Diverse Characteristics of the \textit{cagA} Gene of \textit{Helicobacter pylori} Strains Collected from Patients from Southern Vietnam with Gastric Cancer and Peptic Ulcer\textsuperscript{V}

Bui Xuan Truong,\textsuperscript{1,6} Vo Thi Chi Mai,\textsuperscript{2} Hiroshi Tanaka,\textsuperscript{6} Le Thanh Ly,\textsuperscript{3} Tran Minh Thong,\textsuperscript{4} Hoang Hoa Hai,\textsuperscript{5} Dao Van Long,\textsuperscript{1} Keisuke Furumatsu,\textsuperscript{6} Masaru Yoshida,\textsuperscript{6} Hiromu Kutsumi,\textsuperscript{6} and Takeshi Azuma\textsuperscript{6*}

Department of Gastroenterology, Bach Mai Hospital, Hanoi, Vietnam; Department of Microbiology, Cho Ray Hospital, Ho Chi Minh City, Vietnam; Department of Gastroenterology and Hepatology, Cho Ray Hospital, Ho Chi Minh City, Vietnam; Department of Pathology, Cho Ray Hospital, Ho Chi Minh City, Vietnam; Department of Scientific Study and Management, Cho Ray Hospital, Ho Chi Minh City, Vietnam; and Department of Gastroenterology, Kobe University Graduate School of Medicine, Kobe, Japan

Received 12 March 2009/Returned for modification 20 April 2009/Accepted 11 September 2009

The pathogenesis of gastroduodenal diseases is related to the diversity of \textit{Helicobacter pylori} strains. \textit{CagA}-positive strains are more likely to cause gastric cancer than \textit{CagA}-negative strains. Based on EPIYA (Glu-Pro-Ile-Tyr-Ala) motifs at the carboxyl terminus corresponding to phosphorylation sites, \textit{H. pylori} \textit{CagA} is divided into East Asian \textit{CagA} and Western \textit{CagA}. The former type prevails in East Asia and is more closely associated with gastric cancer. The present study used full sequences of the \textit{cag} gene and \textit{CagA} protein of 22 \textit{H. pylori} strains in gastric cancer and peptic ulcer patients from Southern Vietnam to make a comparison of genetic homology among Vietnamese strains and between them and other strains in East Asia. A phylogenetic tree was constructed based on full amino acid sequences of 22 Vietnamese strains in accordance with 54 references from around the world. The \textit{cag} gene was found in all Vietnamese \textit{H. pylori} strains. Twenty-one of 22 (95.5%) strains belonged to the East Asian type and had similar characteristics of amino acid sequence at the carboxyl terminus to other strains from the East Asian region. From evidence of East Asian \textit{CagA} and epidemiologic cancerous lesions in Vietnam, \textit{H. pylori}-infected Vietnamese can be classified into a high-risk group for gastric cancer, but further studies on the interaction among environmental and virulence factors should be done. Finally, phylogenetic data support that there is a Japanese subtype in the Western \textit{CagA} type.

\textit{Helicobacter pylori} is a gram-negative bacterium that infects about 50% of the world’s population. Infections with \textit{H. pylori} can result in chronic active gastritis and are a risk factor for peptic ulcers, gastric cancer and gastric mucosa-associated lymphoid tissue lymphoma (40, 52, 53, 56). The prevalence of \textit{H. pylori} infections is not the same in different parts of the world. Recent studies reported that humans actually acquired \textit{H. pylori} in the early days of their history, long before the migration of modern humans out of Africa, and the diverse distribution of \textit{H. pylori} today is associated with waves of human migration in the past (19, 32, 36, 61, 62). The rate of \textit{H. pylori} infections is high in Africa, East Asia, and South Asia; however, the incidence of gastric cancer is high in East Asia but not in South Asia or Africa; this may be explained partly by the diversity of \textit{H. pylori} strains in these regions (62).

In cases of gastroduodenal diseases, especially gastric cancer, the pathogenesis involves three major factors: \textit{H. pylori} virulence factors, host factors, and environmental factors (1, 10, 14, 16, 21, 34, 57). Two \textit{H. pylori} virulence factors that have been focused on in many studies all around the world are \textit{VacA} and \textit{CagA}. \textit{VacA} is encoded by the \textit{vacA} gene and found in all \textit{H. pylori} strains. The \textit{vacA} gene is classified into three major genotypes: s1/m1, s1/m2, and s2/m2. s1/m1 strains produce higher levels of cytotoxin than s1/m2 or s2/m2 strains (2, 3). \textit{CagA} is encoded by \textit{cagA}, located in the \textit{cagPAI} (pathogenicity island) region of the \textit{H. pylori} genome, and the presence of this protein is a marker of \textit{cagPAI}. Unlike the \textit{vacA} gene, the \textit{cagA} gene has been found in only 50% to 70% of \textit{H. pylori} strains infecting Western populations; however, 80% to 100% of \textit{H. pylori} strains from East Asia have the \textit{cagA} gene, with the \textit{cagPAI} region, in their genome (17, 49, 51, 61, 64, 65). Patients infected with \textit{H. pylori} strains possessing \textit{CagA} were at greater risk of developing gastric adenocarcinoma than those uninfected or infected with \textit{CagA}-negative strains (11, 41). \textit{CagA} is injected directly from \textit{H. pylori} into intragastric epithelial cells via the type IV secretion system. The injected \textit{CagA} mimics eukaryotic adaptor proteins that recruit multiple host signaling factors into protein complexes that target cellular junctions, cell proliferation, and actin- cytoskeletal rearrangements (54).

The translocated \textit{CagA} undergoes tyrosine phosphorylation by Src and Abl family kinases and binds to SHP-2 in the human gastric mucosa (45, 47, 48). The repeated EPIYA (Glu-Pro-Ile-Tyr-Ala) motifs at the carboxyl-terminal end of \textit{CagA} are the targets of tyrosine phosphorylation (8, 22, 24, 38, 63). \textit{CagA} multimerization plays an important role in the pathophysiological activity of \textit{CagA} in disturbing host cell functions via

\textsuperscript{V} Published ahead of print on 21 October 2009.

\textsuperscript{*} Corresponding author. Mailing address: Department of Gastroenterology, Kobe University Graduate School of Medicine, Kobe, Japan, 7-5-1 Kusunoki-cho, Chuo-ku, Kobe 650-0017, Japan. Phone: 81-78-382-5774, Fax: 81-78-382-6039. E-mail: azumat@med.kobe-u.ac.jp.
SHP-2 deregulation, and EPIYA polymorphisms of CagA greatly influence the magnitude and duration of phosphorylation-dependent CagA activity (37, 42). SHP-2, like its Drosophila melanogaster homolog Corkscrew, is known to play an important role in the mitogenic signal transduction that connects receptor tyrosine kinases and ras (20). It is possible that deregulation of SHP-2 by the translocation of CagA plays a role in the acquisition of a cellular-transformed phenotype at a relatively early stage in the carcinogenesis of gastric carcinoma. A recent study on generating CagA in transgenic mice has provided the first direct evidence of the role of CagA as a bacterium-derived oncprotein that acts in mammals and further indicates the importance of tyrosine phosphorylation, which enables CagA to deregulate SHP-2, in the development of H. pylori-associated neoplasms (39). Based on characteristics of the EPIYA motif, the H. pylori CagA protein could be divided into a Western type and an East Asian type. The East Asian CagA protein exhibits stronger SHP-2-binding activity and is more pathogenic than the Western CagA protein in H. pylori-infected patients (4, 7, 23, 37). Clinical data from East Asia, Japan, and South Korea indicated that the East Asian form of CagA was more closely related to persistent active inflammation, atrophic gastritis, and a higher risk of gastric cancer than the Western form (5, 6, 26, 44). It is quite clear that by studying the characteristics of H. pylori strains, especially the cagA gene and CagA protein, one can categorize H. pylori-infected patients into those at high risk of developing gastric cancer and those not.

Vietnam is a developing country located in Southeast Asia. However, according to historical and migrational evidence, the Vietnamese are more closely related to people from East Asia than people from South Asia. Gastric cancer is one of the five most common cancers in Vietnam, including lung, stomach, liver, recto-colon and naso-pharynx cancer in males. In females, it ranks third behind breast and cervical-uterine cancer. The prevalence of gastric cancer in Northern Vietnam is as high as that in China or Korea. Its prevalence in Southern Vietnam is lower but still higher than that in Thailand and South Asia (46, 50). A few studies reported that cagA was found in nearly 100% of H. pylori-infected Vietnamese (60, 61), but no studies have examined the type of CagA protein or the full sequence of cagA in Vietnamese patients. The present study reports the diverse characteristics of cagA and classification of CagA in H. pylori-infected patients from Southern Vietnam based on the full genomic cagA sequence.

MATERIALS AND METHODS

H. pylori strains. Patients had abdominal symptoms and were clinically examined at the Department of Gastroenterology and further underwent upper gastrointestinal endoscopy in the Endoscopic Unit, Cho Ray Hospital, Ho Chi Minh City (formerly Saigon), Vietnam. All patients were unrelated and Vietnamese in origin. Patients who had received nonsteroidal anti-inflammatory drugs were excluded from the study, and none of the patients had recently been prescribed antibiotics. Two biopsy specimens from the greater curvature of the gastric antral site and one from the greater curvature of the gastric fundic mucosa were obtained during the endoscopic procedure. Immediately, one specimen from the gastric antral mucosa was used for a rapid urea test. Afterward, one specimen each from the antral and fundic mucosa, respectively, was transferred for H. pylori culture at the Department of Microbiology, Cho Ray Hospital, Ho Chi Minh City, Vietnam. Only positive H. pylori strains were used in the cagA sequencing analysis. For patients who had gastric ulcers, additional biopsy specimens at the margin of the lesions and one at each site of antral and fundic mucosa were obtained for further histological examination. Biopsy specimens were fixed in 10% buffered formalin for the differential diagnosis of ulcers from gastric carcinoma. A total of 22 H. pylori strains from 22 patients were studied: 11 patients with gastric cancer, 2 patients with gastric ulcers, and 9 patients with duodenal ulcers. Fourteen patients were males and 8 were females, and the mean age was 53.09 ± 14.76 years (range, 34 to 84 years). This work was performed according to the principles of the Declaration of Helsinki and approved by the Ethics Committee.

Histological analysis. Biopsy specimens were embedded in paraffin and stained with hematoxylin-eosin. The specimens were evaluated by an experienced histopathologist who had no knowledge of the clinical-endooscopic diagnosis. The histological slides were then examined by an independent expert, who had no information on the patients, for confirmation of the diagnosis.

Isolation and culture of H. pylori. Gastric biopsy specimens from each patient were inoculated onto a Trypticase soy agar II–5% sheep blood plate and cultured for 3 to 5 days at 37°C under microaerobic conditions (5% O2, 15% CO2, and 80% N2). A single colony was picked from each primary culture plate, inoculated onto a fresh Trypticase soy agar II plate, and cultured under conditions described previously. H. pylori cells were harvested from each plate, transferred into a brucella broth liquid culture medium containing 10% fetal calf serum, and cultured for 3 days under the conditions described above. The liquid culture samples were stored at −80°C in 0.01 M phosphate-buffered saline containing 20% glycerol before being transferred to Japan. At the Department of Gastroenterology, Kobe University Graduate School of Medicine, the H. pylori strains were once again cultured as described above. Part of the liquid-cultured sample was stored at −80°C in 0.01 M phosphate-buffered saline containing 20% glycerol. DNA from each H. pylori isolate was extracted from the pellet of the liquid culture sample by the protease-phenol-chloroform method, suspended in 300 µl of a TE buffer (10 mM Tris-HCl 1 mM EDTA), and stored at 4°C for the PCR analysis and nucleotide sequencing.

PCR and nucleotide sequencing of the entire cagA gene. Primers for PCR amplification and direct sequencing of the entire coding region of cagA are listed in Table 1. The full-length cagA gene was amplified by PCR under the following conditions: 94°C for 1 min and 25 cycles of 94°C for 30s, 55°C for 30s, and 72°C for 3 min, followed by 72°C for 10 min. PCR products were determined by 2% agarose gel electrophoresis. The gels were stained with ethidium bromide to determine the size of the cagA gene. PCR products then were purified with Centricon-100 concentrator columns (Amicon, Beverly, MA). H. pylori DNA was sequenced using the dye terminator cycle sequencing kit (Applied Biosystems, Foster City, CA) in an ABI Prism 3100-Avant analyzer (Applied Biosystems) according to the manu-

TABLE 1. Oligonucleotide primers used for sequencing of the H. pylori cagA gene

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Corresponding DNA sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>cagA L2 (+)</td>
<td>5'-AAGGAGAAAACATGACTAACAACTATTGG-T3'</td>
<td>-11-19'</td>
</tr>
<tr>
<td>cagA L2 (-)</td>
<td>5'-CTCTTTAAGTTTGGAGACACCCTTGT-T3'</td>
<td>+4-3536</td>
</tr>
<tr>
<td>cagB1 (-)*</td>
<td>5'-CTGCGAAGATTGGTGGAAGA-3'</td>
<td>520-499</td>
</tr>
<tr>
<td>cagM2 (+)*</td>
<td>5'-ATACAGTCCTACCGCTGG-3'</td>
<td>754-772</td>
</tr>
<tr>
<td>cagA S2*</td>
<td>5'-GGCGATGTTGTTGCTGAGC-3'</td>
<td>976-1000</td>
</tr>
<tr>
<td>cagM1 (-)</td>
<td>5'-GTAGCCACATGGTCGCTTGT-T3'</td>
<td>1055-1032</td>
</tr>
<tr>
<td>cagC2 (+)*</td>
<td>5'-GAATTGTCGATAAACCTGA-3'</td>
<td>2059-2080</td>
</tr>
<tr>
<td>cagC3 (-)*</td>
<td>5'-TTTGTGTCGTCACCTGA-3'</td>
<td>2306-2286</td>
</tr>
<tr>
<td>cagC3 F1*</td>
<td>5'-AACCCCTGATGCTGCTAAC-3'</td>
<td>3210-3189</td>
</tr>
</tbody>
</table>

* Primer used for sequencing.

a Nucleotide positions in the HP0547 (cagA) gene of H. pylori 26695 (GenBank accession no. AE000569).

b Product size of 3,576 bp.
manufacturer's recommendations, using the primers marked in Table 1. The sequencing reaction was as follows: 96°C for 5 min and then 25 cycles of 96°C for 10s, 50°C for 5s, and 60°C for 4 min, followed by 4°C for 10 min. The full-length nucleotide sequence of \textit{cagA} was constructed and translated into the amino acid sequence with the standard code using GENETYX software version 7.0 (Genetyx Corporation, Japan).

**Phylogenetic analysis.** To clarify the phylogenetic relationship between the 22 Vietnamese \textit{H. pylori} strains of the present study and previously reported \textit{H. pylori} strains worldwide, sequences of 54 reference strains of the full-length \textit{cagA} gene in GenBank (accession numbers reported in Fig. 1) were downloaded. Of the 54 references, 32 were from Japan, including 14 strains with the Western type of \textit{CagA} (Okinawa strains) and 18 strains with the East Asian type (Fukui strains) reported previously (64), 5 were from Hong Kong (China) and China, 2 were from India (South Asia), 4 were from Australia, 4 were from Europe, and 7 were from America. Full-length sequences of the \textit{cagA} gene of Vietnamese \textit{H. pylori} strains and references were aligned using the CLUSTAL X program. Genetic distances were estimated by Kimura's two-parameter method (30). Phylogenetic trees were constructed by the neighbor-joining method (43). To confirm the reliability of the phylogenetic trees, bootstrap resampling and reconstruction with 1,000 replicates were used. These analyses were carried out using MEGA software version 3.0 (Genetyx Corporation, Japan).

**Phylogenetic analysis.** To clarify the phylogenetic relationship between the 22 Vietnamese \textit{H. pylori} strains of the present study and previously reported \textit{H. pylori} strains worldwide, sequences of 54 reference strains of the full-length \textit{cagA} gene in GenBank (accession numbers reported in Fig. 1) were downloaded. Of the 54 references, 32 were from Japan, including 14 strains with the Western type of \textit{CagA} (Okinawa strains) and 18 strains with the East Asian type (Fukui strains) reported previously (64), 5 were from Hong Kong (China) and China, 2 were from India (South Asia), 4 were from Australia, 4 were from Europe, and 7 were from America. Full-length sequences of the \textit{cagA} gene of Vietnamese \textit{H. pylori} strains and references were aligned using the CLUSTAL X program. Genetic distances were estimated by Kimura's two-parameter method (30). Phylogenetic trees were constructed by the neighbor-joining method (43). To confirm the reliability of the phylogenetic trees, bootstrap resampling and reconstruction with 1,000 replicates were used. These analyses were carried out using MEGA software version 3.0 (Genetyx Corporation, Japan).

**Phylogenetic analysis.** To clarify the phylogenetic relationship between the 22 Vietnamese \textit{H. pylori} strains of the present study and previously reported \textit{H. pylori} strains worldwide, sequences of 54 reference strains of the full-length \textit{cagA} gene in GenBank (accession numbers reported in Fig. 1) were downloaded. Of the 54 references, 32 were from Japan, including 14 strains with the Western type of \textit{CagA} (Okinawa strains) and 18 strains with the East Asian type (Fukui strains) reported previously (64), 5 were from Hong Kong (China) and China, 2 were from India (South Asia), 4 were from Australia, 4 were from Europe, and 7 were from America. Full-length sequences of the \textit{cagA} gene of Vietnamese \textit{H. pylori} strains and references were aligned using the CLUSTAL X program. Genetic distances were estimated by Kimura's two-parameter method (30). Phylogenetic trees were constructed by the neighbor-joining method (43). To confirm the reliability of the phylogenetic trees, bootstrap resampling and reconstruction with 1,000 replicates were used. These analyses were carried out using MEGA software version 3.0 (Genetyx Corporation, Japan).

**Data deposition.** The full sequences of \textit{cagA} in \textit{H. pylori} strains from Southern Vietnam have been deposited in GenBank under the accession numbers (FJ798952 to FJ798973) shown in Table 2.

**RESULTS**

**Characteristic diversity of \textit{CagA}**. The alignment of the deduced amino acid sequence in the 3' region of the \textit{cagA} gene among strains is shown in Fig. 1. The first and second EP-IYa motifs (designated EP-IYa-A and EP-IYa-B, respectively) are present in almost all \textit{CagA} proteins, whereas two major subtypes are observed around the third EP-IYa motif. Western strains have the WSS (Western \textit{CagA}-specific, HP-2 binding sequence) in the third EP-IYa motif (designated EP-IYa-C) (23). The WSS contains D1, D2, and D3 motifs, as defined by Covacci et al. (15), or R1 and WSR (Western specific region) sequences as defined by Yamaoka et al. (58). Strain 26695 (AE000569) with the representative Western \textit{CagA} type (EP-IYa-ABC) is included in Fig. 1a. VN28 \textit{CagA} was classified as type “ABC” since it had two repeats of WSS. The amino acid sequence of \textit{CagA} from \textit{H. pylori} isolated in East Asian countries is quite different from that of the \textit{Western CagA} type. Predominant \textit{East Asia CagA} proteins do not have the WSS, but instead possess a distinct sequence which we designated.
ESS (East Asian CagA-specific, SHP-2-binding sequence) in the corresponding region (23). The ESS contains a JSR (Japanese specific region) sequence previously defined by Yamaoka et al. (58) and also possesses an EPIYA motif, denoted EPIYA-D. F32 (accession no. AF202972), with the representative East Asian type (EPIYA-ABD) is also included in Fig. 1a. VN7-CagA was classified as type “ABD” since it had one ESS. Figure 1b gives amino acid sequences and repeated EPIYA motifs in East Asian CagA. Variations are frequently found between the EPIYA-A motif and EPIYA-B motif. Between the EPIYA-B and EPIYA-D motifs, excluding strain VN20 with the EPIYA-ABD motif and strain VN16 with the EPIYA-ABD motif in Fig. 1b, the sequence of 48 amino acids indicated by double-underlined letters on panel b is quite conserved (QVAKKVSAKIDOLNEATSAINRKRDRNKIAS AGKGVGGFSFGAGGAGGSAGQAS), with only some variation, except in strains VN16 and VN20. In the EPIYA-D region, 8 of 21 (38.1%) strains have an additional residue, a glutamine (E), upstream of the EPIYA-D motif. This variation is specific for Vietnamese strains; however, there was no association between the variation and clinical outcome.

The Vietnamese H. pylori strains and characteristics of their cagA gene are listed in Table 2. The CagA protein in 21 of 22 (95.5%) strains belonged to the East Asian type. EPIYA-ABD motifs were found in 18 strains, and the EPIYA-ABD motif, EPIYA-AB'D motif, and EPIYA-AB'D motif were each found in one strain. As reported previously (64), B' represents EPIYT-B, in which the A (alanine) residue is replaced by T (threonine), and B" represents ESIYA-B, in which the P (proline) residue is replaced by S (serine). Only one H. pylori strain in the study had the Western type of CagA with the EPIYA-ABCC motif, meaning that it has two EPIYA-C motifs. The shortest sequence of the full-length coding region of cagA was 3,480 nucleotides in a strain (VN7) with the EPIYA-ABD motif, and the longest was 3,657 nucleotides in a strain (VN28) with the EPIYA-ABCC motif. The predicted size of CagA in 22 strains ranged from 1,160 to 1,219 amino acids.

The homology of nucleotide sequence and amino acid sequence is shown in Table 3. Within the East Asian CagA group, excepting strain VN32, the highest degree of homology was 99% and the lowest were 93% and 93.1% in full nucleotide sequences and in full amino acid sequences, respectively. Strain VN32 was located in between the East Asian CagA group and Western CagA group in the phylogenetic tree, as reported in Fig. 2, having levels of homology of 92.8% and 92.9% and the lowest degrees of homology—90.7% and 91%—compared with the other strains in the East Asian CagA group for nucleotide sequences and amino acid sequences, respectively. In a comparison between East Asian CagA strains and the Western CagA strain VN28, the highest degrees of homology in nucleotide and amino acid sequences were 90% and 90.3%, respectively, and the lowest were 83.8% for nucleotide sequences and 84.5% for amino acid sequences. A Japanese strain, F32, with the representative East Asian type (EPIYA-ABD) and strain 26695 with the representative Western CagA type (EPIYA-ABC) are included in Table 3. To compare the Japanese strain F32 and strain 26695 with Vietnamese strains, the highest and lowest degrees of homology for nucleotide sequences were 96.3% and 94.4% and 84.9% and 83.8%, respectively; those for amino acid sequences were 96.3% and 94.5% and 85.6% and 84.5%, respectively.

Phylogenetic tree analysis. To more precisely determine the genotype of the CagA protein in Vietnamese H. pylori strains, a phylogenetic analysis based on full CagA amino acid sequences in 22 Vietnamese strains plus 54 reference strains from GenBank was carried out (Fig. 2). Twenty of 22 (90.9%) Vietnamese strains belonged to the East Asian type, in accordance with 18 Japanese strains, mainly from Fukui Province,
and 5 strains from China. One of 22 (4.5%) Vietnamese strains (VN28) belonged to the Western type. From the results in Fig. 2, it is clear that all East Asian CagA strains from Japan, Vietnam, and China are located in the same cluster, except for one strain from Vietnam. However, all East Asian CagA Vietnam strains were integrated closely at one site of the cluster and Japanese strains at another site of the cluster. VietnamHP-No32 was located in between the East Asian and Western cluster, though it has the EPIYA-ABD motif. Also even in Western type CagA strains, most strains from Japan (Okinawa strains) formed a different cluster from strains isolated in the West. We therefore believe that there is a Japanese subtype in the Western CagA type (J-Western CagA subtype).

**DISCUSSION**

*H. pylori* can survive long term outside of a host, suggesting it has flexible evolutionary strategies in accordance with environmental modifications and enabling bacterial transmission with important clinical manifestations (13). The geographical distribution of *H. pylori* has been associated with the migration of modern humans, and the prevalence of gastric cancer in the world not only is related to host and environmental factors but also is related to the diversity of *H. pylori* genomic genotypes (16, 19, 32, 36, 60, 62). *H. pylori* CagA is an oncoprotein that acts in mammals (39) and has a major role in the pathogenesis of gastroduodenal diseases in *H. pylori*-infected patients, especially in cases of gastric cancer (11, 14, 17, 38, 40, 41). Today, the so-called “paradox of gastric cancer” is particularly evident in Asian countries (46). The prevalence of *H. pylori* infections in Japan, Korea, and China is no higher than that in South Asian nations, but the incidence of gastric cancer is significantly higher. This can be partly explained by the diversity of *H. pylori* strains among Asian nations, especially the polymorphisms of EPIYA motifs in the carboxy-terminal region of CagA, the Western CagA type (EPIYA-ABC), and East Asian CagA type (EPIYA-ABD) (4, 6, 7, 59–62, 64). East Asian CagA is more pathological than Western CagA (7, 22, 37, 44). *H. pylori* strains containing many EPIYA-C motifs at the 3′ end of Western CagA could be more likely to cause gastric cancer than strains with fewer EPIYA-C motifs (9).

The prevalence of gastric cancer in Vietnam is higher than that in other countries in the region such as Thailand, Indonesia, the Philippines and also higher than that in South Asian nations such as India, Pakistan, and Nepal (7, 46, 50). The prevalence rates of gastric cancer in males and females in Northern Vietnam are 34.5/100,000 and 16.4/100,000, respectively. The prevalence rates in Southern Vietnam are 16.5/100,000 and 7.5/100,000, respectively (50). This means that the prevalence of gastric cancer in Northern Vietnam may be as high as that in China, Korea, and some regions of Japan. The prevalence of gastric cancer in Southern Vietnam is lower than that in Northern Vietnam, but higher than that in other Southeast Asian nations and in South Asia. According to studies of cancerous lesions among Asian-Pacific ethnicities in the United States, Vietnamese Americans have a high prevalence of gastric cancer (33, 35). Previous studies reported that the cagA gene was common in *H. pylori* from Vietnamese patients both in Vietnam and overseas (60, 61). However, to our knowledge, the present study is the first not only to detail polymorphisms at the carboxy terminal end of CagA but also to fully sequence the cagA gene and CagA protein in *H. pylori* strains from Vietnamese patients. It was found that all of the *H. pylori* strains having cagA in the genome and most others (95.5%) as well had the East Asian form of CagA. As in East Asian countries such as Japan, Korea, and China, the data from the present study and previous studies (60, 61) indicate that most CagA proteins in Vietnamese *H. pylori* strains are of the East Asian type. The full-length open reading frame of cagA in all Vietnamese *H. pylori* strains is over 3,400 (3,480 to 3,657) nucleotides long, with a predicted protein of 1,160 to 1,219.
amino acids. In the present study, in the East Asian CagA group, except for one strain located between the East Asian type and Western type in the phylogenetic analysis, Vietnamese \textit{H. pylori} strains had a high degree of homology in both the nucleotide sequence of \textit{cagA} and the amino acid sequence of CagA (93 to 99\%). Consistent with previous studies of populations from East Asia (26, 27, 51), the present study found that the amino acid sequences of the EPIYA-B and EPIYA-D motifs are quite conserved. A sequence of 18 amino acid sequences (AINRKI DRINKIASAGKG) in the conserved region was designated East Asian-specific immunogens for immunohistochemical genotyping of \textit{cagA} in previous studies (27, 51). Most of the Vietnamese \textit{H. pylori} strains with the East Asian form of CagA have similar characteristics of amino acid sequences in the region to other strains from the East Asian region (Japan and Korea), so the immunohistochemical genotyping for CagA protein reported in previous studies (27, 51) can be applied to \textit{H. pylori} strains from Vietnam. However, in the present study, CagA in Vietnamese \textit{H. pylori} strains was classified based on the full sequence of the \textit{cagA} gene. This method not only is a very good way of genotyping but also can count how many

FIG. 2. A phylogenetic tree constructed based on the complete amino acid sequence of \textit{H. pylori} CagA protein in 22 isolates from Southern Vietnam (VietnamHP no. 7, 6, 17, 16, 4, 23, 27, 33, 2, 5, 15, 31, 24, 25, 20, 3, 8, 36, 22, 14, 32, 28 [given in the order shown in Tables 2 and 3]) compared with complete amino acid sequence of \textit{H. pylori} CagA protein in 54 isolates from the GenBank database, including Japan (32 strains, 14 with the Western CagA type from Okinawa and 18 with the East Asian CagA type from Fukui), China (5 strains), India (2 strains), Australia (4 strains), Europe (4 strains), and America (7 strains). The length of the horizontal bar indicates the number of amino acid substitutions per site.
EPIYA motifs are present at the carboxyl-terminal end of CagA, since the number of EPIYA motifs is associated with clinical outcome in *H. pylori*-infected patients (9, 23).

An increasing number of studies have carried out a phylogenetic analysis based on nucleotide sequences of *H. pylori* strains in different parts of the world. However, when examining the *cagA* gene, most phylogenetic studies have analyzed the 3′ region or S′ region only, with or without nucleotide sequences of the vacA gene or other housekeeping genes (12, 18, 28, 29, 61). Some studies from Japan that carried out a phylogenetic analysis based on full genomic sequences of the *cagA* gene or *cagPAI* region were reported, but unfortunately they focused mainly on intergroup examinations of Japanese *H. pylori* strains (5, 64) or had very few *H. pylori* reference strains from other countries (25). The phylogenetic tree in the present study was based on full nucleotide sequences of *cagA* in 22 strains from Vietnam in comparison with 54 reference strains from East Asia and many parts of the world. So the present study has analyzed the largest number of full nucleotide sequences of *cagA* published to date. The data show that except for one East Asian CagA strain from Vietnam located between East Asian CagA and Western CagA, all East Asian CagA strains (Japan, Vietnam, China, and Hong Kong) are found in the same cluster with a slightly different location of Vietnamese strains and Japanese strains in the cluster. Vietnamese strains are very closely located at one site, and Japanese strains are very closely located at another site. As the number of *H. pylori*-infected patients is huge, we hope that in the future, more full-nucleotide sequences of the East Asian CagA type will be published, and more diverse characteristics of East Asian CagA will be found. It is interesting that in Western type CagA strains, most strains from Japan (Okinawa strains) formed a different cluster from strains isolated in the West. We therefore believe that there is a Japanese subtype in the Western CagA type (J-Western CagA). Since *H. pylori* with the Western CagA type is found in many regions of the world, it is better to carry out further studies on different features of intergroups of Western CagA protein involved in the pathogenesis of gastroduodenal diseases and gastric cancer.

Evidently, East Asian CagA binds more strongly to SHP-2 than Western CagA (6, 23, 37), and *H. pylori*-infected patients with East Asian CagA have a higher risk of gastric cancer (4, 7, 26, 44). The present study found that CagA is present in all *H. pylori* strains from Southern Vietnam and mostly (95.5%) of the East Asian type; this observation supports the notion that *H. pylori*-infected Southern Vietnamese can be considered at high risk of developing gastric cancer. However, the prevalence rates of *H. pylori* infections are ~50% in Southern Vietnam and 70% in Northern Vietnam, yet the prevalence rates of gastric cancer are different between these two regions (50). This cannot simply be explained by the prevalence of *H. pylori* infection. There are about 54 ethnic groups in Vietnam, but more than 84% of the population are of the majority King’s group. Most data in studies of *H. pylori*-infected Vietnamese are from King’s people. Further studies examining environmental factors in *H. pylori*-infected Vietnamese are important. Weather and humidity, lifestyle, and dietary factors such as kinds of vegetables and fruits are quite different between Northern and Southern Vietnam.

In conclusion, the present study found that *cagA* is present in all *H. pylori* strains from Southern Vietnam. Most of the CagA proteins in Vietnamese strains of *H. pylori* are of the East Asian type and have three EPIYA motifs (A, B, and D) in the carboxyl-terminal region. Vietnamese strains contain the same characteristic diversity of amino acid sequence between EPIYA-B and EPIYA-D motifs as *H. pylori* strains from East Asia, excluding a very small proportion with very diverse amino acid sequences in this region. Finally, the phylogenetic analysis supported that there is a Japanese subtype in the Western CagA type (J-Western CagA).

ACKNOWLEDGMENTS

We thank Sueyoshi Kumiko for technical assistance. This work was supported by a Grant-in-Aid for G-COE Research and by a Grant-in-Aid for Scientific Research (B) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

REFERENCES


