Diversity in the cag pathogenicity island of Helicobacter pylori isolates in populations from North and South India

Sushil Kumar,1 Ashok Kumar1 and Vinod Kumar Dixit2

1School of Biotechnology, Banaras Hindu University, Varanasi, India
2Department of Gastroenterology, Institute of Medical Sciences, Banaras Hindu University, Varanasi, India

The cag pathogenicity island (cagPAI) has been reported to be the major virulence determinant in Helicobacter pylori-related diseases. In the present study, the diversity of the cagA gene and the integrity of the cagPAI in 158 H. pylori strains from Varanasi (North India) and Hyderabad (South India) were studied by amplifying the cagA gene (~3.5 kb), followed by PCR-RFLP analysis. The results revealed significant differences in the cagA gene and the integrity of the cagPAI between North and South Indian isolates. Of 158 isolates, 40 (34.8%) from Varanasi and 20 (46.5%) from Hyderabad were found to carry an intact cagPAI. A partially deleted cagPAI was present in 75 (65.2%) isolates from Varanasi and 23 (53.5%) from Hyderabad. None of the isolates showed complete deletion of the cagPAI. Differences in the cagA 5’ and 3’ regions were also noted, and 11 isolates (8 from Varanasi and 3 from Hyderabad) that were cagA negative with primers for the 5’ region turned out to be cagA positive with primers for the 3’ variable region. It is tentatively concluded that the 3’ variable region may be a better marker for cagA typing. The results also showed that the majority of the isolates harboured the Western-type EPIYA motif. PCR-RFLP analysis of the cagA gene showed 29 distinguishable digestion patterns, and cluster analysis of RFLP types from a random selection of 32 isolates placed all of the isolates into 5 groups. These results demonstrate that significant differences in the cagPAI occur among isolates from North and South India, and that RFLP of cagA could be employed for elucidating genetic variations among various isolates of H. pylori.

INTRODUCTION

Helicobacter pylori, a Gram-negative, microaerophilic, spiral-shaped bacterium, colonizes the human stomach and is estimated to inhabit at least half of the world’s human population. Several H. pylori virulence genes that may play a role in pathogenicity have been identified. Of these, the most important determinants are the cytotoxin-associated antigen gene (cagA) and the vacuolating cytotoxin antigen gene (vacA) (Covacci et al., 1999; Graham & Yamaoka, 1998; Higashi et al., 2002). The cagA gene is one of several genes in a pathogenicity island known as cagPAI, which is an approximately 40 kb locus in the H. pylori genome (Censini et al., 1996). It has been reported that most cagA-positive strains producing CagA protein carry all of the other genes of the cagPAI, but all cagA-negative strains lack the cagPAI (Naito et al., 2006).

Based on these findings, there is a general consensus that the presence of cagA is the marker for the presence of the cagPAI (Covacci et al., 1999). However, Censini et al. (1996) have reported partial deletion of the cagPAI and the reason for the genetic rearrangement was explained by the incorporation of an insertion element, IS605, in the cagPAI. Thus, depending on integrity, cagPAI appears to exist in three possible forms: intact, partially deleted and completely deleted.

In comparison with cagPAI-negative strains, infection with cagPAI-positive strains of H. pylori significantly increases the risk of developing severe gastric mucosal inflammation, duodenal ulceration and gastric cancer (GC) (Hatakeyama, 2009; Kumar et al., 2009; Segal et al., 1999). It has been observed that only one-half to two-thirds of Western isolates carry the cagPAI, whereas almost all East Asian strains carry this pathogenicity island. It has also been reported that the cagPAI is highly conserved among Japanese isolates, least conserved in European and African isolates, and is very poorly conserved in Indian isolates (Nomura et al., 1991).

Abbreviations: DU, duodenal ulcer; GC, gastric cancer; GU, gastric ulcer.

The GenBank/EMBL/DDBJ accession numbers for the sequences determined in this study are FJ984536 (cagA), FJ984537 and FJ984538 (cagE) and FJ984539 (cagT).
Screening of cagPAI genes is frequently carried out for their involvement in virulence characteristics and to assess the integrity of the cagPAI (Fig. 1) (Censini et al., 1996; Terry et al., 2005). It was reported that the cagE gene might be a better marker for an intact cagPAI in Japanese and French isolates (Ikenoue et al., 2001; Maeda et al., 1999). However, deletion of the cagE, cagT, cagA, cagG and cagM genes has been reported in several cases of chronic gastritis, gastric ulcer (GU) and GC cases, indicating that the pathogenicity of H. pylori may not be determined by cagPAI genes alone (Covacci et al., 1999).

Data on the prevalence of an intact versus a rearranged cagPAI in H. pylori strains from India are lacking, and it is felt that data on a global scale will be required to understand the role of cagPAI rearrangement in disease outcome. In the present study, an attempt was made to investigate cagA diversity and its integrity in patients from South and North India employing PCR techniques, with emphasis on the amplification of the whole segment of cagA, followed by RFLP analysis.

**METHODS**

**Patients and bacterial strains.** In total, 158 H. pylori strains from Varanasi (North India) and Hyderabad (South India) were isolated from the gastric biopsy samples of patients referred to the gastroendoscopy unit of S. S. Hospital, Banaras Hindu University, Varanasi, India (115 patients), and Deccan College of Medical Sciences and Allied Hospitals, Hyderabad, India (43 patients). Patients ranged from 20 to 75 years old (mean age 45 years). Each patient underwent upper gastroendoscopy for visual examination and biopsy collection. Biopsies from the antral part of the stomach of patients suffering from gastric inflammation (gastritis), GUs and GC, including both malignant and nodular types, and healthy controls (normal gastric mucosa) were taken using an endoscope (CV-70 Videoscope; Olympus). Unless otherwise stated, endoscopic forceps were sterilized in 2% glutaraldehyde solution for 20 min, followed by thorough washing in sterilized distilled water to avoid contamination (Kumar et al., 2009). The case histories of the patients revealed that none had received non-steroidal anti-inflammatory drugs, proton-pump inhibitors or antibiotics during the last 2 months. Written consent was obtained from all patients before collection of the biopsy.

**Isolation and culture of H. pylori.** The biopsy sample of each patient was inoculated onto Brucella agar (Becton Dickinson) supplemented with 7% (v/v) lysed sheep blood, 10 μg vancomycin ml⁻¹, 2.5 IU polymyxin B ml⁻¹, and 5 μg amphotericin B ml⁻¹, and incubated at 37°C, with 100% humidity, under microaerophilic conditions (10% CO₂, 5% O₂ and 85% N₂). Putative colonies appearing after 3–5 days of growth were picked and streaked onto fresh plates. After subculturing two or three times, all of the isolates were grown on Brucella agar supplemented with 7% (v/v) lysed sheep blood (Kumar et al., 2008). Tentative identification of H. pylori was made on the basis of colony morphology (small, translucent colonies), shape (curved) and a Gram-negative stain, and of biochemical tests for urease, catalase and oxidase.

**DNA extraction and amplification of the cagA, cagE and cagT genes.** Genomic DNA was extracted using a DNeasy tissue kit (Qiagen) following the instructions of the manufacturer. The cagA, cagE and cagT genes of pure cultures were amplified using a standard...
set of primers was based on the structure of the cagPAI as shown in Fig. 1. Amplification was performed in a PTC-100 thermal cycler (MJ Research). The PCR mix comprised 0.75 U Taq DNA polymerase (Bangalore Genei), 1 × PCR assay buffer with 1.5 mM MgCl₂, 10 pmol forward and reverse primers (Integrated DNA Technologies), 200 μM each dNTP and 25 ng template DNA in a total volume of 25 μL. Thermal parameters for the amplification were: initial denaturation for 5 min at 94°C; 35 cycles of 30 s at 94°C, 30 s at 52°C and 1 min at 72°C; and final extension for 5 min at 72°C.

The amplified products were electrophoresed on a 1.2% agarose gel in TAE buffer ([40 mM Tris/acetate (pH 8.0), 1 mM EDTA] containing ethidium bromide (0.5 μg ml⁻¹) and visualized with a gel documentation unit (Bio-Rad). Genomic DNA of H. pylori strain 60190 (s1/m1 genotype and cagA positive, obtained from Professor J. C. Atherton, Centre for Biomolecular Sciences, University of Nottingham, UK) was used as a reference.

**Amplification of the large fragment and 3’ variable region of cagA.** Amplification of the large fragment and 3’ variable region was performed using primer set cag5c-F/cagR-R14660 and cagA-F40481/cagA-R41660, respectively. The PCR mix included 1.5 U Taq DNA polymerase (Bangalore Genei), 1 × PCR assay buffer with 1.5 mM MgCl₂, 10 pmol each forward and reverse primer (Integrated DNA Technologies), 200 μM each dNTP and 50 ng template DNA in a total volume of 50 μL. Thermal parameters for the amplification were: initial denaturation for 5 min at 94°C; 35 cycles of 1 min at 94°C, 1 min at 52°C and 3 min at 72°C; and final extension for 10 min at 72°C. Amplification of cagA membrane-targeting signal motif EPIYA-A, -B, -C and -D was performed according to Schmidt et al. (2009). The amplified products were analysed as described above.

**RFLP of cagA.** Forty microlitres of the amplified PCR product was removed, ethanol precipitated and suspended in 20 μl deionised water. Of this 20 μl of purified amplified DNA, 10 μl was digested with each of 10 U HindIII and Alul for 16 h at 37°C following the manufacturer’s instructions (Promega). The digested products were run next to a 100 bp DNA ladder in a 3% agarose gel at 100 V in TAE buffer for 4–5 h. Cluster analysis of RFLP types was performed by the documentation unit (Bio-Rad). Genomic DNA of H. pylori positive, obtained from Professor J. C. Atherton, Centre for Biomolecular Sciences, University of Nottingham, UK) was used as a reference.

Sequencing of cagA, cagE and cagT. Sequencing of the cagA, cagE and cagT genes of selected Varanasi (North India) isolates was performed in an ABI-PRISM 310 genetic analyzer (Applied Biosystems). The primers for sequencing PCR were the same as those used for the PCR assay. PCR and direct sequencing were performed at least twice to determine and confirm the DNA sequences of each isolate. All of the sequences were matched against nucleotide sequences present in GenBank using the BLASTN program (Altschul et al., 1997; www.ncbi.nlm.nih.gov/blast).

**Tests for sensitivity and specificity of primers.** The sensitivity of all of the primers was tested by PCR following dilution of template DNA (concentrations ranging from 0.01 pg μl⁻¹ to 10 ng μl⁻¹) from selected isolates. Reamplification of low-intensity bands was also tested using the same primers. To check the specificity of the primers, template DNA isolated from a large number of bacteria, including Helicobacter species, was tested for amplification of the desired fragment in a PCR assay.

**Statistical analysis.** Fisher’s exact test was used to calculate statistical significance, and a P value of <0.05 was considered significant.

### RESULTS AND DISCUSSION

**Sensitivity and specificity of primers**

The sensitivity test of the primers in the PCR assay revealed that the cagA 5’ end, cagE and cagT genes could be amplified from 2 pg template DNA in a 20 μl reaction mixture, whereas amplification of full-length cagA (~ 3.5 kb), the variable region (3’ end) and the 16S rRNA gene could be achieved with as little as 0.2 pg template DNA. A routine PCR assay also showed that none of the primer pairs produced a smear during reamplification and that a fragment of the desired size was always amplified. Similar to other reports, the specificity test showed that the primers used in this study were specific for strains of H. pylori and the reference strain, with no amplification using template DNA from various other bacteria.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Primer sequence (5’→3’)</th>
<th>Amplicon size (bp)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>16S rRNA-F</td>
<td>TAAGAGATCAGGCCTATGTCC</td>
<td>534</td>
<td>Kumar et al. (2009)</td>
</tr>
<tr>
<td>16S rRNA-R</td>
<td>TCCACCGTCTTTAAGCGCAAT</td>
<td>350</td>
<td>Kumar et al. (2008)</td>
</tr>
<tr>
<td>cag5c-F</td>
<td>GGTGTGACTGATATTTCTCATAT</td>
<td>325</td>
<td>Mukhopadhyay et al. (2000)</td>
</tr>
<tr>
<td>cag3c-R</td>
<td>GGAATCTCTTAACTCAGTTGCGG</td>
<td>1253</td>
<td>Mukhopadhyay et al. (2000)</td>
</tr>
<tr>
<td>cagA5 (F)</td>
<td>AGGATTCATCAAGGTAAACCGAAGC</td>
<td>329</td>
<td>Ikenoue et al. (2001)</td>
</tr>
<tr>
<td>cagA2 (R)</td>
<td>CGGATTTTCCACATCCTTTTGGT</td>
<td>301</td>
<td>Ikenoue et al. (2001)</td>
</tr>
<tr>
<td>cagA-R41660</td>
<td>CGGATTTTCCACATCCTTTTGGT</td>
<td>60</td>
<td>Ikenoue et al. (2001)</td>
</tr>
</tbody>
</table>

F, Forward; R, reverse.
Detection of *H. pylori* in different patients

The results of the upper gastrointestinal endoscopies revealed that, of the 158 patients, 138 had various types of gastrointestinal disease and 20 were healthy (controls) with no evidence of gastric disease. Among the 115 patients from Varanasi (North India), 28 had gastritis, 18 had GC, 31 had duodenal ulcers (DUs), 23 had GUs and 15 were healthy. Similarly, in the 43 patients from Hyderabad (South India), 15 had DUs, 11 had GUs, 8 had GC, 4 had gastritis and 5 were healthy. *H. pylori* was successfully isolated from the biopsy samples of all 158 cases from both regions, and identification of various strains was made on the basis of morphological characteristics, biochemical tests and amplification of the *H. pylori*-specific 16S rRNA gene (Kumar et al., 2008).

Characteristics of the cagPAI region

Five pairs of primers were used to detect the presence of the cagA, cagE and cagT genes in various isolates of *H. pylori*. Standard *H. pylori* strain 60190 (containing the entire cagPAI region) was used as positive control for each PCR assay. The data in Table 2 show the distribution pattern of the cagA 5’ and 3’ ends, and the cagE and cagT genes, in various isolates from different disease outcomes. In the 43 strains from Hyderabad, the cagA 5’ and 3’ regions were present in 34 (79.1%) and 37 (86.0%) strains, respectively. Among the 115 isolates from Varanasi, 80 (69.6%) strains showed the presence of the cagA 5’ region, whilst 88 (76.5%) had the cagA 3’ region (Table 2). The cagE gene was detected in 35 strains from Hyderabad and 82 from Varanasi, and the cagT gene in 72 and 32 strains, respectively (Table 2). With a view to strengthening our findings pertaining to the amplification of cagA, cagE and cagT, DNA sequencing of all three genes was carried out from selected isolates. Matching of sequences with those in GenBank showed homology to the desired level, confirming the fidelity of the amplified product.

The percentage of partially deleted cagPAIs was found to be higher in the strains from Varanasi (65.2%) compared with those from Hyderabad (53.5%), although the percentage of intact cagPAIs was higher in the strains of Hyderabad (Table 3). It was also evident from the findings that the percentage of intact cagPAIs was significantly higher in GC and GU cases, whereas the percentage of partially deleted cagPAIs was higher in gastritis and DU in both regions (Table 3). None of the strains from either region showed complete deletion of the cagPAI. Among the many virulence markers present in the *H. pylori* genome, the cagPAI is a major virulence factor and is associated with severe gastroduodenal pathology (Blaser et al., 1995; Censini et al., 1996; Guillemin et al., 2002). Over the last 10 years, several reports showing rearrangements in this island have appeared (Choi et al., 2007; Covacci et al., 1999; Kauser et al., 2004; Maeda et al., 1999), and the composition of the cagPAI in clinical *H. pylori* isolates has been studied in different populations by various methods, including PCR, dot blotting and long-distance PCR (Ikenoue et al., 2001; Jenks et al., 1998; Maeda et al., 1999; Matteo et al., 2007). The findings of the present study showed marked differences in the integrity and prevalence of the cagPAI of 158 strains isolated from patients of geographically distinct regions (North and South India). Similar to our findings, differences in the percentage of intact cagPAIs in strains isolated from different parts of the world have been reported (Ali et al., 2005; Kauser et al., 2004; Mukhopadhay et al., 2000). As significant geographical differences among strains of *H. pylori* are known to exist, notable differences in the integrity of the cagPAI were expected and our results resemble those of other reports (Ali et al., 2005). Although DU is also considered to be a severe form of gastroduodenal disease, the findings of this study showed that the percentage of DU strains carrying an intact cagPAI was lower, i.e. 29 and 33.3% in the strains from Varanasi and Hyderabad, respectively, Jenks et al. (1998) reported that the presence of certain

Table 2. Distribution pattern of the consensus 5’ region and variable 3’ region of cagA, and the cagE and cagT genes, in various *H. pylori* isolates

Results are shown as the number of isolates (%).

<table>
<thead>
<tr>
<th>Clinical status</th>
<th>No. of <em>H. pylori</em> isolates</th>
<th>cagA 5’ end</th>
<th>cagA 3’ end</th>
<th>cagE</th>
<th>cagT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VAR</td>
<td>HYD</td>
<td>VAR</td>
<td>HYD</td>
<td>VAR</td>
</tr>
<tr>
<td>DU</td>
<td>31</td>
<td>15</td>
<td>17</td>
<td>54.8</td>
<td>12</td>
</tr>
<tr>
<td>GU</td>
<td>23</td>
<td>11</td>
<td>17</td>
<td>73.9</td>
<td>10</td>
</tr>
<tr>
<td>GC</td>
<td>18</td>
<td>8</td>
<td>18</td>
<td>100</td>
<td>7</td>
</tr>
<tr>
<td>Gastritis</td>
<td>28</td>
<td>4</td>
<td>20</td>
<td>71.4</td>
<td>3</td>
</tr>
<tr>
<td>Normal</td>
<td>15</td>
<td>5</td>
<td>8</td>
<td>53.3</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>115</td>
<td>43</td>
<td>80</td>
<td>69.6</td>
<td>34</td>
</tr>
</tbody>
</table>

HYD, Hyderabad; VAR, Varanasi.

* For isolates from VAR, disease versus normal, P = not significant; for isolates from HYD, disease versus normal, P = 0.05.
† For isolates from VAR, disease versus normal, P = 0.04; for isolates from HYD, disease versus normal, P = 0.01.
Differences in the distribution of PAIs between North and South Indian isolates could tentatively be explained by the fact that there are striking differences in the geographical distribution of *cagA* genotypes all over the world. Additionally, this could also be explained in light of a report that the diversity of *H. pylori* will be enhanced if humans differ in their dietary habits (gastric environments), as well as in individual traits such as specific immune responses and/or availability of receptors helpful in adhesion (Ahmed et al., 2003). This is highly applicable in the context of Indian populations, as there is much variation in dietary habits, lifestyle and intake of medicines among the various communities, together with a great ethnic diversity (Ahmed et al., 2003).

**Analysis of the cagA 3′ variable region**

In order to examine variations in the *cagA* 3′ variable region, PCR amplification of this region was performed and the size of all products was determined. A typical PCR amplification of the *cagA* 3′ variable region is shown in Fig. 2. The size of the PCR product of all isolates was in the range of 1042–1360 bp. Detailed analysis showed that the fragment size in strains from the Varanasi and Hyderabad regions was in the range of 1042–1267 bp and 1167–1360 bp, respectively. Moreover, grouping based on size resulted in six groups. Group I contained isolates with a PCR fragment size of 1042 bp and comprised 12 (13.6 %) of the 88 strains from Varanasi. Group II isolates produced a fragment of size 1165 bp and the group comprised 7 (8.0 %) of the strains from Varanasi and 2 (5.4 %) of the 37 strains from Hyderabad. Group III comprised the standard ATCC strain (60190) with a size of 1234 bp. Group IV, with a fragment of 1267 bp, contained 69 (78.4 %) and 33 (89.2 %) strains from Varanasi and Hyderabad, respectively, and formed the largest group. Groups V and VI, with a fragment size of 1285 and 1360 bp, respectively, comprised one (2.7 %) strain each from Hyderabad. It was also evident from our results that eight strains from Varanasi and three from Hyderabad, which were *cagA* negative with primers for the 5′ region, turned out to be

**Table 3. Geographical distribution and relationship between the presence of the cagPAI and clinical status**

Results are shown as number of patients (%).

<table>
<thead>
<tr>
<th>Disease</th>
<th>No. of patients</th>
<th>Intact cagPAI*</th>
<th>Partially deleted cagPAI†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VAR</td>
<td>HYD</td>
<td>VAR</td>
</tr>
<tr>
<td>DU</td>
<td>31</td>
<td>15</td>
<td>9(29.0)</td>
</tr>
<tr>
<td>GU</td>
<td>23</td>
<td>11</td>
<td>12(52.2)</td>
</tr>
<tr>
<td>GC</td>
<td>18</td>
<td>8</td>
<td>12(66.7)</td>
</tr>
<tr>
<td>Gastritis</td>
<td>28</td>
<td>4</td>
<td>5(17.9)</td>
</tr>
<tr>
<td>Normal</td>
<td>15</td>
<td>5</td>
<td>2(13.3)</td>
</tr>
<tr>
<td>Total</td>
<td>115</td>
<td>43</td>
<td>40(34.8)</td>
</tr>
</tbody>
</table>

HYD, Hyderabad; VAR, Varanasi.

*DU versus normal, P= not significant; GU versus normal, P<0.01; GC versus normal, P= not significant; gastritis versus normal, P= not significant; disease versus normal, P= 0.03.

†DU versus normal, P= not significant; GU versus normal, P<0.01; GC versus normal, P= not significant; gastritis versus normal, P= not significant; disease versus normal, P= 0.03.

genes (*cagA, cagE, cagM, cagT, ORF6, ORF10 and ORF13*) in the *cagPAI* is highly associated with DUs. To some extent, our findings are similar to the above report, although for only two genes, i.e. *cagA* and *cagE* in the strains from Hyderabad, where 86.7 % of DU strains carried *cagA* and 80.0 % had *cagE* (Table 2). However, the value was lower in the strains from Varanasi, where *cagA* was found in 64.5 % and *cagE* in 61.3 %. Kauser et al. (2004), while screening the presence of the *cagPAI* in isolates from patients from eight countries, found that the *cagPAI* is disrupted in the majority of isolates throughout the world. According to their study, the *cagPAI* was highly conserved in Japanese isolates (57.1 %), and least conserved in European and African strains. They reported that only 18.6 % of the Peruvian and 12 % of the Indian isolates carried an intact *cagPAI*. In other reports, Mukhopdhyay et al. (2000) reported a level of more than 96 % in Calcutta strains of peptic ulcer and non-ulcer dyspepsia. In contrast, Ali et al. (2005) reported a very low percentage (6.9 %) of intact *cagPAIs* in DU strains from Hyderabad. Our findings differ from those of Ali et al. (2005), as we routinely observed a higher percentage of intact *cagPAI* in DUs. It is indeed hard to correlate disease outcome with *cagPAI* integrity in view of the vast differences prevailing among strains from different countries/regions (Ali et al., 2005; Azuma et al., 2002; Ikenoue et al., 2001; Mukhopdhyay et al., 2000; Rhead et al., 2007). However, it appears that people with strains carrying an intact *cagPAI* are more prone to develop GC and GUs in comparison with those carrying a partially deleted *cagPAI*. It would not be appropriate to link disease outcome to *cagPAI* rearrangement pattern until a large number of strains causing various gastrointestinal diseases have been analysed.

Differences in the distribution of *cagPAIs* between North and South Indian isolates could tentatively be explained by the fact that there are striking differences in the geographical distribution of *cagA* genotypes all over the world. Additionally, this could also be explained in light of a report that the diversity of *H. pylori* will be enhanced if humans differ in their dietary habits (gastric environments), as well as in individual traits such as specific immune responses and/or availability of receptors helpful in adhesion (Ahmed et al., 2003). This is highly applicable in the context of Indian populations, as there is much variation in dietary habits, lifestyle and intake of medicines among the various communities, together with a great ethnic diversity (Ahmed et al., 2003).
cagA positive after amplification with primers for the 3’ variable region. Although the numbers are too low to confirm this result, it appears that the 3’ variable region might be a better marker for the study of cagA typing and genetic variation. However, Rota *et al.* (2001) reported that patients who were cagA negative by amplification with primers for the variable sequence were positive by primers for the consensus region. This ambiguity may be resolved if the study is carried out with a large number of patients, together with selecting a variety of primers from the 5’ and 3’ variable regions.

Analysis of the cagA membrane-targeting signal motif EPIYA (Glu-Pro-Ile-Tyr-Ala) revealed that the cagA variable region of group I contained the EPIYA-AB type, and group II consisted of EPIYA-ABB (five Varanasi isolates and one Hyderabad isolate) and -ABD types (two Varanasi isolates and one Hyderabad isolate). Groups IV and V contained EPIYA-ABC, whereas group VI had the EPIYA-ABCC type. It was also noted that the primers used could detect both Western and East Asian types of cagA. Overall analysis of the data suggested that group II contained the East Asian-type cagA, and groups III, IV, V and VI the Western-type cagA. This distribution pattern of EPIYA motifs is interesting, as the majority of the isolates (104) showed the presence of the Western type, and only three isolates had the East Asian-type EPIYA motif. However, 18 isolates showed an abnormal type of EPIYA motif. Our data are in agreement with previous findings where Western cagA in *H. pylori* isolates from India has been reported (Mukhopadhyay *et al.*, 2000; Schmidt *et al.*, 2009).

**Diversity based on RFLP analysis of cagA**

Having shown the notable differences in status of the cagPAI and cagA 3’ variable region among strains from North and South India, we next looked at diversity on the basis of RFLP analysis of the cagA gene (~3.5 kb). We successfully amplified the full-length cagA gene from 114 isolates (80 from Varanasi and 34 from Hyderabad), and randomly selected 32 strains (plus the reference strain) for RFLP. Of these 32 strains, 11 were from Varanasi and included 2 strains each from normal, gastritis, GU and GC cases, and 3 from DU cases. The remaining 21 isolates were from Hyderabad, of which 4 isolates each were from gastritis and GC, and 2, 6 and 5 isolates were from normal, DU and GU cases, respectively. A typical representation of *HindIII* - and *AluI*-digested fragments of the amplified products (3.5 kb amplicon) from 19 isolates of Hyderabad is shown in Fig. 3. Based on fingerprinting, 29 distinguishable RFLP types were observed from the 32 isolates (Fig. 4). The RFLP types were always more distinct following digestion with *AluI* than with *HindIII*.

To reveal the genetic relatedness among the 32 clinical strains, along with the standard ATCC reference strain (60190), a dendrogram was constructed based on the RFLP patterns of cagA. The similarity coefficient revealed that all 32 strains could be placed into five major groups, named A–E (Fig. 4). Group E contained five strains from Varanasi and showed 54% similarity to the strains placed in groups A, B, C and D. Group D comprised four strains: three from Varanasi and one from Hyderabad, and they shared 60% similarity with the strains from groups A, B and C. Group C contained four strains: two each from Varanasi and Hyderabad, and showed 64.8% similarity with the strains from groups A and B. Group B emerged as an interesting group as it contained 18 strains, all from Hyderabad, and showed 66% similarity with the strains of group A. Group
A consisted of only two strains, one from Varanasi and the reference isolate. Among all the strains, V9 and V14 from group D, and H16, H17 and H18 from group B, showed 100% similarity (Fig. 4). Clinical manifestations revealed that group A contained one strain from a gastritis case and the reference isolate. Group B, with 18 strains from Hyderabad, had two strains from normal, four from gastritis, five from DUs, four from GU and three from GC cases. Group C contained four strains, two each from DU and GU cases. Group D had three strains from GC cases and one from a GU case. Group E contained five strains, two each from normal and gastritis cases, and one from a DU case. The PCR-RFLP method has not been extensively employed for studying the high degree of genomic diversity prevailing in \textit{H. pylori} (Fantry et al., 1996; Steichen et al., 2007; Stone et al., 1997). Stone et al. (1997) used PCR-RFLP for typing the \textit{ureC} gene in order to relate \textit{H. pylori} epidemiologically. Fujimoto et al. (1994) also reported that 25 different isolates could be divided into 25 types when the amplified product of the \textit{ureC} gene was digested with \textit{Hha}I, \textit{Mbo}I and \textit{Mse}IA. The findings of the present study based on the RFLP pattern of the \textit{cagA} gene clearly showed heterogeneity in the \textit{H. pylori} strains and therefore justify the usefulness of this approach. The results are interesting as there are no data on PCR-RFLP types of the \textit{cagA} gene targeting a fragment size of 3.5 kb (Choi et al., 2007; Mukhopadhyay et al., 2000). That \textit{cagA} typing might indeed be useful in revealing genetic diversity among different strains of \textit{H. pylori} was also shown by the degree of genetic relatedness among various isolates. Cluster analysis revealed that: (i) the majority of the isolates from Varanasi (North India) and Hyderabad (South India) belonged to different groups, and (ii) strains from both regions, in spite of showing identical clinical outcomes, fell into different groups, apart from two isolates from group D and three from group B. We thus believe that PCR-RFLP of the \textit{cagA} gene could be conveniently used for the study of genetic diversity among \textit{H. pylori}.

**Fig. 4.** Dendrogram constructed on the basis of the PCR-RFLP types of the \textit{cagA} gene. All 32 clinical isolates and the reference strain could be placed in 5 major groups. Group A had the least number of isolates and group B the most.
In conclusion, our findings clearly demonstrate diversity in cagA, and differences in the distribution and integrity of the cagPAI, among the H. pylori isolates of North and South India. The results also showed that the majority of the isolates harboured the Western-type EPIYA motif. Additionally, RFLP of cagA provided useful data and may be used for genotyping a large number of strains. Data on the geographical distribution of cagPAI rearrangement patterns may provide hitherto unknown facts in relation to bacterial virulence, host genetic predisposition and niche characteristics.

ACKNOWLEDGEMENTS

S.K. is grateful to the University Grants Commission, New Delhi, India, for the award of Junior Research Fellowship. This study was partly supported by a research grant sanctioned to A.K. by the Department of Biotechnology, Government of India, New Delhi (BT/PR 1239/AGR./02/065/98). We are thankful to Professor John C. Atherton (Nottingham, UK) for providing us with the genomic DNA of the reference H. pylori isolate.

REFERENCES


