JOURNAL OF CLINICAL MICROBIOLOGY, June 2004, p. 2508–2517 Vol. 42, No. 6
0095-1137/04/$08.00+0 DOI: 10.1128/JCM.42.6.2508–2517.2004

Distinct Diversity of the *cag* Pathogenicity Island among *Helicobacter pylori* Strains in Japan

Takeshi Azuma,[1]* Akiyo Yamakawa,[1] Shiho Yamazaki,[1] Masahiro Ohtani,[1] Yoshiyuki Ito,[1] Atsushi Muramatsu,[1] Hiroyuki Suto,[1] Yukinao Yamazaki,[2] Yoshihide Keida,[3] Hideaki Higashi,[4] and Masanori Hatakeyama[4]

Second Department of Internal Medicine[1] and Department of Endoscopic Medicine,[2] Faculty of Medical Science, University of Fukui, Fukui, Division of Internal Medicine, Okinawa Chubu Hospital, Okinawa,[3] and Division of Molecular Oncology, Institute for Genetic Medicine and Graduate School of Science, Hokkaido University, Sapporo,[4] Japan

Received 5 February 2004/Returned for modification 29 February 2004/Accepted 2 March 2004

The severity of *Helicobacter pylori*-related disease is correlated with the presence of a *cag* pathogenicity island (PAI). Genetic diversity within the *cag* PAI may have a modifying effect on the pathogenic potential of the infecting strain. We analyzed the complete *cag* PAI sequences of 11 representative Japanese strains according to their *vacA* genotypes and clinical effects and examined the relationship between the diversity of the *cag* PAI and clinical features. The *cag* PAI genes were divided into two major groups, a Western and a Japanese group, by phylogenetic analysis based on the entire *cag* PAI sequences. The predominant Japanese strains formed a Japanese cluster which was different from the cluster formed by Western strains. The diversity of the *cag* PAI was associated with the *vacA* and *cagA* genotypes. All strains with the s1c *vacA* genotype were in the Japanese cluster. In addition, all strains with the East Asian-type *cagA* genotype were also in the Japanese cluster. Patients infected with the Japanese-cluster strain had high-grade gastric mucosal atrophy. These results suggest that a distinct diversity of the *cag* PAI of *H. pylori* is present among Japanese strains and that this diversity may be involved in the development of atrophic gastritis and may increase the risk for gastric cancer.

Helicobacter pylori is a gram-negative microaerophilic bacterium that chronically colonizes the gastric epithelium of more than half of all people worldwide. It is a human pathogen responsible for chronic active gastritis; and infection with this organism is an important risk factor for peptic ulcer, gastric cancer, and gastric mucosa-associated lymphoid tissue lymphoma (20, 22, 28, 29). The CagA protein, encoded by the *cagA* gene, is one of the most studied virulence factors of *H. pylori* and is highly immunogenic. The *cagA* gene is one of several genes in a pathogenicity island (PAI) known as the *cag* PAI. The presence of *cagA* is considered a marker of the presence of *cag* PAI (10). The *cag* PAI is a 40-kb locus in the chromosomal glutamate racemase gene. Its G+C content (35%) differs from the G+C content of the rest of the genome (39%), suggesting that it was acquired from another organism by horizontal transfer (8, 11, 25). At some point during evolution, IS*605*, a mobile sequence encoding two transposases, entered the *H. pylori* genome and in some strains interrupted, multiplied, or deleted parts of the PAI (8). The severity of *H. pylori*-related disease is correlated with the presence of the *cag* PAI. Infection with *cag* PAI-positive *H. pylori* is statistically associated with duodenal ulceration, gastric mucosal atrophy, and gastric cancer (7, 8, 10). The *cag* PAI contains 31 genes, and 6 of the *cag* genes are thought to encode a putative type IV secretion system that specializes in the transfer of a variety of multimolecular complexes across the bacterial membrane to the extracellular space or into other cells (11). Recent studies have indicated that CagA is delivered into epithelial cells by the *cag* type IV secretion system, where it is phosphorylated on tyrosine residues and connected to eukaryotic signal transduction pathways, and is likely to play a major role in *H. pylori*-host cell interactions and pathogenesis (3, 21, 23, 24). Moreover, it was recently discovered (17) that translocated CagA forms a physical complex with the SRC homology 2 domain-containing tyrosine phosphatase SHP-2, which is known to play an important positive role in mitogenic signal transduction, and stimulates phosphatase activity. On the basis of the sequence constituting the SHP-2 binding site, CagA proteins can be subclassified into East Asian and Western types. The East Asian-type CagA possesses stronger SHP-2 binding and transforming activities than the Western-type CagA (16).

H. pylori exhibits a large degree of genomic and allelic diversity. Strain-specific diversity has been proposed to be involved in the organism's ability to cause different diseases. There are also indications of significant geographical differences among strains. Only one-half to two-thirds of Western isolates carry the *cag* PAI. In contrast, nearly all East Asian strains carry the *cag* PAI (19, 26). It has also been reported that large sequence differences distinguish the *cagA* gene fragments from Asian strains and those from other strains. The lineage of *H. pylori* isolates infecting Asian subjects may be different from that of isolates in other parts of the world, or a specific strain may have accumulated in the Asian population (1). The variable genetic structure of the *cag* PAI may influence the clinical outcome of *H. pylori* infection.

Vacuolating cytotoxin (VacA) is another virulence factor. The *vacA* gene is present in all *H. pylori* strains and contains at least two variable parts. The s region (which encodes the signal peptide) coexists as s1 or s2 allelic types. Subtypes s1a, s1b, and

* Corresponding author. Mailing address: Second Department of Internal Medicine Faculty of Medical Science, University of Fukui Matsuoka-cho, Yoshida-gun, Fukui 910-1193, Japan. Phone: 81-776-61-8351. Fax: 81-776-61-8110. E-mail: azuma@fmsrsa.fukui-med.ac.jp.

s1c have been identified among type s1 strains. The m (middle) region occurs as the m1 or m2 allelic type (4, 26, 27, 30). Production of the 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 (4). Because most *vacA* s1 strains are *cagA* positive, these two markers are closely related. It has been reported that the *vacA* s1- and *cagA*-positive genotype is significantly associated with a more severe clinical outcome, such as gastric cancer (26).

The incidence of gastric cancer and the rate of death due to gastric cancer in Japan are high compared with those in other developed countries. However, large intracountry differences in the rates of death from gastric cancer have been reported (18). Fukui is a typical rural prefecture located on the central Japanese mainland (Honshu), while Okinawa consists of islands in the southwestern part of Japan and has a history and food culture different from those in other parts of Japan. The prevalence of atrophic gastritis, a precursor lesion of gastric cancer, is more frequent in Fukui, and the rate of death from gastric cancer is more than 2.4 times higher in Fukui (43.7 per 100,000 population in 1999) than in Okinawa (18.2 per 100,000 population in 1999). In this study, we selected 11 Japanese strains according to their *vacA* genotypes, clinical effects, and geographical locations and determined the complete *cag* PAI gene sequences in order to examine the association between the diversity of genes in the *cag* PAI and clinical outcomes.

MATERIALS AND METHODS

***H. pylori* strains.** *H. pylori* clinical isolates were obtained during upper gastroduodenal endoscopy at the Second Department of Internal Medicine, Faculty of Medical Science, University of Fukui, Fukui, Japan, and Okinawa Chubu Hospital, Okinawa, Japan, respectively. A total of four biopsy specimens were obtained from each patient: two from the greater curvature of the gastric antral mucosa and two from the greater curvature of the gastric fundic mucosa. One specimen each from the antral and fundic mucosae was fixed in 10% buffered formalin (pH 7.2) and subjected to histological analysis. The other antral and fundic mucosal specimens were subjected to culture for *H. pylori*. These studies were performed according to the principles of the Declaration of Helsinki, and consent was obtained from each individual after he or she was given a full description of the nature and protocol of the study.

Histological analysis. Biopsy specimens were embedded in paraffin and stained with hematoxylin-eosin. The specimens were examined without knowledge of the experimental results. The histological features of chronic gastritis in terms of lymphocyte infiltration (inflammation), neutrophil infiltration (activity of gastritis), and mucosal atrophy were graded from 0 to 3 according to the Updated Sydney system (12).

***H. pylori* culture.** Gastric biopsy specimens from each patient were inoculated onto a Trypticase soy agar II (TSA-II)–5% sheep blood plate and cultured under microaerobic conditions (O_2, 5%; CO_2, 15%; N_2, 80%) at 37°C for 5 days. A single colony was picked from each primary culture plate, inoculated onto a fresh TSA-II plate, and cultured under the conditions described above. A few colonies were picked from each plate and transferred into 20 ml of brucella broth liquid culture medium containing 10% fetal calf serum and cultured for 3 days under the same conditions described above. A part of the liquid culture sample was stored at −80°C in 0.01 M phosphate-buffered saline (PBS) 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 TE buffer (10 mM Tris HCl, 1 mM EDTA), and stored at 4°C until PCR analysis and nucleotide sequencing.

Complete *cag* PAI nucleotide sequences of nine representative Japanese strains. The complete *cag* PAIs of two Japanese strains, strains F32 and OK107, were sequenced previously (5). We selected nine additional Japanese strains according to their *vacA* genotypes, clinical effects, and geographical locations. The characteristics of the selected strains are listed in Table 1. The region comprising the entire *cag* PAI of each strain was amplified with *cag* PAI-spanning primer sets designed from the sequence of strain 26695, as reported previously (25) (Table 2). PCR conditions were as follows: heating at 94°C for 5 min, followed by 25 cycles consisting of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s. The tubes were held at 72°C for 7 min before storage at 4°C. The PCR products were then purified with Centricon-100 Concentrator columns (Amicon, Beverly, Mass.). DNA sequencing was performed by the dideoxynucleotide chain termination method with a BigDye Terminator Cycle Sequencing Ready Reaction Mix (Applied Biosystems, Tokyo, Japan) in an ABI PRISM 310 Genetic Analyzer (Applied Biosystems). According to the protocol of the manufacturer, reagent mixtures containing 5 μl of purified PCR product, 3.2 pmol of primer, 8 μl of Terminator Cycle Sequencing Ready Reaction Mix, and 5 μl of sterilized distilled water were prepared. The reaction tubes were placed in the thermal cycler, and the thermal cycling conditions were 96°C for 10 s, 50°C for 5 s, and 60°C for 4 min, which were repeated for 25 cycles. The cycle sequencing reactions were performed for both DNA strands. The nucleotide sequences were aligned and analyzed with GENTYX-Mac software (version 10.0; Software Development Co., Tokyo, Japan). The *cag* PAI gene sequences of strains 26695 (25), J99 (2), and NCTC11638 (8) published previously were also included in the analysis.

TABLE 1. Characteristics of the selected strains

H. pylori strain	Origin	Disease	*vacA* genotype
26695	United Kingdom	Gastritis	s1a/m1
J99	United States	Duodenal ulcer	s1b/m1
11638	Australia	Gastritis	s1a/m1
F16	Fukui	Gastritis	s1c/m1
F17	Fukui	Gastritis	s1c/m1
F28	Fukui	Gastritis	s1c/m1
F32	Fukui	Gastric cancer	s1c/m1
F79	Fukui	Gastric ulcer	s1a/m1
F80	Fukui	Duodenal ulcer	s1b/m1
OK101	Okinawa	Gastritis	s1c/m1
OK107	Okinawa	Gastritis	s1a/m2
OK109	Okinawa	Gastritis	s1a/m1
OK112	Okinawa	Gastritis	s2/m2
OK129	Okinawa	Gastritis	s1c/m2

Phylogenetic analysis. To clarify the phylogenetic relationships between Japanese strains and previously characterized *H. pylori* strains from the West, the nucleotide sequences of the *cag* PAI genes were aligned by using DNASIS Pro software (Hitachi Software Engineering Co., Tokyo, Japan). A phylogenetic tree was constructed by using the unweighted pair group method and the same software.

Detection of *vacA* gene diversity. Genotyping of *vacA* s- and m-region alleles was performed as previously described in detail (4, 18, 19). Briefly, parts of the *vacA* s and m regions were amplified and directly sequenced by using primers SS1-F, SS3-F, S1c-F, and VA1-R for the s region and primers VA3-F, VA4-F, VA3-R, and VA4-R for the m region. The conditions used for PCR and direct sequencing are described above.

In vitro infection. Human gastric epithelial AGS cells were cultured in RPMI 1640 (GIBCO BRL, Grand Island, N.Y.) containing 10% fetal bovine serum (Filtron Pty Ltd., Brooklin, Australia). *H. pylori* (3×10^8 cells) was added to AGS cells (3×10^6 cells per 100-mm dish), which were then cultured in an antibiotic-free medium at a multiplicity of infection of 100. After incubation in a 5% CO_2 atmosphere for 5 h, infected cells were washed three times with 0.01 M PBS (pH 7.5) containing 2 mM Na_3VO_4, lysed in ice-cold 1% Triton X-100 buffer (50 mM Tris-HCl [pH 7.4], 1% Triton X-100, 5 mM EDTA, 1 mM Na_3VO_4, 10 μg of leupeptin per ml, 10 μg of aprotinine per ml, 100 μM *p*-tosyl-L-phenylalanine chloromethyl ketone, 100 μM *p*-tosyl-L-lysine chloromethyl ketone, 1 mM phenylmethylsulfonyl fluoride), and subjected to immunoprecipitation and immunoblotting.

Antibodies. The primary antibodies used for immunoprecipitation and immunoblotting were an anti-CagA polyclonal antibody (Austral Biologicals, San Ramon, Calif.), an antiphosphotyrosine antibody (4G10; Upstate Biotechnology Inc. Lake Placid, N.Y.), and an anti-SHP-2 antibody (C-18; Santa Cruz Biotech. Inc., Santa Cruz, Calif.).

Immunoprecipitation and immunoblotting. The cell lysates were centrifuged at 10,000 × *g* for 10 min at 4°C, and the supernatant was subsequently immunoprecipitated with the anti-CagA polyclonal antibody or control normal immunoglobulin G for 30 min at 4°C, after which protein G-Sepharose beads (Amer-

TABLE 2. *H. pylori cag* PAI-spanning primer sets

Primer position in strain 26695	Designation in the following study:		Primer			Size of PCR product (bp)
	Tomb et al. (25)	Censini et al. (8)	PAI primer	Nucleotide sequence	Primer set	
546899	HP0519		PAI-1 S	5′-ACACTGCCAAGCCCGATGCTGTA-3′	PAI-1 S, PAI-1 AS	1,264
547464	HP0520	*cag* ζ	PAI-2 S	5′-GTGTCTTGAGCGGTGCTATG-3′	PAI-2 S, PAI-2 AS	1,366
547652			HP0520 AS1	5′-CAGTTGGTTCGTTGGTAAC-3′		
547850			HP0521 S1	5′-GCTGTAAGGGCGTTTTAC-3′		
548162			PAI-1 AS	5′-GATACAGCGGTTGCTAGT-3′		
548526			HP0522 S4	5′-GACTTATGCGGATCTCATTA-3′		
548547			PAI-3 S	5′-CATCACAGGCTCATTAGAG-3′	PAI-3 S, PAI-3 AS	1,233
548829			PAI-2 AS	5′-ATCTCTTAGGGGCGAACACACTTC-3′		
549011			HP0522 S2	5′-CCCAACTCGTAGCCAATGAAGAAG-3′		
549312	HP0523	*cag* γ	HP0522 AS3	5′-GATCAAAGTCCCCTCATAGC-3′		
549703			PAI-4 S	5′-GGTTGCGACAATGAAGTG-3′	PAI-4 S, PAI-4 AS	982
549779			PAI-3 AS	5′-CTGTTGTTCAACCCTAGAGAG-3′		
550007			PAI-5 S	5′-GCGCTTACAATGGGGGAATGAA-3′	PAI-5 S, PAI-5 AS	1,327
550032	HP0524	*cag* β (*virD4*)	HP0523 S2	5′-CAACCCTAATGGCGCTTACGTGAA-3′		
550227			HP0524 S4	5′-CTTGAACCCACAGGCACTAAAGA-3′		
550423			HP0524 AS6	5′-CCTATCAAGTGCCACAAG-3′		
550684			PAI-4 AS	5′-GGTAGGAATGGCGCTAAGAC-3′		
550666			PAI-6 S	5′-TCTTAGCGCCATTCCTACCATAACC-3′	PAI-6 S, PAI-6 AS	1,104
550866			HP0524 AS4	5′-TTCTGCCAATCCATGATCCACAGTG-3′		
551230			PAI-7 S	5′-AGTTGATCCCGCTTGCCATAGAAC-3′	PAI-7 S, PAI-7 AS	1,503
551333			PAI-5 AS	5′-CATGTGGACTAAAAAGGGGCTTGAG-3′		
551653			HP0524 S5	5′-CCATAGTGTCAGCTTTAGGGTCA-3′		
551769			PAI-6 AS	5′-CGTTCATTGGCTTGATTGCTCCTAC-3′		
552250	HP0525	*cag* α (*virB11*)	HP0524 AS5	5′-GGCTTTTGGTTGCAAGCTATAC-3′		
552610			PAI-8 S	5′-GGCCAAACGGATAAACGCTTCTTCA-3′	PAI-8 S, PAI-8 AS	1,021
552732			PAI-7 AS	5′-ATTTTAGGGGAACTCAGAAGCAGTG-3′		
553114	HP0526	*cag* Z	PAI-9 S	5′-GACAATCTGCACCCTTTCAC-3′	PAI-9 S, PAI-9 AS	1,122
553630			PAI-8 AS	5′-GTGGCGTTTCAGATCCTAGGGATAG-3′		
553948			HP0526 AS2	5′-GATAGCAACGATCCGCAAGA-3′		
553930	HP0527	*cagY* (*virB10*)	PAI-10 S	5′-CTTGCGGATCGTTGCTATCT-3′	PAI-10 S, PAI-10 AS	1,041
554235			PAI-9 AS	5′-ATCACCACAAGCCCCAAAGGT-3′		
554504			PAI-11 S	5′-CTACCTTTGCCAAGGCCTATGAGT-3′	PAI-11 S, PAI-11 AS	1,283
554970			PAI-10 AS	5′-GAAACAAGCCCTGTCAAACAGG-3′		
555368			PAI-12 S	5′-GGATAACCTTTAGCCGCCATGT-3′	PAI-12 S, PAI-12 AS	1,829
555786			PAI-11 AS	5′-GACAGAGCGGCTATCATGAAGTGT-3′		
555763			PAI-13 S	5′-ACACTTCATGATAGCCGCTCTGTC-3′	PAI-13 S, PAI-13 AS	2,931
557196			PAI-12 AS	5′-GAGAAATTGCTCACCCTTGAATC-3′		
558446			PAI-14 S	5′-CAATCTAGCGCCACTTGAAC-3′	PAI-14 S, PAI-14 AS	623
558693			PAI-13 AS	5′-GAGGACAAAAACCCGTTGAGAG-3′		
558925			PAI-15 S	5′-CACGATAAGAACAGCGACTAC-3′	PAI-15 S, PAI-15 AS	1,134
559068			PAI-14 AS	5′-CTTGACAACCCCACAGAAACTC-3′		
559445	HP0528	*cagX* (*virB9*)	PAI-16 S	5′-GTGGGGTTGTCAAGATGATGATCTG-3′	PAI-16 S, PAI-16 AS	996
560058			PAI-15 AS	5′-CTATGGTGAATTGGAGCGTGTG-3′		
560207			PAI-17 S	5′-CAATGGCGGCATCAGTCATGCTCAA-3′	PAI-17 S, PAI-17 AS	977
560440			PAI-16 AS	5′-ATCAAGCAAAGGCGCTAGAGACTCA-3′		
560909			PAI-18 S	5′-CTTGCATGTCCTCTAGTCGTTCCAT-3′	PAI-18 S, PAI-18 AS	837
561183			PAI-17 AS	5′-ACTTATCGTAGATGCGCCTGACC-3′		
561561	HP0529	*cagW* (*virB8*)	PAI-19 S	5′-TGCCTGCCCCATCAACAATTCCTCT-3′	PAI-19 S, PAI-19 AS	1,008
561745			PAI-18 AS	5′-GAGCGTCAATGCGATCGTTAATACC-3′		
561843			PAI-20 S	5′-TAGCAACAGAGGGCGTTATG-3′	PAI-20 S, PAI-20 AS	1,130
562568			PAI-19 AS	5′-TCAAAGGAGCGGACGCTGCTGTT-3′		
562738			PAI-21 S	5′-GTCCTCAACACCGCCTTTGGTAAA-3′	PAI-21 S, PAI-21 AS	1,024
562972	HP0530	*cagV*	PAI-20 AS	5′-CACAAGTTTAGCCGCTAGCA-3′		
563540			PAI-22 S	5′-GATAGCTTCTGCTCGGACTT-3′	PAI-22 S, PAI-22 AS	970
563761			PAI-21 AS	5′-GTCAAACGCTCCGATGCTAG-3′		
563925	HP0531	*cagU*	PAI-23 S	5′-CTATCAAGGGCTATCACACC-3′	PAI-23 S, PAI-23 AS	1,107
564509			PAI-22 AS	5′-TCTTTGCTCCCCTAAACTCC-3′		
564610			HP0531 S1	5′-GACAAGCCAAACAGAGAATA-3′		
564889	HP0532	*cagT* (*virB7*)	PAI-24 S	5′-GTAGCACTAACGACAAGGTGCT-3′	PAI-24 S, PAI-24 AS	1,087
565031			PAI-23 AS	5′-TGCACCGCCTTGTTTCTTTG-3′		
565573			PAI-25 S	5′-CTTGCATGGCTATGATGTGAG-3′	PAI-25 S, PAI-25 AS	983
565975			PAI-24 AS	5′-TAACGCCCGTTGGCGTTTCTCT-3′		
566268			PAI-26 S	5′-GGGAGCTTAGTGCCATACAA-3′	PAI-26 S, PAI-26 AS	842
566555			PAI-25 AS	5′-GCATACAAACAAGGGAGCGTTAG-3′		

Continued on following page

TABLE 2—*Continued*

P'rimer position in strain 26695	Designation in the following study:		Primer			Size of PCR product (bp)
	Tomb et al. (25)	Censini et al. (8)	PAI primer	Nucleotide sequence	Primer set	
566816			PAI-27 S	5′-CAGAGCGGTCATAATTCAAAGAGC-3′	PAI-27 S, PAI-27 AS	1,342
567109			PAI-26 AS	5′-TCATCTTTCACGCAGAGC-3′		
567090			HP0535 S3	5′-CAACTCTGCGTTCAGTGTGTTGAC-3′		
567212			HP0535 S2	5′-CCAACCAAAGCAGATCCCATGT-3′		
567426			HP0535 AS2	5′-TTGTTGGGTGGCGGAACAAA-3′		
567409			PAI-28 S	5′-TGTTCCGCCACCCAACAAAGAA-3′	PAI-28 S, PAI-28 AS	1,174
568157			PAI-27 AS	5′-CTTATGGGGCAGGGGTGATTTTAG-3′		
568144	HP0537	*cagM*	PAI-29 S	5′-CCCTGCCCCATAAGAAAA-3′	PAI-29 S, PAI-29 AS	1,160
568582			PAI-28 AS	5′-GTATGCGGCTTGTTGGTA-3′		
568783			HP0537 AS2	5′-CCACATTAGCCGACAAAACTCC-3′		
568993			PAI-30 S	5′-GAGGCTCTAGAGAAAGAGAC-3′	PAI-30 S, PAI-30 AS	1,086
569303			PAI-29 AS	5′-GCTAATCGGCTCGCTTTT-3′		
569724	HP0538	*cagN*	PAI-31 S	5′-CGTAGATAGCGATCCTATG-3′	PAI-31 S, PAI-31 AS	1,331
570078			PAI-30 AS	5′-CTCTCAAAGCGTTAGTGG-3′		
570169	HP0539	*cagL*	PAI-32 S	5′-AAGCGGCTAAGCACAAAG-3′	PAI-32 S, PAI-32 AS	1,311
571054			PAI-31 AS	5′-CACAGACGCTTGTAGAAAG-3′		
571040			PAI-33 S	5′-CTACAAGCGTCTGTGAAG-3′	PAI-33 S, PAI-33 AS	880
571479	HP0540	*cagI*	PAI-32 AS	5′-AGAGACCAACCAACAAGTGC-3′		
571741			HP0539 ASI	5′-GATTTGAACGCGCTCATAG-3′		
571729			PAI-34 S	5′-GCGCGTTCAAATCTACTG-3′	PAI-34 S, PAI-34 AS	1,350
571919			PAI-33 AS	5′-CACAAAATGCCCCTATCTTG-3′		
572648	HP0541	*cagH*	HP0540 AS3	5′-GCTTGAACCCGCCTTAAA-3′		
572894			PAI-35 S	5′-TCGCTTGAGTGTCATAGG-3′	PAI-35 S, PAI-35 AS	1,271
573078			PAI-34 AS	5′-CACTCCTGCATGCCCTATTG-3′		
573766			HP0541 AS2	5′-GTGTTGCAGGGCCATTTG-3′		
573832	HP0542	*cagG*	PAI-36 S	5′-GCTTGTGTACCTGCCATGTT-3′	PAI-36 S, PAI-36 AS	783
574164	HP0543	*cagF*	PAI-35 AS	5′-AAATAGCGTGGGGCTTGT-3′		
574444			PAI-37 S	5′-GCTTCAACGCTCATATCAG-3′	PAI-37 S, PAI-37 AS	1,089
574614			PAI-36 AS	5′-CATAAGCGAGGACATGCAGAAC-3′		
574991	HP0544	*cagE* (*virB4.picB*)	PAI-38 S	5′-GAGCGGTAAGGTTTTGTTCGGTGAT-3′	PAI-38 S, PAI-38 AS	1,594
575324			HP0544 S4	5′-GGGCTTCCATCCTGTTTGTAGAGAA-3′		
575532			PAI-37 AS	5′-GTCAGACTTGCGACTCAAAG-3′		
576092			PAI-39 S	5′-GCCGCCCAAGCAAAAGGATTTA-3′	PAI-39 S, PAI-39 AS	1,498
576161			HP0544 AS4	5′-CAATGGGTGGGGAGTATGTCAAGA-3′		
576584			PAI-38 AS	5′-CCAACGCAGCGACTTTCTCTATG-3′		
577004			HP0544 AS5	5′-TGCATGGTGGGGTGAAAGAAGTTTA-3′		
577255			PAI-40 S	5′-CATAACGGGTTCATTGAGAGTGTCT-3′	PAI-40 S, PAI-40 AS	1,051
577589			PAI-39 AS	5′-ATGGGGTGATCCTTACTAACAACTA-3′		
578096	HP0545	*cagD*	PAI-41 S	5′-GCCACAAACACCCCTCTCTTTA-3′	PAI-41 S, PAI-41 AS	921
578305	HP0546	*cagC*	PAI-40 AS	5′-CCAACAAACAACGCTGCTTTC-3′		
578847			PAI-42 S	5′-CAGTCGCCTGACCTCTTTTGAT-3′	PAI-42 S, PAI-42 AS	1,594
579016			PAI-41 AS	5′-CAATCCTTTAATGGCGGTCACCAG-3′		
579342			HP0546 S2	5′-CAAACCCAAGCTGATCAGAGTGAG-3′		
579530			HP0546 AS1	5′-TTTGGTTTGTGTGTGTCATACT-3′		
579910	HP0547	*cagA*	PAI-43 S	5′-AAGGAGAAACAATGACTAACGAAACTATTG-3′	PAI-43 S, PAI-43 AS	1,066
580440			PAI-42 AS	5′-CTGCAAAAGATTGTTTGGCAGA-3′		
580674			PAI-44 S	5′-ATACAAGGCTTACCGCCTG-3′	PAI-44 S, PAI-44 AS	1,675
580975			PAI-43 AS	5′-GTAGCCACATTGTCGCCTTGTTGG-3′		
581449			HP0547 C1(+)	5′-ATTTCAAATACACCAACGCCTCCAA-3′	HP0547 C1(+), HP0547 L2(−)	2,043
581979			PAI-45 S	5′-GAATTGTCTGATAAACTTGAAA-3′	PAI-45 S, PAI-45 AS	1,152
582348			PAI-44 AS	5′-GCTCTACCTTACTGAAATCGCC-3′		
582974			HP0547 C4(−)	5′-AGCTTCTGATACCGCTTGACTG-3′		
583130			PAI-45 AS	5′-GCGTATGTGGCTGTTAGTAGCG-3′		
583193	HP0548	*cagΩ*	PAI-46 S	5′-AAACCCTGAGTGGCTCAAGCTC-3′	PAI-46 S, PAI-46 AS	829
583491			HP0547 L2(−)	5′-TCCTTTAAGATTTTTGGAAACCACCTTTTG-3′		
583718			PAI-47 S	5′-CAGCTCCCACCTAGCATTGAT-3′	PAI-47 S, PAI-47 AS	1,514
584021			PAI-46 AS	5′-GTTGATGCTCCCCTTCAACA-3′		
584386			HP0548 S2	5′-GCGAAGCGATGAGAAGAA-3′		
584646			HP0549 AS1	5′-ATTTCATGCGAGCGGCGATGTG-3′		
584873			PAI-48 S	5′-CTAGCAATTCGCCCTCTA-3′	PAI-48 S, PAI-48 AS	772
585231			PAI-47 AS	5′-CACTAAAGACCCCACCAC-3′		
585644			PAI-48 AS	5′-TTAAAGGCACCGGGAATAGC-3′		

TABLE 3. General comparative features of *cag* PAI

H. pylori strain	Size (bp)	No. of genes	% G+C content	Presence of IS605	CagA type
26695	37,388	29	35.76	−	Western
J99	37,729	27	35.97	−	Western
11638	38,477	33	35.83	+	Western
F16	36,739	27	35.60	−	East Asian
F17	37,119	27	35.52	−	East Asian
F28	36,914	27	35.64	−	East Asian
F32	36,534	27	35.72	−	East Asian
F79	37,153	27	35.64	−	Western
F80	37,209	28	35.89	−	Mixed type
OK101	36,715	27	35.49	−	East Asian
OK107	37,251	27	35.56	−	Western
OK109	36,892	27	35.57	−	East Asian
OK112	36,714	27	36.00	−	Western
OK129	37,342	27	35.60	−	East Asian

sham Pharmacia Biotech. Inc., Piscataway, N.J.) were added for 90 min at 4°C. The immunoprecipitates were washed four times with the lysis buffer and then boiled with a sodium dodecyl sulfate (SDS)-electrophoresis sample buffer (2% SDS, 10% glycerol, 6% 2-mercaptoethanol, 0.003% bromophenol blue, 50 mM Tris-HCl [pH 6.8]) for 5 min. Equal amounts of samples from homogenates or immunoprecipitates were separated by SDS-polyacrylamide gel electrophoresis (7.5% polyacrylamide) and blotted onto an Immobilon P membrane (Millipore Corp., Bedford, Mass.). The membranes were blocked with 1% bovine serum albumin or 5% skim milk in T-PBS (10 mM Tris-HCl [pH 7.5], 100 mM NaCl, 0.5% Tween 20) and incubated with a primary antibody in T-PBS overnight at 4°C. After the membranes were washed with T-TBS, they were incubated with horseradish peroxidase-conjugated goat anti-rabbit or anti-mouse immunoglobulin G polyclonal antibodies in T-TBS for 1 h and visualized with an enhanced chemiluminescence detection system, as directed by the manufacturer (Amersham Pharmacia Biotech. Inc.).

Nucleotide sequence accession numbers. The DNA sequences of the complete *cag* PAI sequences of 11 representative Japanese strains characterized here were deposited in the DDBJ database under accession numbers AB120416 for F16, AB120417 for F17, AB120418 for F28, AB120419 for F32, AB120420 for F79, AB120421 for F80, AB120422 for OK101, AB120423 for OK107, AB120424 for OK109, AB120425 for OK112, and AB120426 for OK129.

RESULTS

Construction of *cag* PAI. All 11 Japanese strains contained the complete *cag* PAI flanked by the same chromosomal genes and the previously described integration signal but lacked the insertion sequence IS*605* elements associated with the *cag* PAI in strain NCTC11638. The general comparative features of the *cag* PAIs are shown in Table 3. The sizes of the complete *cag* PAI ranged from 36,534 to 38,477 bp. The numbers of *cag* PAI genes ranged from 27 to 33. The G+C contents of the full-length *cag* PAI ranged from 35.49 to 36.00%.

Comparative analysis of *cag* PAI genes. The nucleotide sequences of all *cag* PAI genes of the strains were aligned. The degrees of homology of both the nucleotide sequence of the coding regions of the entire *cag* PAI and the amino acid sequences between each isolate are shown in Table 4. The degrees of nucleotide sequence identity ranged from 88.26 to 96.73%, and the degrees of amino acid sequence identity ranged from 87.14 to 96.76%.

Comparison of each *cag* PAI gene showed minor differences among strains. Strains F16, F17, F28, F32, F79, F80, OK101, and OK129 did not have HP0521. Strains OK107 and OK109 contained an HP0521 open reading frame (ORF) (348 bp, compatible with ORF7 of 11638) but did not have a stop codon until the end of HP0522. Therefore, HP0521 was combined with HP0522 in these strains. Strain OK112 did not have HP0533. The orientation of HP0535 was reversed in strains F16, F17, F28, F79, OK101, OK107, OK109, and OK129. None of the Japanese strains except strain F80 had HP0548. The lengths of HP0527 (*virB10*) and HP0547 (*cagA*) varied among the strains (Table 5). To analyze the divergence in each *cag* PAI gene, the nucleotide and amino acid sequence identities were calculated for the strains. The nucleotide and amino acid sequence identities of each *cag* PAI gene were more than 90%, except for HP0523 (*cag*; 73 to 99%) (Table 6), HP0527 (*virB10*; 77 to 95%) (Table 7), HP0535 (*cagQ*; 70 to 100%) (Table 8), and HP0547 (*cagA*; 72 to 97%) (Table 9).

Phylogenetic analysis. The phylogenetic tree of the full-length *cag* PAI demonstrated the genetic relationship among the 11 Japanese strains and 3 previously described strains from the West, strains 26695, J99, and NCTC11638 (Fig. 1). The *cag* PAI genes were divided into two major groups: a Western group and a Japanese group. Two strains, strains F79 and OK107, formed an intermediate type. Two Japanese strains,

TABLE 4. Analysis of *cag* PAI divergence

Strain	% Nucleotide and amino acid sequence identity[a]													
	26695	J99	11638	F16	F17	F28	F32	F79	F80	OK101	OK107	OK109	OK112	OK129
26695		91.93	91.18	91.13	91.52	91.59	92.79	92.67	92.73	92.12	89.98	90.22	93.13	92.88
J99	91.29		91.23	91.15	89.64	91.05	91.28	91.73	92.54	90.77	90.12	90.34	93.19	90.54
11638	91.12	90.78		89.42	88.26	89.70	89.10	90.94	89.91	88.92	88.50	88.90	91.67	89.06
F16	90.38	90.54	88.75		95.05	96.33	95.11	94.82	92.05	95.68	91.14	93.32	92.87	95.78
F17	90.72	88.88	87.14	95.00		95.45	94.61	93.64	92.39	96.20	90.20	91.74	90.92	96.16
F28	90.99	90.34	89.08	95.96	95.16		94.74	95.37	92.28	95.07	92.27	93.72	92.96	96.10
F32	92.34	90.80	88.76	94.88	94.56	94.75		93.43	93.10	96.07	90.24	92.16	92.58	94.56
F79	92.69	91.35	91.09	94.14	92.63	94.66	92.78		92.90	93.58	92.13	92.26	93.48	94.59
F80	92.29	92.17	89.50	91.41	91.65	91.77	92.73	92.32		93.05	89.56	90.10	91.68	93.31
OK101	91.78	90.24	88.40	95.72	95.96	95.03	96.08	92.91	92.67		90.02	92.31	91.91	96.73
OK107	89.29	88.97	87.89	90.19	89.12	91.47	89.63	91.58	88.67	89.22		95.43	92.15	90.92
OK109	89.26	89.08	88.07	93.14	91.57	93.89	92.22	91.58	89.53	92.46	94.10		92.24	93.11
OK112	92.95	92.17	91.34	91.90	90.21	92.36	92.02	93.40	91.30	91.38	90.76	90.77		92.22
OK129	92.35	89.83	88.55	95.57	96.22	96.21	96.35	94.01	92.81	96.76	90.17	93.20	91.64	

[a] Values above the diagonal are for nucleotide sequences, and those below the diagonal are for amino acid sequences.

TABLE 5. Length of each *H. pylori cag* PAI gene

Gene designation by Tomb et al. (25)	Gene orientation	*cag* PAI gene length (bp)													
		26695	J99	11638	F16	F17	F28	F32	F79	F80	OK101	OK107	OK109	OK112	OK129
HP0520	+	345	345	345	345	345	345	345	345	345	345	345	345	345	345
HP0521	+	240	657	348								1,833	1,791	735	
HP0522	+	1,443	1,443	1,443	1,443	1,443	1,443	1,443	1,443	1,443	1,443			1,443	1,443
HP0523	+	507	507	507	507	507	507	507	507	507	507	420	507	507	507
HP0524	−	2,244	2,244	2,244	2,244	2,244	2,244	2,244	2,244	2,244	2,244	2,244	2,244	2,244	2,244
HP0525	−	990	990	990	990	990	990	990	990	990	990	990	990	990	990
HP0526	−	597	597	597	597	597	597	597	597	597	597	597	597	597	597
HP0527	−	5,781	5,457	5,169	5,388	5,460	5,385	6,006	5,439	5,682	5,430	5,391	5,304	5,406	5,781
HP0528	−	1,566	1,566	1,566	1,563	1,563	1,563	1,563	1,563	1,566	1,563	1,566	1,566	1,566	1,566
HP0529	−	1,605	1,605	1,605	1,605	1,605	1,605	1,605	1,605	1,605	1,605	1,605	1,605	1,605	1,605
HP0530	−	756	756	756	756	756	756	756	756	756	756	756	756	756	756
HP0531	+	654	654	648	645	645	645	645	645	654	645	648	645	648	645
HP0532	+	840	840	840	840	840	840	840	840	840	840	840	840	840	840
HP0533	−	87			36	36	36	36	36	36	36	36	36		87
HP0534	−	588	597	597	597	597	597	597	597	597	597	597	597	597	597
HP0535	+				309	303	303		309		378	303	303		303
	−	378	378	378				378		303				378	
HP0536	−	342	342	330	342	342	342	342	342	342	342	321	342	342	342
HP0537	+	1,128	1,128	1,128	1,128	1,128	1,128	1,128	1,128	1,128	1,128	1,128	1,128	1,128	1,128
HP0538	+	918	918	918	918	918	918	918	918	918	918	918	918	918	918
HP0539	−	711	711	711	711	711	711	711	711	711	711	711	711	711	705
HP0540	−	1,143	1,143	1,143	1,143	1,143	1,143	1,143	1,143	1,143	1,143	1,143	1,143	1,143	1,143
HP0541	−	1,110	1,110	1,110	1,110	1,110	1,110	1,110	1,110	1,110	1,110	1,110	1,110	1,110	1,110
HP0542	−	426	426	426	429	426	429	426	426	426	429	426	429	426	426
HP0543	−	804	804	804	804	804	804	804	804	804	804	804	804	804	804
HP0544	−	2,949	2,949	2,943	2,949	2,949	2,949	2,949	2,949	2,949	2,949	2,949	2,949	2,949	2,949
HP0545	−	621	624	627	621	621	621	621	621	621	621	621	621	621	621
HP0546	−	345	345	345	345	342	345	345	345	345	345	342	345	345	345
HP0547	+	3,558	3,501	3,441	3,516	3,825	3,714	3,519	3,648	3,666	3,516	3,753	3,534	3,540	3,516
HP0548	+	825		381						381					

strains F80 and OK112, were included in the same cluster as the Western strains.

A phylogenetic tree constructed on the basis of the *cag* PAI sequences was analyzed according to the *vacA* or *cagA* genotype. All strains with the s1c genotype in the *vacA* signal sequence, which is considered the characteristic genotype in East Asia, were in the Japanese cluster, as determined by *cag* PAI gene sequencing. In addition, all strains with the East Asian-type *cagA* genotype were also in the Japanese cluster. Two intermediate strains, strains F79 and OK107, and all strains in the Western cluster except strain F80 had the Western-type *cagA* genotype (strain F80 had a mixed-type CagA and a mixed Western and East Asian type).

The phylogenetic tree was analyzed according to the clinical features. Although statistical analysis could not be applied to the small number of patients ($n = 11$) evaluated in this study, patients infected with strains in the Japanese cluster tended to have high-grade inflammation and a high level of gastritis

TABLE 6. Analysis of HP0523 (*cag* γ) gene divergence

Strain	% Nucleotide and amino acid sequence identity[a]													
	26695	J99	11638	F16	F17	F28	F32	F79	F80	OK101	OK107	OK109	OK112	OK129
26695		93.14	92.35	90.20	89.80	88.63	89.02	89.41	89.80	89.80	76.67	92.55	91.18	89.22
J99	94.12		90.39	87.25	87.25	86.47	87.06	87.06	87.25	87.25	74.90	90.00	88.82	86.67
11638	92.94	90.59		87.84	87.84	87.65	88.04	88.04	91.57	87.65	76.67	92.16	93.73	87.65
F16	93.53	91.18	90.59		96.08	95.29	94.12	95.29	90.59	95.29	76.08	92.35	90.00	95.49
F17	94.12	91.76	91.18	98.24		96.27	96.27	98.43	93.33	98.82	75.69	92.35	89.02	99.02
F28	92.35	90.59	89.41	96.47	95.88		97.65	95.29	90.98	95.29	77.06	92.35	88.63	95.29
F32	91.18	90.00	89.41	95.88	95.29	97.65		96.47	91.76	96.08	75.69	91.57	88.24	95.29
F79	92.35	90.59	90.59	97.65	98.24	95.29	97.06		92.94	98.04	75.88	92.35	89.02	98.63
F80	92.35	88.24	92.35	97.35	92.94	91.18	91.18	93.53		93.73	77.45	93.14	91.37	92.35
OK101	93.53	91.18	90.59	97.65	99.41	95.29	95.88	98.82	93.53		75.88	92.35	89.02	97.84
OK107	73.53	73.53	74.71	75.29	74.71	75.88	74.12	74.12	74.12	74.12		82.75	78.04	75.49
OK109	92.94	91.76	92.35	95.88	95.29	94.71	93.53	94.71	92.94	94.71	78.24		93.53	92.16
OK112	95.29	91.76	93.53	92.35	92.94	91.18	90.00	91.18	92.94	92.35	75.29	95.29		89.61
OK129	93.53	91.18	90.59	97.65	99.41	95.29	94.71	97.65	92.35	98.82	74.12	94.71	92.35	

[a] Values above the diagonal are for nucleotide sequences, and those below the diagonal are for amino acid sequences.

TABLE 7. Analysis of HP0527 (*cagY*) gene divergence

Strain	% Nucleotide and amino acid sequence identity[a]													
	26695	J99	11638	F16	F17	F28	F32	F79	F80	OK101	OK107	OK109	OK112	OK129
26695		83.47	85.06	84.54	89.85	88.77	91.72	88.59	89.76	90.04	88.45	87.59	89.61	95.18
J99	80.70		87.06	84.06	80.17	87.42	82.82	87.29	83.90	81.26	87.18	87.46	87.35	82.98
11638	84.34	85.67		86.34	82.03	90.45	81.64	91.02	82.33	82.65	90.87	91.06	90.09	85.24
F16	82.10	82.65	84.40		85.16	91.92	84.14	88.86	81.12	83.54	90.34	89.84	90.06	85.57
F17	88.43	77.61	78.68	84.64		84.98	87.95	84.45	87.01	93.65	84.42	84.16	83.51	92.10
F28	88.38	84.66	90.04	90.08	85.06		87.11	93.11	83.55	85.50	95.23	94.40	94.20	90.28
F32	90.92	81.32	80.98	82.75	87.43	87.22		85.50	87.14	88.45	86.21	85.15	85.93	83.30
F79	87.66	85.85	90.90	86.84	83.67	93.34	84.83		84.16	83.32	93.86	91.48	91.75	89.41
F80	86.98	82.96	79.78	78.47	84.94	81.88	85.83	81.92		88.26	83.34	84.92	83.26	90.31
OK101	89.26	79.52	80.68	82.71	92.12	85.45	88.07	82.77	86.93		83.88	84.78	83.64	91.43
OK107	88.02	84.81	90.55	88.34	84.39	95.61	86.18	93.73	81.62	83.39		93.48	94.10	89.48
OK109	86.63	85.17	89.94	87.70	82.22	94.61	85.12	90.75	83.49	84.13	93.33		92.71	88.19
OK112	88.90	84.07	89.52	87.57	82.73	93.90	85.08	90.94	80.79	82.77	93.79	91.91		88.47
OK129	93.83	80.25	84.49	83.62	92.01	90.51	92.96	88.75	88.69	90.72	89.52	87.87	87.81	

[a] Values above the diagonal are for nucleotide sequences, and those below the diagonal are for amino acid sequences.

activity compared to those for patients infected with strains in the Western cluster. All patients infected with Japanese-cluster strains had severe gastric mucosal atrophy. In contrast, patients infected with Western-cluster strains had mild gastric mucosal atrophy (Table 10).

CagA translocation, phosphorylation, and SHP-2 binding activity. All strains in the Japanese cluster possessed CagA translocation and phosphorylation activities and bound to SHP-2. Strains in the Japanese cluster, based on the phylogenetic tree constructed by using the full-length *cag* PAI sequences (strains F16, F17, F28, F32, OK101, OK109, and OK129), had SHP-2 binding activities stronger than that of a Western-cluster strain, strain OK112 (Fig. 2).

DISCUSSION

H. pylori is believed to exhibit a large degree of genomic and allelic diversity. Strain-specific diversity has been proposed to be involved in the organism's ability to cause different diseases (4, 27, 30). The severity of *H. pylori*-related disease is correlated with the presence of the *cag* PAI. Infection with *cag* PAI-positive *H. pylori* is statistically associated with duodenal ulceration, gastric mucosal atrophy, and gastric cancer (7, 8, 10). The *cag* PAI contains genes constituting a type IV secretion system, as well as the *cagA* gene, which encodes the CagA protein. During the bacterium-gastric epithelial cell interaction, *H. pylori* injects CagA directly into the attached cells by means of the bacterial type IV secretion system. The translocated CagA protein undergoes tyrosine phosphorylation in the host cells via an Src family protein, tyrosine kinase (3, 21, 23, 24). It has recently been discovered (17) that translocated CagA forms a physical complex with SRC homology 2 domain-containing tyrosine phosphatase SHP-2 in a phosphorylation-dependent manner and stimulates the phosphatase activity. SHP-2 is known to play an important positive role in mitogenic signal transduction (14). Also, SHP-2 is actively involved in the regulation of spreading, migration, and adhesion of cells (15, 33). Deregulation of SHP-2 by CagA may induce abnormal proliferation and movement of gastric epithelial cells, a cellular condition that eventually leads to severe gastritis and gastric carcinoma. In Japan, nearly 100% of the strains possess CagA (19), and the incidence of atrophic gastritis and gastric cancer is quite high compared with that in Western countries (9). In

TABLE 8. Analysis of HP0535 (*cagQ*) gene divergence

Strain	% Nucleotide and amino sequence identity[a]													
	26695	J99	11638	F16	F17	F28	F32	F79	F80	OK101	OK107	OK109	OK112	OK129
26695		97.38	97.90	74.28	72.44	73.23	97.90	73.49	79.00	89.24	72.97	73.23	97.90	73.23
J99	96.85		97.90	74.80	72.44	73.23	98.43	73.49	77.69	90.03	72.97	73.23	97.90	73.23
11638	97.64	97.64		75.07	72.70	73.49	98.43	73.75	77.95	90.03	73.23	73.49	100.00	73.49
F16	73.23	73.23	74.02		96.79	96.47	74.28	97.12	88.78	81.10	95.51	96.47	75.07	96.47
F17	70.87	71.65	71.65	96.15		98.37	72.44	94.87	89.87	79.27	97.39	98.37	72.70	97.71
F28	70.87	71.65	71.65	95.19	97.06		73.23	95.51	90.85	79.00	98.37	98.69	73.49	98.69
F32	97.64	99.21	98.43	72.44	70.87	70.87		73.49	78.48	89.24	72.97	73.23	98.43	73.23
F79	71.65	71.65	72.44	95.19	92.31	94.23	70.87		87.82	78.74	93.91	94.87	73.75	94.87
F80	77.95	75.59	75.59	84.62	85.29	85.29	76.38	82.69		72.70	90.52	90.85	77.95	90.85
OK101	85.83	86.61	86.61	81.10	78.74	77.95	85.83	77.17	69.29		78.22	79.53	90.03	79.00
OK107	70.08	70.87	70.87	94.23	96.08	97.06	70.08	91.35	84.31	77.17		97.71	73.23	97.71
OK109	72.44	73.23	73.23	97.12	97.06	98.04	72.44	94.23	87.25	79.53	97.06		73.49	99.02
OK112	97.64	97.64	100.00	74.02	71.65	71.65	98.43	72.44	75.59	86.61	70.87	73.23		73.49
OK129	71.65	72.44	72.44	96.15	96.08	97.06	71.65	93.27	86.27	78.74	96.08	99.02	72.44	

[a] Values above the diagonal are for nucleotide sequences, and those below the diagonal are for amino acid sequences.

TABLE 9. Analysis of HP0547 (*cagA*) gene divergence

Strain	% Nucleotide and amino acid sequence identity[a]													
	26695	J99	11638	F16	F17	F28	F32	F79	F80	OK101	OK107	OK109	OK112	OK129
26695		90.77	92.12	86.15	81.70	83.76	86.11	92.83	87.74	86.00	86.70	86.72	94.50	85.88
J99	85.82		89.88	85.83	80.89	82.76	86.00	88.24	85.59	85.77	87.31	85.75	91.28	85.71
11638	89.91	84.06		84.61	79.78	80.90	84.55	90.21	86.59	84.38	83.70	84.45	92.81	84.24
F16	78.51	78.17	77.78		89.63	91.88	97.79	84.93	87.69	97.19	83.85	96.10	85.14	97.24
F17	74.16	73.22	73.32	88.56		92.18	89.25	81.66	84.05	89.26	83.08	88.79	80.26	89.50
F28	75.84	75.44	74.02	90.15	88.63		91.72	84.74	87.76	91.88	84.80	91.33	82.22	91.98
F32	78.00	78.17	77.61	96.59	88.18	89.92		84.86	87.88	97.08	83.88	96.58	85.29	97.30
F79	89.77	82.64	87.77	77.75	73.01	76.40	77.18		90.83	84.63	88.59	85.08	92.23	84.34
F80	82.99	79.80	82.86	82.27	77.47	81.72	82.03	85.95		87.50	87.96	87.34	88.48	87.77
OK101	78.59	78.34	78.03	96.16	88.32	90.15	95.91	77.26	82.35		83.51	96.18	85.07	98.01
OK107	81.67	83.01	78.79	77.31	74.42	77.23	76.91	82.56	81.75	76.91		84.13	85.56	83.80
OK109	78.69	77.84	77.87	94.57	87.48	89.44	94.91	77.19	81.86	95.08	77.00		85.16	96.58
OK112	90.99	86.33	90.10	77.25	72.44	74.00	77.17	88.82	83.66	77.34	79.43	77.10		84.84
OK129	78.93	78.42	78.11	96.16	88.95	90.56	96.34	77.59	82.60	97.70	77.47	95.00	77.51	

[a] Values above the diagonal are for nucleotide sequences, and those below the diagonal are for amino acid sequences.

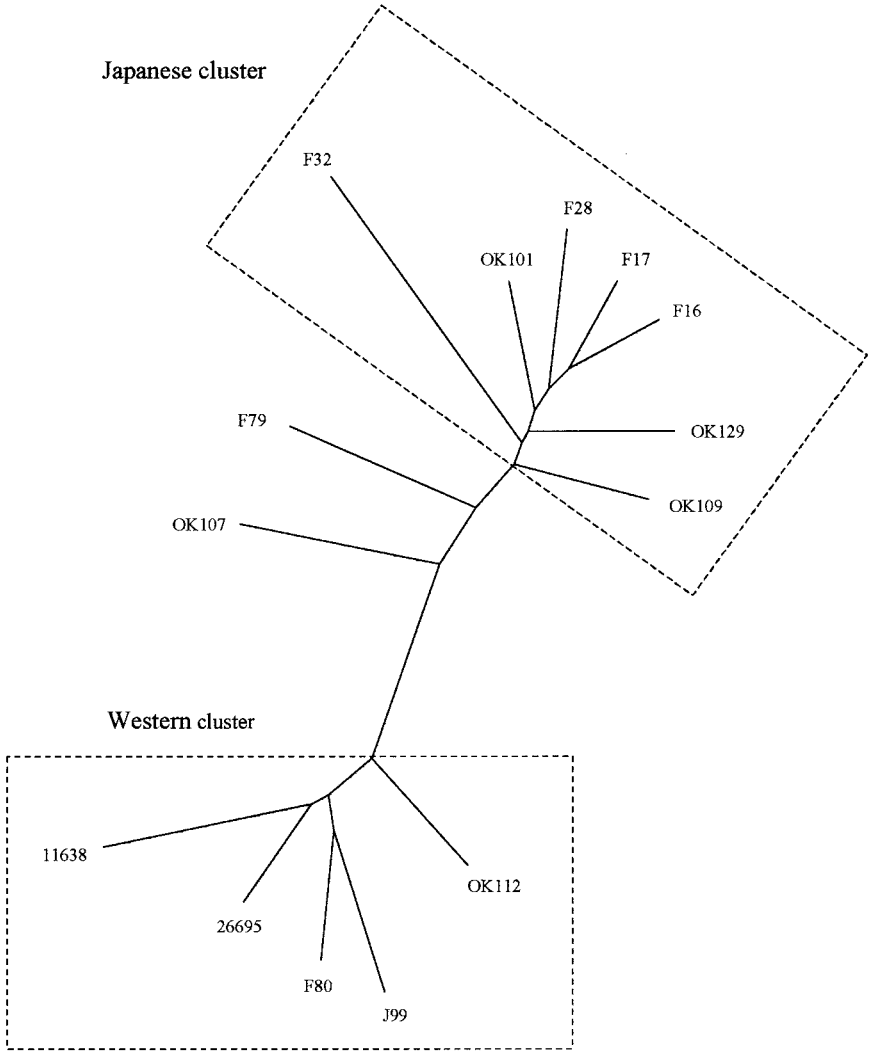

FIG. 1. Phylogenetic tree of the full-length *cag* PAI gene. The *cag* PAI genes were divided into two major groups, a Western group and a Japanese group. Two strains, strains F79 and OK107, formed an intermediate type. Two Japanese strains, strains F80 and OK112, were included in the same cluster as the Western group.

TABLE 10. Relationship between histological features of gastritis and diversity of *cag* PAI

Histological feature and location	Grade of histological feature										
	Japanese-cluster strain							Intermediate strain		Western-cluster strain	
	F16	F17	F28	F32	OK101	OK109	OK129	F79	OK107	F80	OK112
Inflammation											
Antrum	3	3	3	3	3	3	3	3	2	2	3
Body	2	3	3	3	2	3	3	2	1	1	1
Activity											
Antrum	3	3	3	3	2	3	3	3	2	2	2
Body	2	2	2	3	2	2	2	2	1	1	1
Atrophy											
Antrum	3	3	3	3	3	3	3	2	2	1	1
Body	3	3	3	3	3	3	3	2	2	1	1

the present study, we determined the complete *cag* PAI sequences of 11 representative Japanese strains from patients with different pathophysiological conditions in order to examine the association between the diversity of the genes in the *cag* PAI and the clinical outcomes. We demonstrated that the *cag* PAI genes were divided into two major groups, a Western and a Japanese cluster, based on the complete *cag* PAI sequences. Predominant Japanese strains formed a Japanese cluster which was different from the cluster formed by three previously reported Western strains: 26695 from a patient with gastritis in the United Kingdom, J99 from patient with duodenal ulcer in the United Sates, and NCTC11638 from a patient with duodenal ulcer in Australia. The lineage of *H. pylori* isolates infecting Japanese subjects may be different from that of isolates infecting subjects in the West, or a specific strain may have accumulated in the Japanese population. Our results were consistent with those of a recent report by Falush et al. (13). Their analysis of global *H. pylori* samples with the linkage model defined five ancestral populations (Africa 1, Africa 2, East Asia, Europe 1, and Europe 2) (13). The variable genetic structure of the *cag* PAI may influence the clinical outcome of *H. pylori* infection.

We also demonstrated that the diversity of the *cag* PAI is associated with the *cagA* genotype. The CagA protein is highly immunogenic and has a molecular mass of 128 to 140 kDa. The variation in the size of the protein has been correlated with the presence of a variable number of repeat sequences in the 3′ region of the gene (5, 10, 31, 32). The phosphorylation sites are located in the repeat region of CagA (6). Recently, we also discovered that the predominant CagA proteins isolated in East Asia have sequences, which we designated the "East Asian CagA-specific, SHP-2 binding sequence (ESS)," that are distinct from those of CagA proteins isolated in the West. This East Asian-specific sequence conferred stronger SHP-2 binding and transforming activities than the Western CagA sequence by in vitro transfection experiments with a human gastric cancer cell line, AGS (16). In this study, all strains in the Japanese cluster based on the complete *cag* PAI sequences had the East Asian-type *cagA* genotype and had stronger SHP-2 binding activities than Western-type CagA-positive strains. The potential of CagA to disturb host cell functions as a virulence factor could be determined by the degree of SHP-2

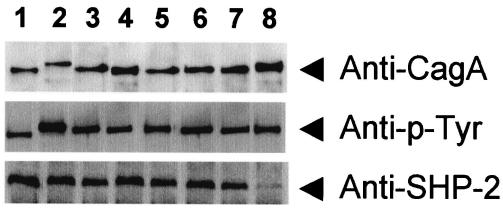

FIG. 2. Immunoblot analysis of AGS cells infected with *H. pylori* by use of anti-CagA, anti-phosphotyrosine (Anti-p-Tyr), and anti-SHP-2 antibodies. Although the band intensities for CagA phosphorylation were similar among the strains, the intensity of the CagA–SHP-2 complex was much higher for Japanese-cluster strains F16 (lane 1), F17 (lane 2), F28 (lane 3), F32 (lane 4), OK101 (lane 5), OK109 (lane 6), and OK129 (lane 7) on the basis of the entire *cag* PAI than for Western-cluster strain OK112 (lane 8).

binding activity. The diversity of the CagA phosphorylation site, which collectively determines the affinity of binding of CagA to SHP-2, may be an important variable in determining the clinical outcome of infection by different *H. pylori* strains. Because SHP-2 plays an important role in both cell growth and motility, deregulation of SHP-2 by CagA may be involved in the induction of abnormal proliferation and movement of gastric epithelial cells, a cellular condition that eventually leads to gastritis and gastric cancer.

The diversity of *cag* PAI was also associated with the *vacA* genotype. All strains with s1c in the signal sequence for the *vacA* genotype, which is considered the characteristic genotype in East Asia (26), were in the Japanese cluster, as determined by *cag* PAI sequencing. 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 (4). It was previously reported (18) that cytotoxin activity is associated with the grade of atrophy. Although there is a genetic linkage between the *cag* PAI and *vacA*, there is a substantial genetic distance between the two genes on the bacterial genome. Thus, the selection of a *vacA* s1c/Japanese *cag* PAI-positive genotype may have a functional basis. In this study, *H. pylori* infection with the Japanese-cluster strain possessing the East Asian CagA and the *vacA* s1c genotype was associated with gastric mucosal atrophy.

In conclusion, our results indicate that *cag* PAI genes are classified into two major groups, a Western group and a Japanese group, and that the structural differences in the *cag* PAIs are substantially associated with the *vacA* s1c genotype and the East Asian *cagA* genotype. The distinct distribution of *cag* PAI diversity in Japan may be involved in the development of atrophic gastritis and may increase the risk of gastric cancer.

ACKNOWLEDGMENTS

This work was supported by the Special Coordination Fund for Promoting Science and Technology, Ministry of Education, Culture, Sports, Science and Technology of Japan, and by a grant-in-aid for Scientific Research (B) from the Japan Society for the Promotion of Science.

REFERENCES

1. **Achtman, M., T. Azuma, D. E. Berg, Y. Ito, G. Morelli, Z. J. Pan, S. Suerbaum, S. A. Thompson, A. van der Ende, and L. J. van Doorn.** 1999. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. Mol. Microbiol. **32:**459–470.

2. **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. de Jonge, 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.
3. **Asahi, M., T. Azuma, S. 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.
4. **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.
5. **Azuma, T., A. Yamakawa, S. Yamazaki, K. Fukuta, M. Ohtani, Y. Ito, M. Dojo, Y. Yamazaki, and M. Kuriyama.** 2002. Correlation between variation of 3′ region of the *cagA* gene in *Helicobacter pylori* and disease outcome in Japan. J. Infect. Dis. **186:**1621–1630.
6. **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.
7. **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 *Helicobcter pylori* strains possessing *cagA* is associated with an increased risk of developing adenocarcinoma of the stomach. Cancer Res. **55:**2111–2115.
8. **Censini, S., C. Lange, Z. Xiang, J. E. Crabtree, P. Ghiara, M. Borodovsky, R. Rappuoli, 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.
9. **Correa, P.** 1988. A human model of gastric carcinogenesis. Cancer Res. **48:**3554–3560.
10. **Covacci, A., S. Censini, M. Bugnoli, R. Petracca, D. Burroni, G. Macchia, A. Massone, E. Papini, Z. Xiang, N. Figura, and R. Rappuoli.** 1993. Molecular characterization of the 128-kDa immunodominant antigen of *Helicobacter pylori* associated with cytotoxicity and duodenal ulcer. Proc. Natl. Acad. Sci. USA **90:**5791–5795.
11. **Covacci, A., J. L. Telford, G. Del Giudice, J. Parsonnet, and R. Rappuoli.** 1999. *Helicobacter pylori* virulence and genetic geography. Science **284:**1328–1333.
12. **Dixon, M. F., R. M. Genta, J. H. Yardley, and P. Correa.** 1996. Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. Am. J. Surg. Pathol. **20:**1161–1181.
13. **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. Katzowitsch, X. Wang, M. Achtman, and S. Suerbaum.** 2003. Traces of human migrations in *Helicobacter pylori* populations. Science **299:**1582–1585.
14. **Feng, G., C. Hui, and T. Pawson.** 1993. SH2-containing phosphotyrosine phosphatase as a target of protein-tyrosine kinases. Science **259:**1607–1611.
15. **Freeman, R., J. Plutzky, and B. G. Neel.** 1992. Identification of a human src homology 2-containing protein-tyrosine-phosphatase: a putative homolog of *Drosophila* corkscrew. Proc. Natl. Acad. Sci. USA **89:**11239–11243.
16. **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.
17. **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.
18. **Ito, S., T. Azuma, H. Murakita, M. Hirai, H. Miyaji, Y. Ito, Y. Ohtaki, Y. Yamazaki, M. Kuriyama, Y. Keida, and Y. Kohli.** 1996. Profile of *Helicobacter pylori* cytotoxin derived from two areas of Japan with different prevalence of atrophic gastritis. Gut **39:**800–806.
19. **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.
20. **Marshall, B. J., C. S. Goodwin, J. R. Warren, R. Murray, E. D. Blincow, S. J. Blackbourn, M. Phillips, T. E. Waters, and C. R. Sanderson.** 1988. Prospective double blind trial of duodenal ulcer relapse after eradication of *Campylobacter pylori*. Lancet **ii:**1437–1442.
21. **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.
22. **Parsonnet, J., G. D. Friedman, D. P. Vandersteen, Y. Chang, J. H. Vogelman, N. Orentreich, and R. K. Sibley.** 1991. *Helicobacter pylori* infection and the risk of gastric carcinoma. N. Engl. J. Med. **325:**11127–11131.
23. **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.
24. **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.
25. **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. Fitzegerald, N. Lee, M. D. Adams, E. K. Hickey, D. E. Berg, J. D. Gocayne, T. R. Utterback, J. D. Peterson, J. M. Kelley, 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.
26. **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.
27. **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.
28. **Warren, J. R., and B. J. Marshall.** 1983. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. Lancet **i:**1273–1275.
29. **Wotherspoon, A. C., C. Ortiz-Hidalgo, M. R. Falzon, and P. G. Isaacson.** 1991. *Helicobacter pylori*-associated gastritis and primary B-cell gastric lymphoma. Lancet **338:**1175–1176.
30. **Xiang, Z., S. Censini, P. F. Bsyeli, 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.
31. **Yamaoka, Y., H. M. T. El-Zimaity, O. Gutierrez, N. Figura, J. G. Kim, T. Kodama, K. Kashima, D. Y. Graham, and J. K. Kim.** 1999. Relationship between the *cagA* 3′ repeat region of *Helicobacter pylori*, gastric histology, and susceptibility to low pH. Gastroenterology **117:**342–349.
32. **Yamaoka, Y., T. Kodama, K. Kashima, D. Y. Graham, and A. R. Sepulveda.** 1998. Variants of the 3′ region of the *cagA* gene in *Helicobacter pylori* isolates from patients with different *H. pylori*-associated diseases. J. Clin. Microbiol. **36:**2258–2263.
33. **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.