Distinct Diversity of *vacA*, *cagA*, and *cagE* Genes of *Helicobacter pylori* Associated with Peptic Ulcer in Japan

Shiho Yamazaki,¹ Akiyo Yamakawa,¹ Tomoyuki Okuda,¹ Masahiro Ohtani,² Hiroyuki Suto,² Yoshiyuki Ito,¹ Yukinao Yamazaki,² Yoshihide Keida,³ Hideaki Higashi,⁴ Masanori Hatakeyama,⁴ and Takeshi Azuma^{5*}

Second Department of Internal Medicine, Faculty of Medical Sciences, University of Fukui, Fukui, Japan¹; Department of Endoscopic Medicine, Faculty of Medical Sciences, University of Fukui, Fukui, Japan²; Division of Internal Medicine, Okinawa Chubu Hospital, Okinawa, Japan³; Division of Molecular Oncology, Institute for Genetic Medicine and Graduate School of Science, Hokkaido University, Sapporo, Japan⁴; and Frontier Medical Science in Gastroenterology, International Center for Medical Research and Treatment, Kobe University School of Medicine, Kobe, Japan⁵

Received 24 November 2004/Returned for modification 23 January 2005/Accepted 17 May 2005

Colonization of the stomach mucosa by *Helicobacter pylori* is a major cause of acute and chronic gastric pathologies in humans. Several *H. pylori* virulence genes that may play a role in its pathogenicity have been identified. The most important determinants are *vacA* and *cagA* in the *cag* pathogenicity island (*cagPAI*) genes. In the present study, to consider the association of molecular genetics between *vacA* and the *cagPAI* regarding clinical outcome, we selected *H. pylori* strains with various genotypes of *vacA* in Japan and sequenced full-length *vacA*, *cagA*, and *cagE* genes. Sequencing of *vacA* and *cagA* genes revealed variable size, whereas the *cagE* gene was well conserved among strains. Each of the phylogenetic trees based on the deduced amino acid sequences of VacA, CagA, and CagE indicated that all three proteins were divided into two major groups, a Western group and an East Asian group, and the distributions of isolates exhibited similar patterns among the three proteins. The strains with s2 and s1a/m1a *vacA* genotypes and the Western-type 3' region *cagA* genotype were classified into the Western group, and the strains with the s1c/m1b *vacA* genotype and the East Asian-type 3' *cagA* genotype were included in the East Asian group. In addition, the prevalence of infection with the Western group strain was significantly higher in patients with peptic ulcer (90.0%, 9/10) than in patients with chronic gastritis (22.7%, 5/22) ($\chi^2 = 12.64$, $P = 0.00057$). These data suggest that the molecular genetics of *vacA* and *cagPAI* are associated and that the Western group with *vacA* and *cagPAI* genes is associated with peptic ulcer disease.

Helicobacter pylori, a spiral, gram-negative, microaerophilic bacterium, colonizes at least half of the world's population and is recognized as a major cause of chronic gastritis and peptic ulcer and as an important risk factor for gastric cancer (24, 30, 33). On the basis of various epidemiological studies, *H. pylori* was classified as a class I carcinogen in humans by a working group of the World Health Organization International Agency for Research on Cancer (19).

Several *H. pylori* virulence genes that may play a role in its pathogenicity have been identified. Of these, the most important determinants are *vacA* and *cagA* genes. *H. pylori* strains have been divided into two broad families, type I and type II, which are based on whether or not they possess the *vacA* and *cagA* genes. Type I strains have the ability to produce VacA and CagA, while type II strains lack that ability (36). Type I strains are regarded as having greater pathogenicity and potential to cause development of disease. It is also known that *H. pylori* can be divided into distinct populations with different geographical distributions (13, 15).

VacA protein induces the formation of intracellular vacu-

oles in eukaryotic cells in vitro. The *vacA* gene contains at least two variable parts. Two sequence families, called s1 and s2, have been identified in different isolates. Among type s1 strains, subtypes s1a, s1b, and s1c have been identified. The m-region (middle) occurs as the m1 or m2 allelic type. Among type m1 strains, subtypes m1a and m1b have been identified (5, 20, 34, 36). Production of vacuolating cytotoxin is related to the mosaic structure of *vacA*. In general, type s1/m1 and s1/m2 strains produce high and moderate levels of toxin, respectively, whereas s2/m2 strains produce little or no toxin (5).

CagA protein, which is encoded by the *cagA* gene, is a highly immunogenic protein. It is thought that *H. pylori* strains possessing *cagA* are associated with significantly increased risk for the development of atrophic gastritis and gastric cancer (11, 18, 26). The *cagA* gene is located at one end of a 40-kb DNA insertion called the *cag* pathogenicity island (*cagPAI*) and may have originated from a non-*Helicobacter* source (12). The *cagPAI* contains 31 putative genes, 6 of which are thought to encode components of a bacterial type IV secretion system (2, 13). Recent studies have indicated that CagA is directly injected into epithelial cells via the type IV secretion system and undergoes tyrosine phosphorylation in the host cells (4, 9, 23, 27, 28). Furthermore, it was recently confirmed (16) that phosphorylated CagA forms a physical complex with SHP-2 (Src homology 2 domain-containing protein tyrosine phosphatase), which is known to play a positive role in mitogenic signal

* Corresponding author. Mailing address: Frontier Medical Science in Gastroenterology, International Center for Medical Research and Treatment, Kobe University School of Medicine, Kusunoki-cho, Chuo-ku, Kobe 650-0017, Japan. Phone: 81-78-382-6308. Fax: 81-78-382-6309. E-mail: azumat@med.kobe-u.ac.jp.

TABLE 1. Oligonucleotide primers used in typing of *H. pylori vacA* alleles

Region	Primer	Sequence ^a	Size and location of PCR product
s1a	vacA s1a-F VA1-R	5'-CTC TCG CTT TAG TAG GAG C-3' 5'-CTG CTT GAA TGC GCC AAA C-3'	213 bp (843–1055) ^b
s1b	SS3-F VA1-R	5'-AGC GCC ATA CCG CAA GAG-3'	187 bp (869–1055) ^c
s1c	vacA s1c-F VA1-R	5'-CTC TCG CTT TAG TGG GGY T-3'	213 bp (843–1055) ^c
s2	SS2-F VA1-R	5'-GCT AAC ACG CCA AAT GAT CC-3'	199 bp (433–631) ^d
m1a	VA3-F VA3-R	5'-GGT CAA AAT GCG GTC ATG G-3' 5'-CCA TTG GTA CCT GTA GAA AC-3'	290 bp (2741–3030) ^b
m1b	VAm-F3 VAm-R3	5'-GGC CCC AAT GCA GTC ATG GAT-3' 5'-GCT GTT AGT GCC TAA AGA AGC AT-3'	291 bp (2741–3031) ^c
m2	VA4-F VA4-R	5'-GGA GCC CCA GGA AAC ATT G-3' 5'-CAT AAC TAG CGC CTT GCA C-3'	352 bp (976–1327) ^c (2284–2635) ^d

^a Y, C or T.^b Nucleotide positions in the *vacA* gene of *H. pylori* 60190 (GenBank accession no. HPU05676) (14).^c Corresponding to nucleotide positions of *H. pylori* 60190.^d Nucleotide positions in the *vacA* gene of *H. pylori* Tx30a (GenBank accession no. HPU29401) (5).^e Nucleotide positions in the *vacA* gene of *H. pylori* 87–203 (GenBank accession no. HPU05677) (14).

transduction (39), and deregulates its enzymatic activity. Deregulation of SHP-2 by CagA may induce abnormal proliferation and movement of gastric epithelial cells. On the basis of the sequence constituting the SHP-2 binding site, CagA proteins can be subclassified into Western and East Asian types. The East Asian-type CagA possesses stronger SHP-2 binding and transforming activities than the Western-type CagA (17). It has also been reported that large sequence differences distinguish the *cagA* gene fragments from Asian strains and other strains (1, 37).

It is thought that there is a genetic linkage between *vacA* and *cagA* allelic geographic variations (35), although these genes are located separately on the genome (3, 31). Therefore, in this study, to consider the association of the molecular genetics of *vacA* and the *cagPAI* regarding clinical outcome, we selected *H. pylori* strains according to various genotypes of *vacA* and sequenced full-length *vacA*, *cagA*, and *cagE*, which is located near *cagA* in the *cagPAI*. The *cagE* gene product is a component of the type IV secretion system (13). Furthermore, we analyzed the phylogenetic relationship among the entire *VacA*, *CagA*, and *CagE* amino acid sequences.

MATERIALS AND METHODS

***H. pylori* strains.** Two different areas of Japan were selected as sources for strains. Fukui is a typical rural prefecture located on the central Japanese mainland (Honshu), while Okinawa consists of islands in the southwestern part of Japan. These two areas are separated by more than 1,300 km. The clinical outcomes of *H. pylori* infections are quite different between Fukui and Okinawa. The prevalence of atrophic gastritis, a precursor lesion of gastric cancer, is more frequent in Fukui, and the death rate from gastric cancer is more than 2.4 times higher in Fukui than in Okinawa. In contrast, the prevalence of duodenal ulcer is more frequent in Okinawa, and the ratio of incidence of duodenal ulcer and gastric ulcer is more than 1.8 times higher in Okinawa than in Fukui. A total of 220 *H. pylori* clinical isolates (115 Fukui strains and 105 Okinawa strains) were obtained during upper gastroduodenal endoscopy at the Second Department of

Internal Medicine, Faculty of Medical Sciences, University of Fukui, and Okinawa Chubu Hospital, Okinawa, respectively. The 115 patients in Fukui consisted of 41 patients with chronic gastritis (22 men and 19 women; mean age, 57.7 years), 26 with gastric cancer (14 men and 12 women; mean age, 60.2 years), 26 with duodenal ulcer (17 men and 9 women; mean age, 49.9 years), 18 with gastric ulcer (10 men and 8 women; mean age, 55.3 years), and 4 with gastroduodenal ulcer (4 men; mean age, 58.5 years). The 105 patients in Okinawa consisted of 58 patients with chronic gastritis (20 men and 38 women; mean age, 59.6 years), 4 with gastric cancer (3 men and 1 woman; mean age, 70.3 years), 24 with duodenal ulcer (20 men and 4 women; mean age, 53.8 years), 18 with gastric ulcer (11 men and 7 women; mean age, 59.2 years), and 1 with gastroduodenal ulcer (1 man; age, 40 years).

***H. pylori* culture conditions.** 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, 5% CO₂, 90% N₂). A single colony was picked from each primary culture plate, inoculated onto a fresh trypticase soy agar II plate, and cultured under the conditions described above. *H. pylori* cells were harvested from each plate, transferred into a brucella broth liquid culture medium containing 10% fetal calf serum, and cultured for 24 h with agitation under the same conditions described above. A part of the bacterial suspension 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 bacterial suspension by use of the protease/phenol-chloroform method, suspended in TE buffer (10 mM Tris-HCl, pH 8.0, and 1 mM EDTA), and stored at 4°C until PCR amplification was performed.

PCR-based typing of *vacA*. Genotyping of the *vacA* gene was performed by PCR amplification in the s-region (the signal peptide-encoding region) and the m-region (middle region) described previously (5, 20, 34) using the primers shown Table 1. PCR products were separated by 2% agarose gel electrophoresis and examined under UV illumination.

Nucleotide sequencing of the entire *vacA*, *cagA*, and *cagE* genes. We selected 33 *H. pylori* strains (13 Fukui and 20 Okinawa strains) according to the *vacA* genotypes. The 13 patients in Fukui consisted of 10 with chronic gastritis, 1 with duodenal ulcer, 1 with gastric ulcer, and 1 with gastric cancer. The 20 patients in Okinawa consisted of 12 with chronic gastritis and 8 with duodenal ulcer. The 11 strains (F16, F17, F28, F32, F79, F80, OK101, OK107, OK109, OK112, and OK129) for which the entire *cagPAI* sequences were previously reported were included in this study (8). Primers for PCR amplification and direct sequencing of the entire coding regions of *vacA*, *cagA*, and *cagE* are shown in Table 2. The region containing full-length *vacA* was amplified by PCR under the following

TABLE 2. Oligonucleotide primers used for DNA sequencing of *H. pylori vacA*, *cagA*, and *cagE*

Gene	Primer ^a	Sequence ^a	Corresponding DNA sequence (size of PCR product)
<i>vacA</i>	VAS-1F	5'-AGC CGA TAG CAT CAG AGA AGA AC-3'	95–117 (4,281 bp) ^b
	VAS-11R	5'-TGT GGT GTA TGC GTT GTA GGG GTT-3'	4375–4352 ^b
	<i>vacA</i> s1a-F*	**	452–470 ^b
	SS3-F*	**	478–495 ^c
	<i>vacA</i> s1c-F*	**	452–470 ^c
	SS2-F*	**	433–452 ^d
	<i>vacA</i> S5*	5'-GCT AAC CGC ACC ACG AGA GT-3'	959–968 ^b
	<i>vacA</i> AS5*	5'-TTG TCT GTA ACG CCG CTA AAA-3'	2009–1999 ^b
	VA3-F*	**	2373–2392 ^b
	VAm-F3*	**	2373–2394 ^c
	VA4-F*	**	976–994 ^e
	VA3-R*	**	2663–2644 ^b
	VAm-R3*	**	2664–2642 ^c
	VA4-R*	**	1327–1309 ^e
	<i>vacA</i> S3*	5'-TAT TGA AAG CGT GTT TGA AT-3'	3069–3088 ^b
	<i>vacA</i> 3'F*	5'-CAT TGT GGG CGG TTT TGG AAG-3'	4156–4177 ^b
<i>cagA</i>	<i>cagA</i> L2(+)	5'-AAG GAG AAA CAA TGA CTA ACG AAA CTA TTG-3'	(–11)–19 (3,576 bp) ^f
	<i>cagA</i> L2(–)	5'-TCC TTT AAG ATT TTT GGA AAC CAC CTT TTG-3'	(+4)–3536 ^d
	<i>cagA</i> B1(–)*	5'-CTG CAA AAG ATT GTT TGG CAG A-3'	520–499 ^f
	<i>cagA</i> M2(+)*	5'-ATA CAA GGC TTA CCG CCT G-3'	754–772 ^f
	<i>cagA</i> S2*	5'-GGC AAT GGT GGT CCT GGA GCT AGG C-3'	976–1000 ^f
	<i>cagA</i> M1(–)*	5'-GTA GCC ACA TTG TCG CCT TGT TGG-3'	1055–1032 ^f
	<i>cagA</i> C2(+)*	5'-GAA TTG TCT GAT AAA CTT GAA A-3'	2059–2080 ^f
	<i>cagA</i> C2(–)*	5'-TTT GCT TGC GTT ACC TTG CTG-3'	2306–2286 ^f
	<i>cagA</i> C3(–)*	5'-GCG TAT GTG GCT GTT AGT AGC G-3'	3210–3189 ^f
	<i>cagA</i> 3'F1*	5'-AAA CCC TGA GTG GCT CAA GCT C-3'	3273–3294 ^f
<i>cagE</i>	<i>cagE</i> S1	5'-GAG CGG TAA GGT TTT GTT CGG TGA T-3'	(–164)–(–140) (3,315 bp) ^g
	<i>cagE</i> AS3	5'-CCA ACA AAC AAC GCT GCT TTC-3'	(+199)–(+179) ^g
	<i>cagE</i> AS4*	5'-CAA TGG GTG GGG AGT ATG TCA AGA-3'	1007–984 ^g
	<i>cagE</i> AS1*	5'-CCA ACG CAG CGA CTT TCT CTA TG-3'	1430–1408 ^g
	<i>cagE</i> AS5*	5'-TGC ATG GTG GGG TGA AAG AAG TTT A-3'	1850–1826 ^g
	<i>cagE</i> AS2*	5'-ATG GGG TGA TCC TTA CTA ACA ACT A-3'	2435–2411 ^g

^a An asterisk indicates that a primer was used only for DNA sequencing. Double asterisks indicate that primers from Table 1 were chosen according to s-region (forward primer) and m-region (forward and reverse primers) genotypes.

^b Nucleotide positions in the *vacA* gene of *H. pylori* 11638 (GenBank accession no. HPU07145) (21, 25).

^c Corresponding to nucleotide position of *H. pylori* 11638.

^d Nucleotide positions in the *vacA* gene of *H. pylori* Tx30a (GenBank accession no. HPU29401) (5).

^e Nucleotide positions in the *vacA* gene of *H. pylori* 87–203 (GenBank accession no. HPU05677) (14).

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

^g Nucleotide positions in the HP0544 (*cagE*) gene of *H. pylori* 26695 (GenBank accession no. AE000568).

conditions: 95°C for 5 min; 25 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 4.5 min; followed by 72°C for 7 min. The regions containing full-length *cagA* and *cagE* were amplified under the same conditions except for the extension time (72°C for 4.5 min): in this case, *cagA* for 4 min and *cagE* for 3.5 min. PCR products were then purified with Centricon-100 Concentrator columns (Amicon, Beverly, Massachusetts). DNA direct sequencing was performed using a BigDye Terminator v.3.1 cycle sequencing kit (Applied Biosystems, Foster City, California) and an ABI PRISM 3100-Avant genetic analyzer (Applied Biosystems) according to the manufacturer's recommendations. The full-length amino acid sequences of each gene were constructed and translocated from the nucleotide sequence and aligned and analyzed by use of GENETYX-Mac version 11.2.3 (Software Development, Tokyo, Japan).

Phylogenetic analysis. To clarify the phylogenetic relationships between Japanese isolates and previously characterized *H. pylori* strains from the West, sequences of *VacA*, *CagA*, and *CagE* were aligned (GENETYX-Mac). Phylogenetic trees were constructed using the unweighted pair group method using the same software. The previously published gene sequences of strain 26695 (GenBank accession number AE000598, *vacA* [HP0887]; AE000569, *cagA* [HP0547]; and AE000568, *cagE* [HP0544]), strain J99 (AE001511, *vacA* [jhp0819]; AE001483, *cagA* [jhp0495]; and AE001482, *cagE* [jhp0492]), NCTC11638 (HPU07145, *vacA*; AF282853, *cagA* and *cagE*), and NCTC11637 (AF049653, *vacA*; AF202973, *cagA*) were included in the analysis.

Statistical analysis. The associations between the diversity of *cagA* and *vacA* genes and clinical outcome were analyzed with the chi-square test and Fisher's exact probability test.

Nucleotide sequence accession numbers. The DNA sequences of *cagA*, *cagE*, and *vacA* of each strain characterized here were deposited in the GenBank database. Accession numbers are shown in Table 3.

RESULTS

Diversity of *vacA*. The list of *vacA* genotypes is shown in Table 4. We previously reported that the predominant *vacA* genotype in Japan was s1c/m1b (40). In Fukui, strains with the s1c/m1b genotype account for 81.7% of all strains, whereas in Okinawa, there are various *vacA* genotypes, although the predominant *vacA* genotype is also s1c/m1b (71.4%) (Table 5). In the present study, we selected 33 *H. pylori* strains according to the *vacA* genotypes. However, the s1b/m2 and s1c/m1a genotypes and hybrid m1/m2 genes were not present, although we checked a total of 220 strains (115 Fukui and 105 Okinawa strains) (Table 5).

TABLE 3. List of the nucleotide sequence accession numbers in the DDBJ database

Strain	<i>vacA</i>	<i>cagA</i>	<i>cagE</i>
F13	AB190958	AB090073	AB191062
F15	AB190959	AB090074	AB191063
F16	AB190960	AB090075	AB191064
F17	AB190961	AB090076	AB191065
F18	AB190962	AB090077	AB191066
F26	AB190963	AB090078	AB191068
F32	AB190964	AF202972	AB191071
F36	AF049642	AB090080	AB191072
F37	AF071095	AB190934	AB191073
F55	AF049632	AB090081	AB191076
F79	AF071097	AB090082	AB191077
F80	AB190965	AB090083	AB191078
F92	AB190966	AB090084	AB191079
OK101	AB190967	AB090085	AB191080
OK107	AB190968	AB090086	AB191081
OK109	AB190969	AB090087	AB191082
OK111	AB190970	AB190937	AB191083
OK118	AB190971	AB190939	AB191086
OK129	AB190972	AB090089	AB191087
OK130	AB190973	AB190940	AB191088
OK139	AB190974	AB190941	AB191090
OK144	AB190975	AB190942	AB191091
OK155	AB190976	AB190943	AB191092
OK158	AB190977	AB190944	AB191093
OK159	AB190978	AB190945	AB191094
OK160	AB190979	AB190946	AB191095
OK180	AB190981	AB190948	AB191096
OK181	AB190982	AB190949	AB191097
OK185	AB190983	AB190950	AB191098
OK187	AB190984	AB190951	AB191099
OK194	AB190985	AB190952	AB191100
OK204	AB190986	AB190954	AB191101
OK210	AB190988	AB190956	AB191102
26695	AE000598	AE000569	AE000568
J99	AE001511	AE001483	AE001482
NCTC11638	HPU07145	AF282853	AF282853
NCTC11637	AF049653	AF202973	AB191103

Sequencing of *vacA* genes revealed a variably sized ~3,867- to 3,981-bp open reading frame (ORF), encoding proteins of ~1,289 to 1,327 amino acids. The homology in the signal sequence region among F37 (s1a), F80 (s1b), F32 (s1c), and OK155 (s2) is shown in Fig. 1a and Table 6 (values above the dashes). The homology in the m-region (PCR fragment for genotyping of *vacA*) among F37 (m1a), F32 (m1b), and OK155 (m2) is shown in Fig. 1b and Table 6 (values below the dashes). In addition, the homologies of total amino acid residues and nucleotide sequences among F37, F80, F32, and OK155 are shown in Table 7. Furthermore, Atherton et al. (5) compared the *VacA* sequences between strain 60190 (m1) and strain Tx30a (m2) and found that there was only 59.0% amino acid identity in the middle of *VacA* (strain 60190, 244-amino-acid region, residues 509 to 752; strain Tx30a, 253-amino-acid region, residues 522 to 774). Therefore, we checked the homology of F32 (m1b) to OK210 (m2) in the m-region (F32, 244-amino-acid region, residues 512 to 755; OK210, 253-amino-acid region, residues 533 to 785), and it was found to be 58.9% (Fig. 2). Also, the homology of F32 (m1b) to F37 (m1a, 244-amino-acid region, residues 517 to 760) was 85.3%.

Diversity of CagA. The alignment of the deduced amino acid sequence in the 3' region of the *cagA* gene (CagA repeat

domain) for strains 26695 and F32, which have typical Western and East Asian CagA proteins, respectively, is shown in Fig. 3. We previously demonstrated that CagA can be divided into Western and East Asian types (6, 17). Briefly, tyrosine phosphorylation of CagA occurs at the unique Glu-Pro-Ile-Tyr-Ala (EPIYA) motifs present several times in the C-terminal region (10, 16, 29). These EPIYA motifs are involved in the interaction of CagA with SHP-2. The first and second EPIYA motifs (designated EPIYA-A and EPIYA-B, respectively) are present in almost all Western and East Asian CagA proteins, although the subsequent amino acid sequences are quite different between Western and East Asian CagA. The Western-type CagA possesses WSS (Western CagA-specific SHP-2 binding sequence), while the East Asian-type CagA possesses a distinct sequence, ESS (East Asian CagA-specific SHP-2 binding sequence), in the corresponding region. The EPIYA motifs included in WSS or ESS were designated EPIYA-C or EPIYA-D, respectively. Strain 26695 has a single WSS and is thus classified as the "A-B'-C" type, and F32 has a single ESS and is thus classified as the "A-B-D" type.

On the basis of the above typing, the list of CagA genotypes is shown in Table 4. "B'" means EPIYT-B, in which the A (alanine) residue is replaced by T (threonine); "B"" means ESIYA-B, in which the P (proline) residue is replaced by S (serine); "B"" means ESIYT-B, in which the P and A residues are replaced by S and T; and "A'" means NPIYA-A, in which the E (glutamic acid) residue is replaced by N (asparagine). Although the number of repeats of the EPIYA motif was different among strains, all strains can be divided into the Western (12 strains) or East Asian (20 strains) types, except for the OK181 strain, which had neither the WSS nor the ESS sequence and was classified as "A-B'." B' was characteristic of the Western type because it appeared in all strains which possessed Western-type CagA, except for NCTC11637, NCTC11638, OK155, and OK160 in the present study. OK155 and OK160 had B"" instead of B'. B" appeared in the four East Asian-type CagA-possessing strains (F26, F55, OK159, and OK204). A' appeared in the F37 strain only. Sequencing of *cagA* genes revealed a variable size of ~3,444- to 3,825-bp ORFs encoding proteins of ~1,148 to 1,275 amino acids. The homology of OK130 (Western type) to F32 (East Asian type) in total amino acid residues was found to be 80.3%, and in the CagA repeat domain (Fig. 3) (OK130, 109-amino-acid region, residues 884 to 992; F32, 114-amino-acid region, residues 870 to 983) it was calculated to be 53.0%.

The F80 and OK204 strains had a mixed-type CagA (Fig. 4). Insertion of a 13-amino-acid (residues 823 to 835 of 26695) region, which exists forward of the CagA repeat domain, is characteristic of the Western type. Both strains had an insert region, but they also had the ESS (F80 was classified into the A-B'-D-D type, and OK204 was classified into the A-B"-D type). OK204 had the complete ESS, while in the F80 ESS, the first half of the region was the same as the WSS sequence, and the following EPIYA-D region was the ESS sequence. So, F80 CagA may be closer to the Western type than to OK204 CagA.

Diversity of *cagE*. The *cagE* gene was well conserved among strains. In all isolates except for the F36 strain, sequencing of *cagE* genes revealed a 2,949-bp ORF encoding proteins of 983 amino acids, which was the same as the previously published strains 26695, J99, and NCTC11637. The F36 strain possessed

TABLE 4. Genetic properties of 33 isolates of *H. pylori*

Strain	Diagnosis ^a	<i>vacA</i>			<i>cagA</i>			<i>cagE</i>	
		Genotype	Geographic type	Size of predicted protein (bp)	EPIYA type ^b	Geographic type	Size of predicted protein (bp)	Geographic type	Size of predicted protein (bp)
F79	GU	s1a/m1a	Western	1,290	AB'CC	Western	1,216	East Asian	983
F37	AG	s1a/m1a	Western	1,295	A'BD	East Asian	1,172	East Asian	983
F16	AG	s1a/m1b	East Asian	1,291	ABD	East Asian	1,172	East Asian	983
OK109	AG	s1a/m1b	East Asian	1,289	ABD	East Asian	1,178	East Asian	983
OK118	AG	s1a/m1b	East Asian	1,291	ABD	East Asian	1,176	East Asian	983
OK158	DU	s1a/m1b	East Asian	1,296	ABD	East Asian	1,173	East Asian	983
OK159	AG	s1a/m1b	East Asian	1,296	AB'D	East Asian	1,172	East Asian	983
OK194	AG	s1a/m1b	East Asian	1,296	ABD	East Asian	1,177	East Asian	983
OK107	AG	s1a/m2	Western	1,303	AB'CCC	Western	1,251	Western	983
OK139	AG	s1a/m2	Western	1,303	AB'C	Western	1,183	Western	983
OK144	DU	s1a/m2	Western	1,303	AB'C	Western	1,183	Western	983
OK160	AG	s1a/m2	Western	1,303	AB'''C	Western	1,183	Western	983
OK185	DU	s1a/m2	Western	1,303	AB'C	Western	1,183	Western	983
OK204	DU	s1a/m2	Western	1,303	AB'D ^c	Western	1,188	Western	983
OK210	DU	s1a/m2	Western	1,323	AB'C	Western	1,188	East Asian	983
F80	DU	s1b/m1a	Western	1,295	AB'DD ^c	Western	1,222	Western	983
OK111	AG	s1b/m1b	East Asian	1,295	AB'CC	Western	1,214	Western	983
F13	AG	s1c/m1b	East Asian	1,296	ABD	East Asian	1,175	East Asian	983
F15	AG	s1c/m1b	East Asian	1,291	ABD	East Asian	1,172	East Asian	983
F17	AG	s1c/m1b	East Asian	1,297	AABD	East Asian	1,275	East Asian	983
F18	AG	s1c/m1b	East Asian	1,291	ABABD	East Asian	1,239	East Asian	983
F26	AG	s1c/m1b	East Asian	1,296	ABAB'D	East Asian	1,269	East Asian	983
F32	GCA	s1c/m1b	East Asian	1,291	ABD	East Asian	1,173	East Asian	983
F36	AG	s1c/m1b	East Asian	1,294	ABD	East Asian	1,170	East Asian	981
F55	AG	s1c/m1b	East Asian	1,291	AB'B'D	East Asian	1,232	East Asian	983
F92	AG	s1c/m1b	East Asian	1,291	ABABD	East Asian	1,270	East Asian	983
OK101	AG	s1c/m1b	East Asian	1,296	ABD	East Asian	1,172	East Asian	983
OK129	AG	s1c/m2	Western	1,323	ABD	East Asian	1,172	East Asian	983
OK130	DU	s1c/m2	Western	1,301	AB'C	Western	1,183	Western	983
OK180	DU	s1c/m2	Western	1,303	AB'C	Western	1,183	Western	983
OK181	AG	s1c/m2	Western	1,303	AB'	Western	1,148	Western	983
OK187	DU	s1c/m2	Western	1,303	AB'C	Western	1,189	East Asian	983
OK155	AG	s2/m2	Western	1,327	AB'C	Western	1,184	East Asian	983
26695	G	s1a/m1a	Western	1,290	AB'C	Western	1,186	Western	983
J99	DU	s1b/m1a	Western	1,288	B'C	Western	1,167	Western	983
NCTC11638	G	s1a/m1a	Western	1,296	AB	Western	1,147	Western	981
NCTC11637	G	s1a/m1a	Western	1,290	ABCCC	Western	1,247	Western	983

^a AG, atrophic gastritis; G, gastritis; GU, gastric ulcer; DU, duodenal ulcer; GCA, gastric cancer.^b B', EPIYT; B'', ESIYA; B''', ESIYT; A', NPIYA.^c Mixed-type CagA repeat domain.

a 2,943-bp ORF, which encoded 981 amino acids because of a deletion of the first 2 amino acids and was the same as the previously published NCTC11638. The homology between each isolate in amino acids was very high, at over 98% (Table 4).

Phylogenetic analysis of the *VacA*, *CagA*, and *CagE* sequences. The phylogenetic tree of the *VacA*, *CagA*, and *CagE* proteins demonstrated a genetic relationship among 33 clinical isolates and 4 previously published strains from the West: 26695, J99, NCTC11638, and NCTC11637 (Fig. 5). All three proteins were divided into two major groups, a Western group

and an East Asian group (Table 4). The distribution of the isolates was almost the same among each protein, with several exceptions.

For *CagA*, OK181, which was the A-B' type, was included in a Western group. F80 and OK204 were also contained a Western group, although they possessed ESS (Fig. 4). Since their *VacA* and *CagE* also contained a Western group, both strains may have originated from a Western area rather than an East Asian area. In the other strains, the EPIYA type accorded with the geographic type based on the phylogenetic tree.

TABLE 5. Distribution of the *vacA* allele types

Allele type	No. of strains (%) with:											Total
	s1a/m1a	s1a/m1b	s1a/m2	s1b/m1a	s1b/m1b	s1b/m2	s1c/m1a	s1c/m1b	s1c/m2	s2/m2	UD ^a	
Fukui	2 (1.8%)	9 (7.8%)	0 (0%)	1 (0.9%)	0 (0%)	0 (0%)	0 (0%)	94 (81.7%)	0 (0%)	0 (0%)	9 (7.8%)	115
Okinawa	0 (0%)	5 (4.8%)	9 (8.6%)	0 (0%)	1 (0.9%)	0 (0%)	0 (0%)	75 (71.4%)	6 (5.7%)	5 (4.8%)	4 (3.8%)	105
Total	2 (0.9%)	14 (6.3%)	9 (4.1%)	1 (0.5%)	1 (0.5%)	0 (0%)	0 (0%)	169 (76.8%)	6 (2.7%)	5 (2.3%)	13 (5.9%)	220

^a UD, undetectable.

a

				s1a-F	ss3-F
F37	s1a	1	ATGGAAATACAACAAACACACCGCAAATCAATCGCCCTCTGGTTTCTCTCGCTTTAGTAGGAGCATTGGTCAGCATCAC		

F80	s1b	1	ATGGAAATACAACAAACACACCGCAAATGAATCGCCCTTTAGTTTCTCTCGTTTACGAGGAGCGTTGATTAGTGCCAT		

F32	s1c	1	ATGGAAATACAACAAACACACCGCAAATCAATCGCCCTCTGGTTTCTCTCGCTTTAGTGGGGCTATTGGTCAGCATCAC		

OK155	s2	1	ATGGAATTACAACAAACACACCGCAAATCAATCGCCCTATTATCTCCCTCGCTTTAGTGGGGCATTAAATTGGCACCGA		
				s1c-F	
			A-----CGCAACAAAGTCATGCC--GCC	102	
			* * * * *		
			A-----CGCAAGAGAGTCATGCC--GCC	102	
			* * * * *		
			A-----CCACAAAAAGTCATGCC--GCC	102	
			* * * * *		
			ACTAGGGGCTAATACGCCAAATGATCCCATACAGCGAGAGTCGCGCC	129	
				ss2-F	

b

				VA3-F
F37	m1a	1912	TCAACCCAGAAAAACCCCTTGGGCACATCAAACTTATGTTTAATAATCTAACCCCTGGGTCAAATGCGGTTCATGGACTA	
			* * * * *	
F32	m1b	1897	GGACCTCAAGGAAGCCCTTGGGGCACGGCAAAGCTTATGTTTAATAATCTAACCCCTAGGCCCAATGCAGTCATGGATTA	
			*** * * * * *	
OK155	m2	1984	GGAGCCCAGGAAACATTGCCCGCAAACAGGGCTTATGTTTAACAACCTAACCCCTAAATAGCAACGCAAGCATGGATTA	
				VA4-F
			TAGTCAATTTTCAAATTTAACCATCCAAGGGGATTTTATCAACAATCAAGGCACATCAACTATCTGGTCCGAGGCGGGA	

			TAGTCAATTTTCAAATGTAACCATTCAGGGGAATTTTATCAACAATCAAGGCACATCAATTTACTTGGTTCGAGGTGGGA	
			* * * * *	
			TGGTAAGGATTGGATTTAACCATTCAAAGGATTTACCAATAATCAAGGCGTGATGAATCTTTTGTCCAAGATGGGC	
				VA3-R
			AAGTGGCAACCTTAAATGTAGGCAATGCAGCAGCTATGATGTTTAATAATGATATAGACAGCGCGACCGGATTTTACAAA	
			* * * * *	
			ACATAGAAACCTTAAATGTAGGCAATGCAGCAGTCATGAGTTTAAATAACGATATAGATAGCGGACTGGATTTTACAAA	

			GTGTAGCGCCCTTAAATATAGGCAATACAGCAACCATGAAGTTTAATAATGTAGTGATAGCGGACCGGGTTTACCAA	
				VA4-R
			CCGCTCATCAAGATTAACAGCGCTCAAGATCTCATTAAAAATACAGAGCATGTTTATTGAAAGCGAAAATCATTGGTTA	

			CCGCTTATCAAAATTAACAGCGCTCAAGATCTCATTAAAAATAAAGAACATGTCTTATTGAAAGCGAAAATCATCGGTTA	
			*** * * * *	
			CCGATCATTAGATCAATAACGCTCAAAATCTCACTAAAAACAAAGAACATGTTTATTGTAAGCGCAAAACATTGATTA	
				VA3-R
			T-----GGTAATGTTTCTACAGGTACTAATGGC	2259
			* * * * *	
			T-----GAAAATGCTTCTTTAGGCACTAACAGC	2244
			* * * * *	
			TAATTTAGTAGGAGTGCAAGGCGCTAGTTATGACAAT	2340
				VA4-R

FIG. 1. Nucleotide sequence alignment of *vacA* genes. (a) *vacA* s-region sequences showing differences among s1a, s1b, s1c, and s2 alleles. (b) *vacA* m-region sequences showing differences among m1a, m1b, and m2 alleles. Numbers at the heads and ends of sequences represent the positions of nucleotides in each strain. Stars denote identity. A hyphen indicates the absence of a nucleotide.

TABLE 6. Analysis of divergence in the *vacA* signal sequence and m-region

Strain (type)	% Nucleotide sequence identity ^a with:			
	F37 (s1a)	F80 (s1b)	F32 (s1c)	OK155 (s2)
F37 (m1a)	—	86.3	93.1	62.8
F80 (m1a)	92.5	—	80.4	62.8
F32 (m1b)	84.8	89.4	—	64.3
OK155 (m2)	70.5	72.3	69.2	—

^a Values above and to the right of the dashes are for the signal sequence region, and those below and to the left of the dashes are for the m-region.

In *VacA*, the phylogenetic analysis indicated a clear separation between the m1 and m2 sequences. All m2 strains are classified into the Western cluster. Furthermore, in the m1 cluster, m1a and m1b clusters were clearly distinguished. All m1a and m1b strains are classified into the Western and East Asian clusters, respectively. Moreover, in the m1a and m1b clusters, s1a or s1c and s1b clusters were divided. Subtypes of s1a and s1c were not separated. F37 and OK129 were included in a Western group, although both of their CagA types were East Asian. This resulted from their subtypes in the m region: F37 was m1a and OK129 was m2. On the other hand, OK111 (*vacA* genotype, s1b/m1b) was included in an East Asian group, although the CagA type was the Western type, resulting from the subtypes of m1b.

For CagE, F79, OK155, OK187, and OK210 were included in the East Asian group although the CagA type was the Western type. In the other strains, except the above seven strains (F37, F79, OK111, OK129, OK155, OK187, and OK210) (78.8%, 26/33), the geographic types corresponded among the three proteins.

Interestingly, the prevalence of infection with the Western cluster strain in both CagA and *VacA* was significantly higher

TABLE 7. Analysis of *vacA* divergence

Strain	% Nucleotide and amino acid sequence identity ^a with:			
	F37 (s1a/m1a)	F80 (s1b/m1a)	F32 (s1c/m1b)	OK155 (s2/m2)
F37	—	93.7	90.6	78.8
F80	94.4	—	89.9	77.1
F32	91.6	91.4	—	78.7
OK155	84.1	83.0	84.4	—

^a Values above and to the right of the dashes are for amino acid sequences, and those below and to the left of the dashes are for nucleotide sequences.

in patients with peptic ulcer (90.0%, 9/10) than in patients with chronic gastritis (22.7%, 5/22) ($\chi^2 = 12.64$, $P = 0.00057$).

DISCUSSION

It is known that in East Asian countries, such as Japan and Korea, the incidence of gastric carcinoma is among the highest in the world. There are also indications of significant geographic differences among *H. pylori* strains (13, 15). *H. pylori* strains have been classified into type I and type II, depending on the presence of *VacA* and CagA (36), and type I strains are thought to be more virulent than type II strains. It has been confirmed that almost all Japanese isolates are type I strains (20). It is also thought that there are correlations between the *vacA* and *cagA* allelic geographic variations, and both genes are classified as virulence markers (18, 35). In the present study, we examined the molecular genetic relationships among *vacA*, *cagA*, and *cagE* in Okinawa and Fukui, Japan. We selected 33 *H. pylori* isolates according to the genotypes of *vacA* in most Okinawa isolates, since there are various *vacA* genotypes; although the s1c/m1b type is predominant in Okinawa (71.4%), almost all isolates in Fukui (81.7%) express the s1c/m1b type (40). Since Okinawa is an area which has had active interna-

60190 m1a	509	ID--TGNGGFNT--LDFSGVTNKVNINKLITASTNVAVKNFNINELIVKTNGSVGEYTHFSEIDIGSQSRINTVRLETGTRSIFSGGVKFKSGEKLVIDE
F32 m1b	512	ID--TGNGGFNT--LDFSGVTNKVNINKLITASTNVAIKNFNINELLVKTNGISVGEYTNFSEIDIGNQSRINTVRLETGTRSIFSGGVKFKGGEKLVIDE
OK210 m2	533	IDATKSDNGLNTSLDFSGVTDKVNINKLITTSATNVNINKNFIDIKELVVTTTRVQSFQGYTIFGENIGDKSRIGVVSLOTGYSPAYSGGVTFKGGKKLVIDE
Tx30a m2	522	IDATKSDNGLNTSLDFSGVTDKVNINKLITTAATNVNINKNFIDIKELVVTTTRVQSFQGYTIFGENIGDKSRIGVVSLOTGYSPAYSGGVTFKGGKKLVIDE
FYYPWNYFDARNIKNVEITRKFFASSTPENPFGTSKLMFNNTLGGQNAVMDYSQFSNLTIQGDFINNQGTINYLVRGGKVATLVNGNAAMMFNNNDIDSA		
*** ***** * *		
FYYPWNYFDARNIKNVEITNKLAFGPGQSPWGTAKLMFNNTLGGQNAVMDYSQFSNLTIQGDFINNQGTINYLVRGGNIETLVNGNAAMMFNNNDIDSA		
* ***** * *		
LYHAPWNYFDARNVTDVEINKILFGAPGYIAGTKGLMFNNLTLSNASMDYKDLDLTIQGHFTNNQGTMLNFVQDGRVATLVNAGHQASMIFFNNVVDSDA		

LYHAPWNYFDARNVTDVEINKILFGAPGNIAGTKGLMFNNLTLSNASMDYKDLDLTIQGHFTNNQGTMLNFVQDGRVATLVNAGHQASMIFFNNVVDST		
TGfYKPLIKINSAQDLIKNTEHVLLKAKIIGY--GNVS--TGtNGISNVNLE		752
***** * *		
TGfYKPLIKINSAQDLIKNTEHVLLKAKIIGY--ENAS--LGtNSISNANLI		755
***** * *		
TGfYKPLIKINNAQNLTKNKEHVLVKARNIDYNLVGVQGASYNISASNTNLQ		785

TGfYKPLIKINNAQNLTKNKEHVLVKARNIDYNLVGVQGASYNISASNTNLQ		774

FIG. 2. Amino acid sequence alignment of the *VacA* m-region showing differences among the m1a, m1b, and m2 subtypes. Numbers at the heads and ends of sequences represent the positions of the amino acid residues of each strain. Stars denote identity. A hyphen indicates the absence of an amino acid residue.

FIG. 3. Amino acid sequence alignment of the EPIYA regions (CagA repeat domains) for Western (strain 26695) and East Asian (strain F32) CagA proteins. The Western CagA-specific sequence (WSS) and the East Asian CagA-specific sequence (ESS) are boxed. Numbers at the heads of sequences represent the positions of amino acid residues of both strains. Stars denote identity. A hyphen indicates the absence of an amino acid residue.

tional communication with the West historically and had a relatively large American presence in the last half-century, different types of *H. pylori* may be present. On the other hand, the low genetic diversity in *vacA* genes of *H. pylori* isolates in Fukui may be due to the typical Japanese traits. Japan is a country consisting of a relatively homogeneous population and is geographically isolated. Therefore, the opportunity for the transfer of DNA between strains of different genotypes may be

FIG. 4. Amino acid sequence alignment of the CagA repeat domain for strains F80 and OK204, which possess a mixed-type CagA. Numbers at the heads and ends of sequences represent the positions of the amino acid residues of each strain. Stars denote identity. A hyphen indicates the absence of an amino acid residue.

FIG. 5. Phylogenetic tree of the complete amino acid sequences of VacA (a), CagA (b), and CagE (c). Scale bars represent calculated distances. All three proteins were divided into two major groups, a Western group and an East Asian group. §, ¶, and † indicate isolates from duodenal ulcer, gastric ulcer, and gastric cancer, respectively. Other strains were isolated from gastritis.

lower than that in Western countries. Van Doorn et al. also reported that the subtype s1c was prevalent in East Asia but appeared to be very rare in other parts of the world (35).

The phylogenetic analysis of VacA demonstrated that all m2 strains are classified into the Western group. Ji et al. also investigated the phylogenetic analysis of a set of eight Chinese and six Western *vacA* genes and reported that while the m-region of m1 alleles has an evolutionary topology similar to the conserved regions of the genes, the m-region of m2 alleles has an evolutionary history independent of the rest of the gene. They suggested that the m2 m-region spread after the separation of Chinese and Western strains (22). In addition, s1a/m1a, s1b/m1a, and s1b/m1b strains, which appear to be very rare in Japan, are also involved in the Western group. In the present study, we calculated the homology in the s-region and m-region among s1a, s1b, s1c, and s2 or m1a, m1b, and m2 subtypes. In the s-region, there was more than 80% homology among s1 subtypes, and the correlation of s1a and s1c subtypes was higher than that of s1b, whereas there was only about 60% homology between the s1 and s2 subtypes. In the m-region, we had similar results; the correlation of m1a and m1b subtypes was higher, but that of the m1 and m2 subtypes was lower. Atherton et al. also compared the VacA sequences between strain 60190 (m1) and strain Tx30a (m2) and found that there was only 59.0% amino acid identity in the middle of VacA. The differences between s1 and s2 and m1 and m2 may be linked to strong and weak toxicities (5). Whereas different rates of evo-

lution in different regions of the same gene might be expected, dramatically different patterns of evolution between different regions of the same set of genes are best explained by recombination. Hence, it is likely that the mosaicism in the *vacA* gene is maintained in the population by genetic recombination, and since there is a single copy of the *vacA* gene in *H. pylori*, this implies horizontal transfer of DNA (22).

It has been considered that *cagA*-positive *H. pylori* strains are associated with increased risk for the development of atrophic gastritis and gastric carcinoma (11, 18, 26), and large sequence differences distinguish the *cagA* gene fragments from East Asian and Western strains (1, 37). Recently, it was demonstrated by using in vitro transfection experiments of a human gastric cancer cell line (AGS) that the East Asian-specific sequence of CagA protein confers stronger SHP-2 binding and transforming activities than does the Western-specific sequence (17), meaning that the potential of East Asian CagA to disturb host cell functions as a virulence factor may be higher than that of Western CagA. In clinical cases, we have reported that the CagA-SHP-2 complex is found in in vivo human gastric mucosa (38), and the grades of inflammation, activity of gastritis, and atrophy are significantly higher in patients with gastritis infected with the East Asian CagA-positive *H. pylori* strains than in patients with gastritis infected with the *cagA*-negative or Western CagA-positive strains (6). In this study, we observed two major subtypes, Western and East Asian CagA, according to the EPIYA region sequence constituting the

SHP-2 binding site as previously reported (6, 7, 40). The EPIYA type corresponds with the cluster of the phylogenetic tree of CagA in most strains. We also examined the homology in the CagA repeat domain between Western and East Asian strains, and there is less homology (only 53%) when comparing the total sequence, indicating that large differences in the sequences may affect the strength or weakness of the toxicity. Furthermore, there were two strains possessing a mixed-type CagA in the study. Although it is possible that interstrain gene transfer and recombination occurred in the strains, they may be derived from a Western area, because both VacA and CagE were included in the Western type (as described below).

cagE, which is a homolog of the *ptlC* and *virB4* genes of *Bordetella pertussis* and *Agrobacterium tumefaciens*, respectively (12), is located in upper proximity to *cagA* in the *cagPAI*, and the product is a component of the type IV secretion system (32). It has been reported that CagE is involved in the induction of interleukin-8 expression in gastric epithelial cells (32). In this study, we showed that the *cagE* gene is well conserved among strains. Although we could not find the Western or East Asian group-specific profile in the full-length amino acid sequences, the protein is clearly divided into two groups in the phylogenetic tree.

The phylogenetic analysis of the three proteins demonstrates the genetic relationship. The distributions of strains exhibited almost the same patterns among the three proteins, indicating that there is a genetic linkage between *vacA* and the *cagPAI*, although there is substantial distance between *vacA* loci and *cag* genes on the bacterial genome (3, 31). Thus, the selection of a *vacA/cagPAI* genotype may have a functional basis, perhaps in counterbalancing proinflammatory and anti-inflammatory characteristics of strains to facilitate long-term equilibrium in the human gastric mucosa. Interestingly, peptic ulcer strains were associated with the Western cluster in both CagA and VacA in this study. The strain diversity observed in Okinawa may affect the difference in the prevalence of diseases related to *H. pylori* infection between Fukui and Okinawa. Further analysis of strain diversity and clinical relevance is needed to clarify the pathogenicity of *H. pylori*.

ACKNOWLEDGMENTS

This work was supported by a grant-in-aid for Scientific Research (B) from the Japan Society for the Promotion of Science.

We thank Yusuke Kashiwazaki for technical assistance.

REFERENCES

- Achtman, M., T. Azuma, D. E. Berg, Y. Ito, G. Morelli, Z. Pan, S. Suerbaum, S. A. Thomson, A. van der Endo, and L. J. van Doorn. 1999. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol. Microbiol.* 32:459–470.
- Akopyants, N. S., S. W. Clifton, D. Kersulyte, J. E. Crabtree, B. E. Youree, C. A. Reece, N. O. Bukanov, E. S. Drazek, B. A. Roe, and D. E. Berg. 1998. Analyses of the *cag* pathogenicity island of *Helicobacter pylori*. *Mol. Microbiol.* 28:37–53.
- Alm, R. A., L. S. Ling, D. T. Moir, B. L. King, E. D. Brown, P. C. Doig, D. R. Smith, B. Noonan, B. C. Guild, B. L. deJonge, G. Carmel, P. J. Tummino, A. Caruso, M. Uria-Nickelsen, D. M. Mills, C. Ives, R. Gibson, D. Merberg, S. D. Mills, Q. Jiang, D. E. Taylor, G. F. Vovis, and T. J. Trust. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 397:176–180.
- Asahi, M., T. Azuma, S. Ito, Y. Ito, H. Suto, Y. Nagai, M. Tsubokawa, Y. Tohyama, S. Maeda, M. Omata, T. Suzuki, and C. Sasakawa. 2000. *Helicobacter pylori* CagA protein can be tyrosine phosphorylated in gastric epithelial cells. *J. Exp. Med.* 191:593–602.
- Atherton, J. C., P. Cao, R. M. Peek, M. K. Tummuru, M. J. Blaser, and T. L. Cover. 1995. Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*: association of specific *vacA* types with cytotoxin production and peptic ulceration. *J. Biol. Chem.* 270:17771–17777.
- Azuma, T., S. Yamazaki, A. Yamakawa, M. Ohtani, A. Muramatsu, H. Suto, Y. Ito, M. Dojo, Y. Yamazaki, M. Kuriyama, Y. Keida, H. Higashi, and M. Hatakeyama. 2004. Association between diversity in the Src homology 2 domain-containing tyrosine phosphatase binding site of *Helicobacter pylori* CagA protein and gastric atrophy and cancer. *J. Infect. Dis.* 189:820–827.
- Azuma, T. 2004. *Helicobacter pylori* CagA protein variation associated with gastric cancer in Asia. *J. Gastroenterol.* 39:97–103.
- Azuma, T., A. Yamakawa, S. Yamazaki, M. Ohtani, Y. Ito, A. Muramatsu, H. Suto, Y. Yamazaki, Y. Keida, H. Higashi, and M. Hatakeyama. 2004. Distinct diversity of the *cag* pathogenicity island among *Helicobacter pylori* strains in Japan. *J. Clin. Microbiol.* 42:2508–2517.
- Backert, S., E. Ziska, V. Brinkmann, U. Zimny-Arndt, A. Fauconier, P. R. Jungblut, M. Naumann, and T. F. Meyer. 2000. Translocation of the *Helicobacter pylori* CagA protein in gastric epithelial cells by a type IV secretion apparatus. *Cell. Microbiol.* 2:155–164.
- Backert, S., S. Moese, M. Selbach, V. Brinkmann, and T. F. Meyer. 2001. Phosphorylation of tyrosine 972 of the *Helicobacter pylori* CagA protein is essential for induction of a scattering phenotype in gastric epithelial cells. *Mol. Microbiol.* 42:631–644.
- Blaser, M. J., G. I. Perez-Perez, H. Kleanthous, T. L. Cover, R. M. Peek, P. H. Chyou, G. N. Stemmermann, and A. Nomura. 1995. Infection with *Helicobacter pylori* strains possessing *cagA* is associated with an increased risk of developing adenocarcinoma of the stomach. *Cancer Res.* 55:2111–2115.
- Censini, S., C. Lange, Z. Xiang, J. E. Crabtree, P. Ghiara, M. Borodovsky, R. Rappuori, and A. Covacci. 1996. *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc. Natl. Acad. Sci. USA* 93:14648–14653.
- Covacci, A., J. L. Telford, G. DelGiudice, J. Parsonnet, and R. Rappuori. 1999. *Helicobacter pylori* virulence and genetic geography. *Science* 284:1328–1333.
- Cover, T. L., M. K. Tummuru, P. Cao, S. A. Thompson, and M. J. Blaser. 1994. Divergence of genetic sequences for the vacuolating cytotoxin among *Helicobacter pylori* strains. *J. Biol. Chem.* 269:10566–10573.
- Falush, D., T. Wirth, B. Linz, J. K. Pritchard, M. Stephens, M. Kidd, M. J. Blaser, D. Y. Graham, S. Vacher, G. I. Perez-Perez, Y. Yamaoka, F. Megraud, K. Otto, U. Reichard, E. Katzwitsch, X. Wang, M. Achtman, and S. Suerbaum. 2003. Traces of human migrations in *Helicobacter pylori* populations. *Science* 299:1582–1585.
- Higashi, H., R. Tsutsumi, S. Muto, T. Sugiyama, T. Azuma, M. Asaka, and M. Hatakeyama. 2002. SHP-2 tyrosine phosphatase as an intracellular target of *Helicobacter pylori* CagA protein. *Science* 295:683–686.
- Higashi, H., R. Tsutsumi, A. Fujita, S. Yamazaki, M. Asaka, T. Azuma, and M. Hatakeyama. 2002. Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. *Proc. Natl. Acad. Sci. USA* 99:14428–14433.
- Huang, J. Q., G. F. Zheng, K. Sumanac, E. J. Irvine, and R. H. Hunt. 2003. Meta-analysis of relationship between *cagA* seropositivity and gastric cancer. *Gastroenterology* 125:1636–1644.
- International Agency for Research on Cancer Working Group. 1994. Schistosomes, liver flukes and *Helicobacter pylori*. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. IARC Monogr. Eval. Carcinog. Risks Hum. 61:1–124.
- Ito, Y., T. Azuma, S. Ito, H. Miyaji, M. Hirai, Y. Yamazaki, F. Sato, T. Kato, Y. Kohli, and M. Kuriyama. 1997. Analysis and typing of the *vacA* gene from *cagA*-positive strains of *Helicobacter pylori* isolated in Japan. *J. Clin. Microbiol.* 35:1710–1714.
- Ito, Y., T. Azuma, S. Ito, H. Suto, H. Miyaji, Y. Yamazaki, Y. Kohli, and M. Kuriyama. 1998. Full-length sequence analysis of the *vacA* gene from cytotoxic and noncytotoxic *Helicobacter pylori*. *J. Infect. Dis.* 178:1391–1398.
- Ji, X., F. Frati, S. Barone, C. Pagliaccia, D. Burrone, G. Xu, R. Rappuoli, J. M. Reyat, and J. L. Telford. 2002. Evolution of functional polymorphism in the gene coding for the *Helicobacter pylori* cytotoxin. *FEMS Microbiol. Lett.* 206:253–258.
- Odenbreit, S., J. Puls, B. Sedlmaier, E. Gerland, W. Fischer, and R. Haas. 2000. Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. *Science* 287:1497–1500.
- Parsonnet, J., G. P. Friedman, D. P. Vandersteen, Y. Chang, J. H. Vogelstein, N. Orentreich, and R. K. Sibley. 1991. *Helicobacter pylori* infection and the risk of gastric carcinoma. *N. Engl. J. Med.* 325:1127–1131.
- Phadnis, S. H., D. Ilver, L. Janzon, S. Normark, and T. U. Westblom. 1994. Pathological significance and molecular characterization of the vacuolating toxin gene of *Helicobacter pylori*. *Infect. Immun.* 62:1557–1565.
- Rokkas, T., S. Ladas, C. Liatsos, E. Petridou, G. Papatheodorou, S. Theodoris, A. Karameris, and S. Raptis. 1999. Relationship of *Helicobacter pylori* CagA status to gastric cell proliferation and apoptosis. *Dig. Dis. Sci.* 44:487–493.
- Segal, E. D., J. Cha, J. Lo, S. Falkow, and L. S. Tompkins. 1999. Altered states: involvement of phosphorylated CagA in the induction of host cellular

- growth changes by *Helicobacter pylori*. Proc. Natl. Acad. Sci. USA **96**:14559–14564.
28. Stein, M., R. Rappuoli, and A. Covacci. 2000. Tyrosine phosphorylation of the *Helicobacter pylori* CagA antigen after *cag*-driven host cell translocation. Proc. Natl. Acad. Sci. USA **97**:1263–1268.
29. Stein, M., F. Bagnoli, R. Halenbeck, R. Rappuoli, W. J. Fantl, and A. Covacci. 2002. c-Src/Lyn kinases active *Helicobacter pylori* CagA through tyrosine phosphorylation of the EPIYA motifs. Mol. Microbiol. **43**:971–980.
30. The EUROGAAT Study Group. 1993. An association between *Helicobacter pylori* infection and gastric cancer. Lancet **341**:1359–1362.
31. Tomb, J. F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzgerald, N. Lee, M. D. Adams, E. K. Hickey, D. E. Berg, J. D. Gocayne, T. R. Utterback, J. D. Peterson, J. M. Kelly, M. D. Cotton, J. M. Weidman, C. Fujii, C. Bowman, L. Watthey, E. Wallin, W. S. Hayes, M. Borodovsky, P. D. Karp, H. O. Smith, C. M. Fraser, and J. C. Venter. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature **388**:539–547.
32. Tummuru, M. K., S. A. Sharma, and M. J. Blaser. 1995. *Helicobacter pylori* *picB*, a homologue of the *Bordetella pertussis* toxin secretion protein, is required for induction of IL-8 in gastric epithelial cells. Mol. Microbiol. **18**:867–876.
33. Uemura, N., S. Okamoto, S. Yamamoto, N. Matsumura, S. Yamaguchi, M. Yamakido, K. Taniyama, N. Sasaki, and R. J. Schlemper. 2001. *Helicobacter pylori* infection and the development of gastric cancer. N. Engl. J. Med. **345**:784–789.
34. van Doorn, L. J., C. Figueiredo, R. Sanna, S. Pena, P. Midolo, E. K. Ng, J. C. Atherton, M. J. Blaser, and W. G. Quint. 1998. Expanding allelic diversity of *Helicobacter pylori* *vacA*. J. Clin. Microbiol. **36**:2597–2603.
35. van Doorn, L. J., C. Figueiredo, F. Megraud, S. Pena, P. Midolo, D. M. Queiroz, F. Carneiro, B. Vanderborght, M. D. Pegado, R. Sanna, W. De Boer, P. M. Schneeberger, P. Correa, E. K. Ng, J. Atherton, M. J. Blaser, and W. G. Quint. 1999. Geographic distribution of *vacA* allelic types of *Helicobacter pylori*. Gastroenterology **116**:823–830.
36. Xiang, Z., S. Censini, P. F. Bayeli, J. L. Telford, N. Figura, R. Rappuoli, and A. Covacci. 1995. Analysis of expression of *cagA* and *vacA* virulence factors in 43 strains of *Helicobacter pylori* reveals that clinical isolates can be divided into two major types and that *cagA* is not necessary for expression of the vacuolating cytotoxin. Infect. Immun. **63**:94–98.
37. Yamaoka, Y., E. Orito, M. Mizokami, O. Gutierrez, N. Saitou, T. Kodama, M. S. Osato, J. G. Kim, F. C. Ramirez, V. Mahachai, and D. Y. Graham. 2002. *Helicobacter pylori* in North and South America before Columbus. FEBS Lett. **517**:180–184.
38. Yamazaki, S., A. Yamakawa, Y. Ito, M. Ohtani, H. Higashi, M. Hatakeyama, and T. Azuma. 2003. The CagA protein of *Helicobacter pylori* is translocated into epithelial cells and binds to SHP-2 in human gastric mucosa. J. Infect. Dis. **187**:334–337.
39. Yu, D., C. Qu, O. Henegariu, X. Lu, and G. Feng. 1998. Protein-tyrosine phosphatase Shp-2 regulates cell spreading, migration, and focal adhesion. J. Biol. Chem. **273**:21125–21131.
40. Zhou, W., S. Yamazaki, A. Yamakawa, M. Ohtani, Y. Ito, Y. Keida, H. Higashi, M. Hatakeyama, J. Si, and T. Azuma. 2004. The diversity of *vacA* and *cagA* genes of *Helicobacter pylori* in East Asia. FEMS Immunol. Med. Microbiol. **40**:81–87.