isolates in different countries have been reported in isolation (5, 22, 23, 26, 28, 37, 40, 47). However, there are no comprehensive data available on the abundance of intact versus rearranged cag-PAI in strains inhabiting geographically and culturally distinct patient populations on a global scale. Such data on geographical distribution of cag-PAI rearrangement patterns may provide a window into the trinity of bacterial virulence, host genetic predisposition, and niche characteristics in a better way. These concerns, as well as our continued efforts (1, 9, 10, 21) to study the complexities associated with the infection dynamics and evolutionary genetics of this interesting pathogen with regard to host-pathogen interactions at all levels, led to the inception of the present study. We have used here simple and nested PCR-based genotyping approaches to demonstrate for the first time with a large number of strains that the cag-PAI is not perpetually intact in majority of H. pylori in the world.

MATERIALS AND METHODS

Patients and bacterial isolates. cag-PAI analyses were performed on a total of 335 H. pylori DNAs obtained from patient isolates from eight different countries: Spain, Peru, South Africa, England, Ireland, Japan, Costa Rica, and India. These isolates were recovered from unrelated patients undergoing upper gastrointestinal endoscopy after informed consent was obtained. Bacterial isolates were cultured from antral biopsies as described previously (1). Genomic DNA samples were prepared, quantified, and preserved according to the methods elsewhere (1). Most of the strains we used were described in other studies, and their microbiological and clinical characteristics have been elaborated elsewhere (1, 9, 25, 35, 48). To remove biases, isolates were picked randomly based merely on geographic representation. Classification and tabulation based on clinical correlation was attempted only after the experiments were completed. Isolates including genomic DNA preparations belonging to patients from Peru, Spain, Japan, England, and South Africa were a gift from Asish Mukhopadhyay and Douglas E. Berg (Washington University, St. Louis, Mo.). Genomic DNAs from Irish H. pylori isolates were obtained from single antral gastric biopsies from patients experiencing peptic ulcer disease at the Meath-Adelaide and St. James’s Hospitals in Dublin, Ireland.

cag-PAI PCR analysis. PCR analyses were carried out with eight oligonucleotide pairs (Table 1) as described by Ikenoue et al. (22) to amplify five different loci spread over the cag I and cag II regions. The presence or absence of the gene cagA, the promoter regions of cagE, cagG, and cagT, and the left end of cag-PAI (LEC) was detected. All PCR were performed in a total volume of 10 µl containing 0.25 µl of genomic DNA (50 ng/µl), 10× PCR buffer (Applied Biosystems), 2 mM MgCl2 (MBI), 80 µM dNTPs, 10 µM concentrations of each forward and reverse primer, and 1 U of Taq polymerase (Applied Biosystems). The total volume was made up with autoclaved Milli-Q water. H. pylori 26695 DNA was taken as a positive control. A negative control lacking either DNA or primers was also included. The absence of cag-PAI was confirmed in isolates reported to be negative for the same by using previously described primers flanking the cag-PAI (11, 22, 25).

Appropriate measures were taken to avoid contamination. PCR was performed in a 9700 thermal cycler (Applied Biosystems). Amplification conditions were optimized as follows: initial denaturation for 5 min at 94°C was followed by 40 cycles of denaturation at 90°C for 30 s, annealing at 52°C for 30 s, and extension at 72°C for 1 min. After a final extension at 72°C for 1 min, the amplified products were resolved on a 1% agarose gel in 1× TAE buffer (40 mM Tris-acetate-1 mM EDTA [pH 8.0]) containing ethidium bromide (0.5 µg/ml) and then visualized under UV light.

All of the cag-PAI rearrangement profiles generated in the present study were archived in the form of a searchable database (http://www.cdfd.org.in/amplibase/HP).

RESULTS

PCR amplification of cag-PAI regions. We determined intactness or otherwise of the cag-PAI genes by using specific flanking and internally anchored PCR primers. This was essentially to know the geographic prevalence of strains with different cag-PAI rearrangement patterns and whether these patterns could be associated with the clinical outcome of the infection. The reasons for spanning the entire cag-PAI region with eight specific PCR primers have been already explained elsewhere (28) based on Southern blot analysis and, more recently, by Ikenoue et al. (22). In our experiments, all of the PCR results were 100% reproducible on all occasions. All of the samples were independently tested at the Centre for DNA Fingerprinting and Diagnostics (CDFD) and the Deccan College of Medical Sciences. Our results, upon crosschecking, were found to be the same within and between the two labs.

PCR amplification protocols were portable and easy to perform. The PCR amplification procedure was tested on different positive and negative control samples and also with paired isolates and serial isolates obtained from the same patients months apart. Overall, the PCR technique that we used was robust, reproducible, portable, easy to perform, and highly specific for the corresponding regions of the cag-PAI (Fig. 2).
TABLE 1. Details of the PCR primers used for rearrangement analysis within cag-PAI

<table>
<thead>
<tr>
<th>Primer pairs</th>
<th>Sequence</th>
<th>Amplicon size (bp)</th>
<th>Locus name</th>
<th>Coordinates in HP026695 genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>cagAF1</td>
<td>AACAGGACAAGTATGCTAAGC</td>
<td>349</td>
<td>HP0547/JHP0495</td>
<td>582638-583137</td>
</tr>
<tr>
<td>cagAR1</td>
<td>TATATTGCTGCAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cagAF2</td>
<td>GATAACAGGCAAGCTTTTGA</td>
<td>701</td>
<td>HP0547/JHP0495</td>
<td>580092-580440</td>
</tr>
<tr>
<td>cagAR2</td>
<td>TCTGCACAACACATTTTGCAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cagAP-F1</td>
<td>GTTGGTTAAAATGGAATCG</td>
<td>730</td>
<td>Promoter region of HP0547/JHP0495</td>
<td>579697-580440</td>
</tr>
<tr>
<td>cagAR2</td>
<td>TCTGCACAACACATTTTGCAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cagAP-F2</td>
<td>CTACTGGTCCCAAACCATTTTT</td>
<td>1181</td>
<td>Promoter region of HP0547/JHP0495</td>
<td>579231-580440</td>
</tr>
<tr>
<td>cagAR2</td>
<td>TCTGCACAACACATTTTGCAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cagEF1</td>
<td>GCGATTGTTATTGTGTGATG</td>
<td>329</td>
<td>HP0544/JHP0492</td>
<td>577492-579697</td>
</tr>
<tr>
<td>cagER1</td>
<td>GAAGTGGTTAAAAAATCAATGCC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cagTF1</td>
<td>CCATGTTTATTACGGCGTTG</td>
<td>301</td>
<td>HP0532/JHP0522</td>
<td>565171-565471</td>
</tr>
<tr>
<td>cagTR1</td>
<td>CCATGTTTATTACGGCGTTG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LecF1</td>
<td>ATAGGCTGTGAATTGAG</td>
<td>384</td>
<td>JHP0521/JHP0522</td>
<td>547222-547242</td>
</tr>
<tr>
<td>LecR1</td>
<td>ATAGGCTGTGAATTGAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LecF2</td>
<td>ATAGGCTGTGAATTGAG</td>
<td>877</td>
<td>JHP0520/JHP521</td>
<td>547222-547242</td>
</tr>
<tr>
<td>LecR2</td>
<td>ATAGGCTGTGAATTGAG</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the strains with an incomplete or rearranged cag-PAI, PCRs were repeated for all of the regions to reconfirm the results.

The percentage of PCRs positive for the cagE and cagT genes was always (82%) than any other cag-PAI region studied (Fig. 1 to 3). These genes were more frequently intact compared to the cagA gene (72.8%). Deletions were frequently observed in the three ORFs located in the extreme left of the...
cag-PAI, namely, JHP520, -521, and -522 (annotated in the J99 genome), coding for the predicted ribose 5-phosphate isomerase and a membrane protein. In addition, the promoter region of the cagA gene was also disrupted in the majority of the isolates. Within the cagA gene, deletions were more frequently observed involving the 3' region of the gene corresponding to H. pylori 399 ORF JHP0495 than at the 5' end of this ORF. A very high amount of nucleotide diversity has already been reported in this region (25). All of the results corresponding to rearrangement patterns of cag-PAI and the constituent genes are shown in Fig. 1 and 3.

**Biogeographic distribution of cag-PAI rearrangements.** Our analyses revealed that the highest proportion of the strains from Japan (57.1%) were harboring an intact cag-PAI. A good fraction of Peruvian strains (18.6%), carried an intact cag-PAI. Only 12% of the 63 isolates recovered from northern, southern, and western parts of India had an intact cag-PAI. Fifty percent of the strains from West India (Pune) were devoid of the entire island of genes analyzed. Three of the ten (30%) strains from far-northern India (Ladakh) had an intact cag-PAI. Frequency of rearrangements in the cagA gene and the LEC region was higher for Indian isolates from Pune and Delhi than for isolates from Hyderabad and Ladakh. Conservation of cag-PAI from strains in Europe and Africa was minimum. Majority of strains from Ireland were found to be rearranged except for the 15.8% strains that carried an intact cag-PAI. Only 2 of the 66 English strains analyzed had an intact cag-PAI. None of the strains from Costa Rica had intact cag-PAI. Strains with intermediate genotype (deletions within the cag-PAI) were more common compared to strains with intact cag-PAI in all of the geographic regions studied. cagA gene was present in 72.8% of strains world over while both cagE and cagT genes were present in 82% of these isolates, indicating that the presence of cagA gene alone is not a marker for an intact cag-PAI. Conservation of cagE region was 100% in Japanese and Spanish strains and lowest in Ireland. Conservation of cagA promoter region was also highest among Japanese strains and was lowest in Cost Rican and Indian strains. The LEC region was found to be more frequently rearranged in isolates from Costa Rica than in other isolates in the study. The extreme right end of the cagA gene was found to be frequently rearranged in all of the African isolates studied compared to all other isolates. However, the proximal portion of this gene was present in 100% of the African isolates, the
highest level among all of the isolates analyzed from different regions. In all, a total of 33.33% of the strains representing Asian gene pool (Japan, native Peruvian, and Ladakh) were intact with respect to all of the genomic landmarks analyzed. Conversely, only 10.25% of the strains representative of all other gene pools (Indo-European, African, and American) were found to be harboring an intact PAI.

**Corroboration with clinical outcome.** Although no association could be discerned between the presence of cagA gene and the clinical outcome, the deletion frequency of both cagE and cagT genes was higher in nonulcer dyspepsia than in cases of ulcers and cancers. cagA, cagE, and cagT were found to be more frequently rearranged in isolates linked to rather benign infections than in isolates recovered from patients suffering from severe ulcers and gastric cancers worldwide. The cagA gene was intact in 60% of the isolates from benign cases, whereas it was intact in >82% of the isolates recovered from patients with ulcers and cancers. The cagT gene was conserved in 100 and 80% of the isolates analyzed from cases of severe gastroduodenal pathology and gastritis, respectively. Conversely, the cagA promoter and LEC regions were rearranged more frequently in isolates linked to severe pathology than isolates responsible for a more benign outcome worldwide (Fig. 4). Many of the Japanese and Peruvian strains linked to nonulcer gastritis were also found to be intact with respect to all of the regions analyzed. Surprisingly, none of the five cancer strains from Costa Rica were found to be intact.

**DISCUSSION**

Pathogenic *H. pylori* may have evolved from its benign counterparts by acquiring large amounts of genetic information in the form of PAIs that would have conferred upon it an increased fitness for colonizing new host niches (19). The ancient circumstances associated with such acquisition and the impact of the same on human gastric disease epidemiology is a subject of intense argument. Our study sought to perform an extensive survey of the genetic rearrangements within the cag-PAI by using a large number of *H. pylori* strains from across the globe. It appears that rearrangement in cag-PAI is quite a prevalent phenomenon, and its constituent genes are under selection pressure more than any other regions in the chromosome. A recent study, in which paired antral and corpus isolates differed with respect to polymorphism of cagA locus, supports this idea (9). However, these paired isolates were identical with respect to...
to other important loci, such as the vacA, ribAP, vacAP, and cagA-glr motifs, ~50 fluorescent amplified fragment length polymorphism loci, and many RAPD [random(y) amplified polymorphic DNA] loci (9). Mechanisms underlying such rearrangements are unknown. Insertion elements have been reported to cause rearrangements in the H. pylori genome (3, 50). The core sequence of the 31-bp direct repeat flanking the cag-PAI has sequence homology with the flanking ends of IS605 element (11). Insertion of the latter in functionally silent regions of PAIs has been reported (20, 42). Strains lacking the cag-PAI contained the IS605 (5, 23, 28), although its location within the cag-PAI could not be demonstrated. In contrast, certain studies have failed to identify IS605 elements within the cag-PAI and their association to the strains with rearranged genes (23, 28, 37, 47). Apart from this, one can also guess the involvement of certain dedicated gene products (such as DNA helicases), encoded by chromosomal regions flanking the cag-PAI, in propagating the island (16). However, in our study the rearrangement profiles obtained for regions within the cag-PAI were independent of the insertion deletion and substitution activities (in-del-sub) involving putative helicase (hel), IS605 and IS606 at the extreme right end of the cag-PAI (data not shown). These in-del-sub profiles have been previously reported for a majority of isolates from our collection (10, 21, 25).

Genotype-phenotype correlation in case of cag-PAI genotypes is a difficult exercise, although there are reports describing the association between the presence of the cag-PAI and disease type (5, 6, 34). The pathogenic role of the cag-PAI as a whole or in part, in disease development, is not yet completely understood (8, 12, 32, 39). H. pylori strains differing in virulence potentials can colonize in a mixed infection and emergence of recombinant strains, with deletion of cag-PAI, cannot be ruled out. It is therefore possible to isolate strains with rearranged cag-PAI from gastric ulcer and cancer patients. Isolates linked to severe pathology (including gastric cancers) in our study did not reveal an intact cag-PAI. In case of chronic infections, it is possible that strains with intact cag-PAI may subsequently rearrange the island long after the initial damage (leading to severe pathology after many years). On the basis of our analyses, the cag-PAI was found to be most intact in Japanese strains. This is in agreement with reports of an increased occurrence of gastric cancer and augmented severity of gastric ulcers in this part of the world (27, 31, 44). However, it will be premature to link disease outcome to rearrangement patterns within cag-PAI unless a large number of strains with precisely defined clinicopathological and pathological features are analyzed. Also, the cag-PAI may not be the principal virulence factor, as suggested by the absence or sporadic distribution of the cag-PAI genes among strains from varied clinical outcome. This may be due to the fact that the development of ulcer disease is a complex process that also involves factors other than the cag-PAI, such as the constituents of the outer membrane, vacA, and the nuclear activating protein, etc. Nonetheless, the PCR-based rearrangement profiling described here could provide the clinician with an educated guess on the present proinflammatory power of the infecting strain. This will be immensely helpful in case of patients colonized only with a single strain and not undergoing eradication therapy. The genotypic signatures thus developed may also help track the routes of infection in communities and may be used as reference material in the form of strain databases. Finally, the present study of H. pylori cag-PAI genotypes from different human populations adds to our understanding of bacterial-host interactions and the ecology of the bacterium in mixed infections. It will also be helpful to confirm the importance of H. pylori geographical genomics in studying gastro-intestinal pathology in humans.

ACKNOWLEDGMENTS

We thank Seyed E. Hasnain for his patronage, encouragement, guidance, and support. We thank Douglas E. Berg (Washington University, St. Louis, Mo.) for sharing some genomic DNA of H. pylori isolates from his collection of strains. We thank Robert H. Gilman and John Albertson for DNA samples from Peruvian and English strains, respectively. We thank Martin Blaser, Richard Novick, and Mark Achtman for allowing us to use Ladhak isolates included in this study and Cyril J. Smyth for his patronage. We also thank G. Balakrish Nair (ICDDR,B, Bangladesh) for guidance.

This study was funded in part by a grant from the Department of Biotechnology, Government of India, to N.A., C.M.H., B.D., and Y.S. I.M.C. was a visiting scientist at DCMS and CDFD and was supported by the Irish Health Research Board. N.A. is a staff scientist and currently leads the pathogen evolution program at CDFD.

REFERENCES


