JOURNAL OF CLINICAL MICROBIOLOGY, July 1999, p. 2306–2311 Vol. 37, No. 7
0095-1137/99/$04.00+0

Distinct Variants of *Helicobacter pylori cagA* Are Associated with *vacA* Subtypes

LEEN-JAN VAN DOORN,[1]* CÉU FIGUEIREDO,[1,2] RICARDO SANNA,[1]
MARTIN J. BLASER,[3] AND WIM G. V. QUINT[1]

Delft Diagnostic Laboratory, 2625 AD Delft, The Netherlands[1]; IPATIMUP and Faculty of Medicine, University of Porto, Porto, Portugal[2]; and Vanderbilt University, Division of Infectious Diseases, and Veterans Affairs Medical Center, Nashville, Tennessee[3]

Received 3 November 1998/Returned for modification 27 February 1999/Accepted 26 March 1999

The diversity of the cytotoxin-associated gene (*cagA*) of *Helicobacter pylori* was analyzed in 45 isolates obtained from nine countries. We examined variation in the 5′ end of the *cagA* open reading frame as determined by PCR and sequencing. Phylogenetic analysis revealed the existence of at least two distinct types of *cagA*. One variant (*cagA1*) was found exclusively in strains from Europe, the United States, and Australia, whereas a novel variant (*cagA2*) was found in strains from East Asia. The greatest diversity between *cagA1* and *cagA2* was found in the first 20 amino acids of the *cagA* open reading frame, where several consistent insertions or deletions were observed. Additional *cagA* sequence variants that could be classified as separate subtypes were found in two of three Peruvian and in five of seven U.S. strains tested. The calculated isoelectric point of the first 154 amino acids of the *cagA1* variants (7.52 ± 1.54) was significantly higher than that of the first 154 amino acids of the *cagA2* variants (5.61 ± 0.94; $P < 0.001$). Most *cagA2* strains contained *vacA* subtype s1c ($P < 0.001$), and in *vacA* m1 strains *cagA1* was more frequently observed than *cagA2*. These results show the epidemiological relationship between *cagA* and *vacA* at the subtype level and indicate the existence of distinct *H. pylori* lineages that are not uniformly distributed over the globe.

Helicobacter pylori is a medically important bacterium that is involved in the pathogenesis of peptic ulcer disease and that is associated with gastric carcinoma. The ecological niche of *H. pylori* is the human stomach, where it establishes a long-term colonization of the mucosa (8). During the past decade, products of several *H. pylori* genes that are markers for differences in clinical outcomes in colonized persons have been identified (7).

A cytotoxin that may damage epithelial cells by inducing the formation of vacuoles is encoded by *vacA* (22). Although *vacA* is present in all *H. pylori* strains, it contains at least two variable parts (4). The s region (which encodes the signal peptide) exists as s1 or s2 allelic types. Among type s1 strains, subtypes s1a, s1b, and s1c have been identified, and the m region (the middle region) occurs as m1, m2a, or m2b allelic types (43). The particular *vacA* s/m genotype is a marker of the pathogenicity of an individual strain, since in vitro production of the cytotoxin, in vivo epithelial damage, and development of peptic ulcer disease are all related to the *vacA* genotype (4, 5).

The cytotoxin-associated gene (*cagA*), which is not present in every *H. pylori* strain (11, 12, 37), is a marker for a genomic pathogenicity (*cag*) island of about 40 kbp (2, 35). The presence of this island is associated with more severe clinical outcomes (6, 21, 28) as well as with protection from esophageal diseases (10, 46). The function of CagA has not yet been determined. The *cag* island contains other genes that encode proteins that enhance the interaction of the strain with host cells, for example, by induction of cytokine production (2, 9, 38). There is a close association between the presence of *cagA* and *vacA* type s1, because most s1 strains are *cagA* positive (4).

The use of DNA fingerprinting techniques has revealed substantial genetic heterogeneity among different clinical isolates (1, 24, 41), and the total genetic variation within *H. pylori* is greater than that in other bacteria that have been studied (14). However, there is heterogeneity in the degree of variation; e.g., the variability of single-copy genes encoding enzymes may be restricted by functional constraints on the encoded proteins. Since *cagA* is an important marker for *H. pylori* strains that are highly interactive with the host, the present study aimed to investigate the existence of distinct *cagA* variants. Earlier studies indicated that the 5′ part of *cagA* is more conserved than the middle and 3′ end of the gene (11, 37). Since *cagA* is considered an important pathogenicity marker, the sequence heterogeneity of the 5′ end of *cagA* as well as the relationship with *vacA* variants was investigated among *H. pylori* strains of various geographic origins. We found that distinct *cagA* variants exist, with specific geographic distributions and particular associations with *vacA* genotypes.

MATERIALS AND METHODS

***H. pylori* isolates.** A total of 45 *H. pylori* isolates were obtained from different countries, including Australia ($n = 5$), China ($n = 4$), Hong Kong ($n = 9$), Japan ($n = 5$), The Netherlands ($n = 6$), Peru ($n = 3$), Portugal ($n = 3$), Thailand ($n = 3$), and the United States ($n = 7$). These were randomly selected from a large number of cultures representing cultures of isolates from each country. Each isolate was obtained from a different patient who underwent gastroscopy and gastric biopsy for dyspeptic symptoms. After primary isolation and identification of the gastric organisms as *H. pylori*, the strains were frozen at −70°C until their use in these studies. Subsequently, bacteria were cultured on Trypticase soy agar plates containing 5% sheep blood (Becton Dickinson) for 3 to 5 days at 37°C under microaerobic conditions (5% O_2, 10% CO_2, 85% N_2). The *H. pylori* cells were harvested from plates by suspension in 2 ml of sterile 0.9% NaCl solution and were pelleted by centrifugation at 10,000 × *g* for 2 min. The cells were resuspended in 400 μl of 10 mM Tris-HCl (pH 8.0), 5 mM EDTA, 0.1% sodium dodecyl sulfate, and 0.1 mg of proteinase K per ml and were incubated for 2 to 4 h at 55°C. Proteinase K was inactivated by incubation at 95°C for 10 min. The lysate was clarified by centrifugation at 14,000 × *g* for 2 min. The clarified supernatant was diluted 1/100 in sterile water and was directly used for PCR.

PCRs. All primers used in this study are shown in Table 1. Primer set cagP1-B1 and primer set cagF2-B1 were used to amplify the 5′ 450 bp of the *cagA* open reading frame (ORF). Primers cagAF and cagAR were used as a universal *cagA*

* Corresponding author. Mailing address: Delft Diagnostic Laboratory, R. de Graafweg 7, 2625 AD, Delft, The Netherlands. Phone: 31-15-2604577. Fax: 31-15-2604550. E-mail: L.J.van.Doorn@ddl.nl.

TABLE 1. PCR primers for amplification of *cagA* used in this study

Primer designation (polarity)	Primer sequence (5′→3′)	Position
cagP1 (+)	CCATTTTAAGCAACTCCATAAACC	552–575[a]
cagF1 (+)	TGGGTAAAAATGTGAATCGT	858–877[a]
cagF2 (+)	AAGATACCGATAGGTATGAA	1032–1051[a]
B1 (−)[b]	TCTGCCAAACAATCTTTTGCAG	1557–1578[a]
cagAF (+)	Bio[c]-TTGACCAACAACCACAAACCGAAG	1090–1113[a]
cagAR (−)	Bio-CTTCCCTTAATTGCGAGATTCC	1251–1272[a]
cagEF (+)[d]	GCGATTGTTATTGTGCTTGTAG	16891–16870[e]
cagER (−)[d]	GAAGTGGTTAAAAAATCAATGCCCC	16563–16587[e]
cagTF (+)[d]	CCATGTTTATACGCCTGTGT	442–461[e]
cagTR (−)[d]	CATCACCACACCCTTTTGAT	723–742[e]

[a] Positions corresponding to the *cagA* ORF of *H. pylori* ATCC 53726 (GenBank accession no. L11714).
[b] Tummuru et al. (37).
[c] Bio, biotin moiety at the 5′ end of the primer.
[d] Ikenoue et al. (17).
[e] Positions according to the sequence with GenBank accession no. U60176.

detection primer set. Separate PCRs were used to detect the presence of *cagE* and *cagT* in the *cag* pathogenicity island.

All PCR mixtures consisted of 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM deoxynucleotides, 25 pmol each of the forward and reverse primers, and 1.5 U of AmpliTaq Gold DNA polymerase (Perkin-Elmer) in a final volume of 50 μl. One microliter of DNA from a culture lysate was used in each PCR. The mixture was covered with mineral oil to prevent evaporation. The PCR program comprised 9 min of predenaturation at 94°C to activate the AmpliTaq Gold DNA polymerase, followed by 40 cycles of 30 s at 94°C, 45 s at 50°C, 45 s at 72°C, and a final incubation at 72°C for 5 min.

Sequence analysis of *cagA*. The PCR products, synthesized with either primers cagP1 and B1 or primers cagF2 and B1, were inspected by electrophoresis on 2% agarose gels. The PCR products were sequenced with the Thermo-Sequenase cycle sequencing kit (Amersham) by using Cy-labeled primers, followed by electrophoresis on an ALF-express automatic sequencer (Pharmacia Biotech). Sequences were analyzed with PC-Gene software (Intelligenetics Inc.) and ClustalX alignment software. Phylogenetic analyses were performed with Windows Easy Tree software (version 1.31) (12a). Pairwise sequence comparisons were made by using the Jukes and Cantor parameters, and matrices of sequence distances were produced. Phylogenetic relationships were further analyzed by the neighbor-joining method with 500 bootstrap steps, and the program Treecon (39) was used to create a graphic output.

Reverse hybridization LiPA. The presence of *cagA* and the subtypes of the *vacA* s and m regions were determined by multiplex PCR followed by reverse hybridization by a line probe assay (LiPA), as described earlier (32, 42). This assay consists of a nitrocellulose strip that contains poly(dT)-tailed oligonucleotide probes for *cagA* and the *vacA* s and m region genotypes (s1a, s1b, s1c, m1, m2a, and m2b) immobilized as parallel lines. Briefly, PCR products from *cagA* and the *vacA* s and m regions (containing biotin at the 5′ end of each primer) were mixed and denatured by addition of 400 mM NaOH and 10 mM EDTA in a plastic trough. Hybridization buffer and a LiPA strip were added, hybridization was performed, and the strip was stringently washed. Hybrids were detected by addition of conjugate (streptavidin-alkaline phosphatase) and substrate (4-nitroblue tetrazolium chloride and 5-bromo-4-chloro-3-indolylphosphate). Hybrids are visible as purple probe lines. Interpretation of the hybridization patterns was performed visually.

Statistical analysis. Data were analyzed by the χ^2 test or Fisher's exact test.

RESULTS

Nucleotide sequence analysis and distinct variants of *cagA*. The sequence heterogeneity of *cagA* was analyzed for 45 *H. pylori* isolates randomly selected from a larger panel of cultures obtained from nine different countries in Europe, North America, South America, and Asia by PCR and direct sequencing. Primer set cagP1 and B1 (1,026 bp) or primer set cagF2 and B1 (546 bp) was used to amplify the 5′ end of the *cagA* ORF. Neither of the two primer sets permitted amplification of *cagA* sequences for all 45 strains. However, with both primer sets, the sequences encoding the first 154 amino acids of the *cagA* ORF from all strains could be determined. *cagE* and *cagT* sequences also could be amplified from each of the 45 strains, suggesting that instances when the *cagA* pathogenicity island is not intact are uncommon.

Phylogenetic analyses of *cagA* sequences revealed the existence of several distinct groups (Fig. 1). The first group, designated *cagA1*, comprised 19 strains, mainly from Europe and Australia, and the reference sequence from U.S. strain ATCC 53726 (GenBank accession no. L11714), as well as a reference strain from Italy (GenBank accession no. X70039), which is virtually identical to ATCC 53726 in that region. The second group was designated *cagA2* and exclusively comprised 19 strains from East Asia. The patient from whom strain AU4 was obtained lived in Australia but also originated from East Asia.

In two of four Peruvian and in five of seven U.S. strains, variant *cagA* sequences that belonged to neither *cagA1* nor *cagA2* were found. On the basis of the distribution of molecular distances, as shown in the phylogenetic tree (Fig. 1), three possible additional *cagA* subtypes comprising sequences from Peru (*cagA3*) and the United States (*cagA4* and *cagA5*) could be distinguished. Formal classification of these sequences would require a larger number of strains from these parts of the world.

Analyses of the *cagA* sequences. The average G+C content was 36.3% and was not significantly different between the *cagA1* (37.3%) and *cagA2* (36.0%) variants. These values are similar to the 35% G+C content of the *cag* pathogenicity island and lower than the average 39% G+C content of the entire genome, as reported earlier (36).

On the basis of the alignment of the 45 novel and 1 reference *cagA* sequences, a set of general PCR primers (cagAF and cagAR) which permitted detection of *cagA* in all strains studied was designed. Strains that were negative with various other combinations of the *cagA* primer sets also remained negative with the novel cagAF and cagAR primer set (data not shown).

Pairwise comparisons of all sequences showed that nucleotide sequences were more than 95% conserved within these two groups (Table 2). Transitions were on average 2.2-fold more frequent than transversions. As expected, nonsynonymous substitutions (K_a) were less common than synonymous substitutions (K_s). That the mean synonymous substitution rate between *cagA1* and *cagA2* strains was greater than that within each group is consistent with the phylogenetic tree in Fig. 1, indicating that the deep branching represents the most fundamental dichotomy in the populations. That within each branch the synonymous substitution rates are similar suggests that the two branches are about the same age. Furthermore, the K_a/K_s ratio serves as an index of the functional constraints on the protein. The similarity of these ratios within *cagA1* and *cagA2* indicates that they are subject to similar constraints. That comparison between *cagA1* and *cagA2* yields a higher ratio indicates that to a certain degree the constraints are different.

Taking all 46 *cagA* sequences together, the nucleotide and amino acid sequences were 92.8 and 83.1% conserved, respectively (Table 2). The deduced amino acid sequences of *cagA1* and *cagA2* were more than 94% conserved within each group. Analysis of representative amino acid sequences of the major *cagA* variants confirmed that *cagA* does not have a cleaved signal sequence (Fig. 2). The ATG start codon was completely conserved in all 46 sequences studied. No conserved cysteine residues are present in this part of the CagA protein. A potential N-linked glycosylation site at Asn3 was conserved in 43 (93%) of the 46 sequences. Potential phosphorylation sites were found in several sequences, but none of these was conserved.

Compared to sequences from the *cagA1* group, *cagA2* se-

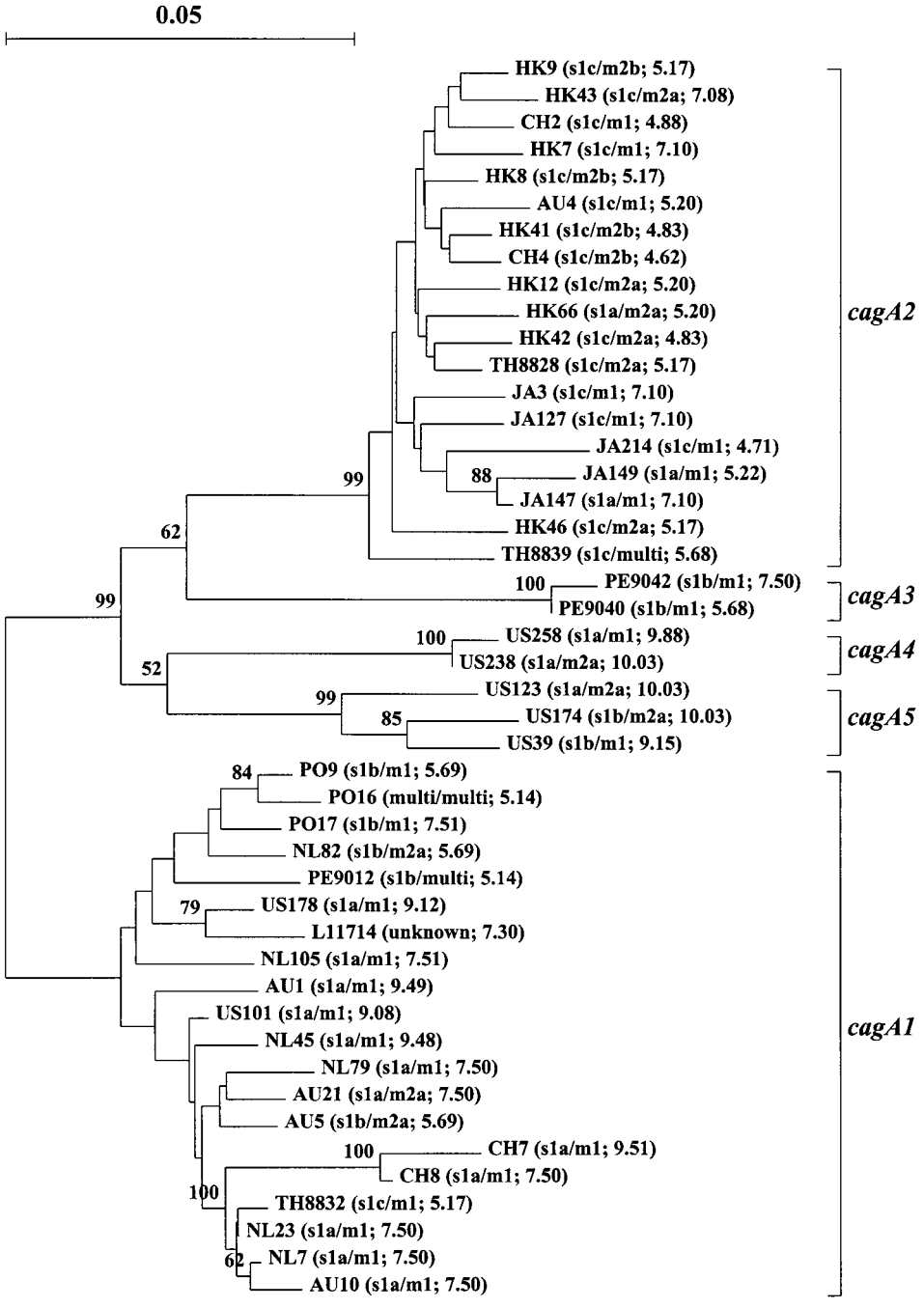

FIG. 1. Phylogenetic tree of *cagA* nucleotide sequences encoding the N-terminal 149 to 154 amino acids of the *cagA* ORF, including one reference sequence (GenBank accession no. L11714). Clusters of subtypes *cagA1* to *cagA5* are indicated, and bootstrap values greater than 50 are shown. Letters in the names of the isolates indicate the country of origin: Australia (AU), Costa Rica (CR), China (CH), Hong Kong (HK), Japan (JA), The Netherlands (NL), Peru (PE), Portugal (PO), Thailand (TH), and the United States (US). The *vacA* s and m subtypes and the isoelectric point of the deduced protein sequence are shown in parentheses after each sequence name (Multi indicates the presence of multiple *vacA* subtypes). A reference bar is shown for molecular distance.

quences contain an insertion of five amino acids (TXTPD in 13 of 19 strains) between residues 9 and 10. Four strains (JA127, JA147, JA214, and AU4) show a deletion of four amino acids (PQTQ in L11714) between residues 14 and 19. Thus, the first 20 amino acids of the *cagA* ORF are highly polymorphic. There were no consistent differences between the hydrophilicity/hydrophobicity patterns of the deduced CagA protein sequences. However, the calculated isoelectric points of the deduced 154 N-terminal amino acids of CagA ranged from 4.62 (CH4) to 10.03 (US174 and US123), as indicated in Fig. 1. The average isoelectric point of the sequences from the *cagA1* group (7.52 ± 1.54) was significantly higher ($P < 0.001$ by the *t* test) than that of the sequences from the *cagA2* group (5.61 ± 0.94). The deduced protein sequences from the seven strains of subtypes *cagA3* to *cagA5* had even higher isoelectric points (8.88 ± 1.68).

TABLE 2. Phylogenetic analysis of 46 *cagA* sequences[a]

Sequence compared	Distance[b]	Homology[c]	Transitions (T_s)[d]	Transversions (T_v)[e]	T_s:T_v	K_s[f]	K_a[g]	K_a/K_s ratio
cagA1 (n = 20)	0.032 ± 0.012	96.2 ± 2.5 (94.2 ± 2.3)	0.024 ± 0.009	0.008 ± 0.006	3.0	0.060 ± 0.024	0.023 ± 0.011	0.38
cagA2 (n = 19)	0.027 ± 0.011	95.9 ± 1.8 (94.8 ± 2.7)	0.020 ± 0.006	0.007 ± 0.003	2.9	0.064 ± 0.026	0.018 ± 0.007	0.28
cagA1 + *cagA2* (n = 39)	0.068 ± 0.039	93.6 ± 3.5 (90.5 ± 5.6)	0.045 ± 0.023	0.024 ± 0.016	1.9	0.110 ± 0.052	0.053 ± 0.033	0.48
All *cagA* sequences (n = 46)	0.077 ± 0.037	92.8 ± 3.3 (83.1 ± 5.5)	0.053 ± 0.025	0.025 ± 0.014	2.1	0.121 ± 0.051	0.060 ± 0.032	0.50

[a] Based on the 5′ 462 bp from the *cagA* ORF in each of the 46 *H. pylori* strains.
[b] Average substitution rates (corrected with the Jukes and Cantor parameters) ± standard deviation.
[c] Average percentage similarity ± standard deviation based on pairwise comparisons of sequences of the corresponding subtypes. Homologies are based on nucleotide sequences; values in parentheses are based on deduced amino acid sequences.
[d] Transitions are nucleotide replacements of one purine by another purine.
[e] Transversions are nucleotide replacements of a purine by a pyrimidine or vice versa.
[f] Synonymous substitutions do not result in an amino acid change in the encoded open reading frame.
[g] Nonsynonymous substitutions result in an amino acid change.

Association of *cagA* variants with *vacA* types or subtypes. All 45 strains were analyzed for subtypes of *cagA* (*cagA1* to *cagA5*), the *vacA* s region (s1a, s1b, s1c, s2), and the *vacA* m region (m1, m2a, m2b) by PCR-LiPA (Table 3). Multiple *vacA* genotypes were observed in three strains (PE9012, PO16, and TH8839) and were excluded from the analysis. Of the 17 *cagA1* strains with a single *vacA* genotype, 16 (94%) were s1a or s1b and only 1 (6%) was s1c. Conversely, of the 18 *cagA2* strains with a single *vacA* genotype, 15 (83%) were s1c, whereas only 3 (17%) were s1a or s1b. This relationship between the *vacA* s-region subtype and the *cagA* subtype was highly significant ($P < 0.001$). Of the 17 *cagA1* strains, 14 were m1 and 3 were m2. Conversely, of the 18 *cagA2* strains, 8 were m1 and 10 were m2 ($P = 0.035$). The numbers of strains of the *cagA3* to *cagA5* genotypes were too small for association analysis.

DISCUSSION

Knowledge of the existence of different *H. pylori* genotypes may become clinically important. In Western populations, strains that contain *cagA* (indicating the presence of the pathogenicity island) are more strongly associated with more severe disease than strains that lack *cagA* (7, 12, 30, 34). Similarly, type s1 *vacA* strains are more often associated with disease than type s2 strains (3, 5), and responses to anti-*Helicobacter* therapy also may vary for strains of different genotypes (45). Adhesion of *H. pylori* to human gastric epithelium, mediated by histo-blood group antigens, also appears to be related to distinct allelic variants of the bacterial *babA* gene (18). In contrast, in Asian populations the association of *cagA* positivity and disease risk is much weaker or not present (13, 26, 27, 47). The existence of distinct variants of certain genes of the *cag* pathogenicity island may offer an explanation for this discrepancy. Recently, analysis of *H. pylori* strains from East Asian patients suggested that in some strains, only part of the *cag* pathogenicity island is present (17, 23). Some strains contained *cagA* but appeared to lack *cagE* and/or *cagT*, which was located in other parts of the pathogenicity island. That each of the 45 strains analyzed in the present study contained *cagA* and *cagE*, as well as *cagT*, suggests the presence of a complete pathogenicity island in these strains. Polymorphisms may have been responsible for failure to detect these other *cag* island genes in the previous studies (17).

The present study extended previous analyses of *cagA* heterogeneity (25, 40) and examined the relationship with *vacA* subtypes. The overall genetic heterogeneity of *H. pylori* is very high (14), and several studies reveal extreme variability, even within a single patient (19, 35, 41). The existence of distinct *cagA* variants that appear to be highly conserved (>95% at the nucleotide level and >94% at the amino acid level) confirms the observations of van der Ende et al. (40), who analyzed a

```
                    *        20         *        40         *        60         *        80
L11714 (1)  : MTNETIDQ-----QPQTEAAFNPQQFINNLQVAFLKVDNAVASYDPDQKPIVDKNDRDNRQAFEGISQLREEYSNKAIKN :  75
AU21 (1)    : ..........-.....................................S..................DR.............. :  75
CH7 (1)     : ....P.T.------..R..V.....................T.....................D..........R..... :  75
AU4 (2)     : ........TTTPD.----TD.V..R.....................................K........A..... :  76
JA147 (2)   : ........TTTPDR----TD.V..R.....................................K........A..... :  76
PE9040 (3)  : .A......TEKPD.TPN.T.....................S.....................D.........A..... :  80
US238 (4)   : ........TRTPD.T...T.....R.............S.....................K........A..... :  80
US123 (5)   : ........-----.....................V.....................K........A..... :  75
US39 (5)    : ........TRTPD.T...T.....................................K........A..... :  80

                    *       100         *       120         *       140         *
L11714 (1)  : PTKKNQYFSDFINKSNDLINKDNLIVVESSTKSFQKFGDQRYRIFTSWVSHQNDPSKINTRCIRNFMEHTIQPP : 149
AU21 (1)    : ........................D.....................................S......NI.... : 149
CH7 (1)     : ......................A..E.................Q.................S......NI.... : 149
AU4 (2)     : .A.....................R.A.D..VD..........Q.......L.K.......QQT.....NI.... : 150
JA147 (2)   : ........................A.D..VE..R.......Q.......L.K.......QQ......NI.... : 150
PE9040 (3)  : ........................DTD..........T...Q..MN.....K.......QK......NI.... : 154
US238 (4)   : ............S.........G..DTG..I.....................................S..... : 154
US123 (5)   : .A.......S..S..S......S..DTG..I.......T...Q..MN.....K.......QK..G...NI.... : 149
US39 (5)    : .A.......S..S..S......S..DTG..I.......T...Q..MN.....K.......QK.Q....NI.... : 154
```

FIG. 2. Alignment of representative N-terminal amino acid sequences (residue 1 to residues 149 to 154) of the distinct *cagA* variants, as deduced from nucleotide sequences. To obtain proper alignment, a hyphen indicates the absence of an amino acid residue. Identical amino acids are indicated by a dot. The deduced amino acid sequence from GenBank accession no. L11714 is also shown as a reference. The numbers in parentheses indicate the *cagA* subgroup.

TABLE 3. Relationship between *cagA* and *vacA* subtypes in 45 *H. pylori* strains

cagA genotype	No. of strains with the following *vacA* genotype:							
	s1a		s1b		s1c		Multiples[a]	Total
	m1	m2	m1	m2	m1	m2		
cagA1[b]	11	1	2	2	1	0	2	19
cagA2[c]	2	1	0	0	6	9	1	19
cagA3 to *cagA5*	1	2	3	1	0	0	0	7
Total	14	4	5	3	7	9	3	45

[a] Strains containing multiple *vacA* subtypes (PE9012, PO16, and TH8839).
[b] The presence of *cagA1* is associated with *vacA* type s1a or s1b ($P < 0.001$) and type m1 ($P = 0.035$).
[c] The presence of *cagA2* is associated with *vacA* type s1c ($P < 0.001$) and type m2 ($P = 0.035$).

more downstream portion of *cagA*. That the 5′ end of the *cagA* ORF was in frame in all 45 strains is consistent with a functional CagA protein. The nucleotide substitution rates (transitions to transversions and nonsynonymous to synonymous substitution rates) observed in *cagA* also are consistent with functional constraints at the protein level, implying that the *cagA* gene product has an in vivo role. The K_a/K_s ratio for *cagA* was much higher than the average value of 0.04 for *Escherichia coli* and *Salmonella* genes (31). One explanation could be the strong immunogenicity of the CagA protein, since most patients colonized by a *cagA*-positive *H. pylori* strain show high titers of anti-CagA antibodies (12, 29). A rabbit antiserum raised to a single CagA protein permitted detection of *cagA*-positive strains from several geographic regions (15), but whether variant-specific antibodies to CagA exist remains unknown. However, the efficacy of detection of anti-CagA antibodies by a single CagA variant should be evaluated separately with strains from patients from different geographic areas.

That the substitution rates among *cagA1* and *cagA2* were very similar suggests that the evolutionary ages for these variants are similar. On the basis of phylogenetic analyses, the number of recombinants between the distinct *cagA* subtypes appears to be limited. The heterogeneity of the single-copy genes that encode enzymes is usually restricted due to functional constraints on the encoded proteins, and substitutions tend to be synonymous. Virulence-associated genes may be subjected to selection for differences in biological specificity, resulting in phenotypically distinct variants distinguishable at the genetic level. Thus, the 5′ heterogeneity of *cagA* may be related to differences in biological activity of the CagA protein.

Although only the 154 N-terminal amino acids of the *cagA* ORF were analyzed, there were considerable differences in the isoelectric points of the deduced peptides. *cagA* transcription can be increased after exposure of *H. pylori* cells to acid pH, and CagA surface exposure also increased (20). Since *H. pylori* colonizes within the acidic environment of the gastric mucus layer, the isoelectric point may be important, especially if the protein is secreted or surface exposed, as CagA is. Differences in the CagA isoelectric points also may reflect particular adaptations to gastric acidity present in host populations.

Although several sets of PCR primers for *cagA* amplification have been described (13, 16, 37, 48), the utility of specific primers should be evaluated for each geographic area. The new *cagA* primers that we designed were deduced from the alignment of 46 sequences and permitted efficient amplification from all tested strains, indicating their broad applicability for analysis of *H. pylori* isolates of various geographic origins.

Analysis of *H. pylori* isolates of diverse geographic origins permitted a comprehensive description of the *cagA* and *vacA* variants. We confirmed the association between the presence of *cagA* and *vacA* s1 (44) and now show that there is also a highly significant association between the distinct *cagA* and *vacA* s1 subtypes. The different *vacA* s1-region subtypes are related to the level of cytotoxin production (4), but the evolutionary and functional relationships between *vacA* and *cagA* have not been elucidated. That the association between *cagA* and *vacA* variants was strong, but not perfect, is consistent with the recombinatorial genetic population structure of *H. pylori* (33). The observed relationship between variants of *cagA* and *vacA* provides further evidence for the existence of distinct clonal groupings of *H. pylori*. These data appear to be in contrast to those from studies of other genetic loci that showed that *H. pylori* does not have a strong clonal structure (14) and that recombination is frequent (33). That there are distinct East Asian and Western branches could indicate that *cagA* has been present in *H. pylori* for at least tens of thousands of years, before human populations segregated in Europe and Asia. On the basis of earlier studies we speculate that recombination at the *cagA* locus has been ongoing within each isolated geographic pool of strains. Further subtyping of *H. pylori* strains should facilitate insights into the evolution, epidemiology, and clinical significance of particular variants.

ACKNOWLEDGMENTS

C. Figueiredo is supported by PRAXIS XXI. This work was supported in part by R01 DK 53707 from the National Institutes of Health.

REFERENCES

1. **Akopyants, N. S., T. U. Bukanov, T. U. Westblom, S. Kresovich, and D. E. Berg.** 1992. DNA diversity among clinical isolates of *Helicobacter pylori* detected by PCR-based RAPD profile. Nucleic Acids Res. **20:**S137–S142.
2. **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:**37–54.
3. **Atherton, J. C.** 1997. The clinical relevance of strain types of *Helicobacter pylori*. Gut **40:**701–703.
4. **Atherton, J. C., P. Cao, R. M. J. 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. **Atherton, J. C., R. M. J. Peek, K. T. Tham, T. L. Cover, and M. J. Blaser.** 1997. Clinical and pathological importance of heterogeneity in *vacA*, the vacuolating cytotoxin gene of *Helicobacter pylori*. Gastroenterology **112:**92–99.
6. **Beales, I. L., J. E. Crabtree, D. Scunes, A. Covacci, and J. Calam.** 1996. Antibodies to *cagA* protein are associated with gastric atrophy in *Helicobacter pylori* infection. Eur. J. Gastroenterol. Hepatol. **8:**645–649.
7. **Blaser, M. J.** 1995. Intrastrain differences in *Helicobacter pylori*: a key question in mucosal damage? Ann. Med. **27:**559–563.
8. **Blaser, M. J.** 1997. Ecology of *Helicobacter pylori* in the human stomach. J. Clin. Invest. **100:**759–762.
9. **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.
10. **Chow, W. H., M. J. Blaser, W. J. Blot, M. D. Gammon, T. L. Vaughan, H. A. Risch, G. I. Perez-Perez, J. B. Schoenberg, J. L. Stanford, H. Rotterdam, et al.** 1998. An inverse relation between *cagA*+ strains of *Helicobacter pylori* infection and risk of esophageal and gastric cardia adenocarcinoma. Cancer Res. **58:**588–590.
11. **Covacci, A., S. Censini, M. Bugnoli, R. Petracca, D. Burroni, G. Macchia, A. Massone, E. Papini, Z. Xiang, and N. Figura.** 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.
12. **Cover, T. L., Y. Glupczynski, A. P. Lage, A. Burette, M. K. Tummuru, G. I.**

Perez-Perez, and M. J. Blaser. 1995. Serologic detection of infection with *cagA*$^+$ *Helicobacter pylori* strains. J. Clin. Microbiol. **33:**1496–1500.
12a. **Dopazo, J.** 1997. Windows Easy Tree software, version 1.31. [Online.] http://www.tdi.es. [30 April 1999, last date accessed.]
13. **Go, M. F., and D. Y. Graham.** 1996. Presence of the *cagA* gene in the majority of *Helicobacter pylori* strains is independent of whether the individual has duodenal ulcer or asymptomatic gastritis. Helicobacter **1:**107–111.
14. **Go, M. F., V. Kapur, D. Y. Graham, and J. M. Musser.** 1996. Population genetic analysis of *Helicobacter pylori* by multilocus enzyme electrophoresis: extensive allelic diversity and recombinational population structure. J. Bacteriol. **178:**3934–3938.
15. **Hook-Nikanne, J., G. I. Perez-Perez, and M. J. Blaser.** 1997. Antigenic characterization of *Helicobacter pylori* strains from different parts of the world. Clin. Diagn. Lab. Immunol. **4:**592–597.
16. **Husson, M. O., F. Gottrand, A. Vachee, L. Dhaenens, E. M. De La Salla, D. Turclk, and M. L. H. Houcke.** 1995. Importance in diagnosis of gastritis of detection by PCR of the *cagA* gene in *Helicobacter pylori* strains isolated from children. J. Clin. Microbiol. **33:**3300–3303.
17. **Ikenoue, T., S. Maeda, K. Ogura, H. Yoshida, F. Kanai, Y. Shiratori, and M. Omata.** 1998. Simple and practical methods to study *cag* pathogenicity island: usage of *cagE* and *cagT* gene amplifications. Gastroenterology **114:** A157 (abstr. G0641).
18. **Ilver, D., A. Arnqvist, J. Ogren, I. M. Frick, D. Kersulyte, E. T. Incecik, D. E. Berg, A. Covacci, L. Engstrand, and T. Boren.** 1998. *Helicobacter pylori* adhesin binding fucosylated histo-blood group antigens revealed by retagging. Science **279:**373–377.
19. **Jorgensen, M., G. Daskalopoulos, V. Warburton, H. M. Mitchell, and S. L. Hazell.** 1996. Multiple strain colonization and metronidazole resistance in *Helicobacter pylori*-infected patients: identification from sequential and multiple biopsy specimens. J. Infect. Dis. **174:**631–635.
20. **Karita, M., M. K. Tummuru, H. P. Wirth, and M. J. Blaser.** 1996. Effect of growth phase and acid shock on *Helicobacter pylori cagA* expression. Infect. Immun. **64:**4501–4507.
21. **Kuipers, E. J., G. I. Perez-Perez, S. G. Meuwissen, and M. J. Blaser.** 1995. *Helicobacter pylori* and atrophic gastritis: importance of the *cagA* status. J. Natl. Cancer Inst. **87:**1777–1780.
22. **Leunk, R. D., P. T. Johnson, B. S. David, W. G. Kraft, and D. R. Morgan.** 1988. Cytotoxic activity in broth culture filtrates of *Campylobacter pylori*. J. Med. Microbiol. **26:**93–99.
23. **Maeda, S., T. Ikenoue, K. Ogura, H. Yoshida, F. Kanai, N. Kato, Y. Shiratori, and M. Omata.** Structure of *cag* pathogenicity island in Japanese and Western *Helicobacter pylori* isolates. Gastroenterology **114:**A210 (abstr. G0861).
24. **Marshall, D. G., D. C. Coleman, D. J. Sullivan, H. Xia, C. A. O'Morain, and C. J. Smyth.** 1996. Genomic DNA fingerprinting of clinical isolates of *Helicobacter pylori* using short oligonucleotide probes containing repetitive sequences. J. Appl. Bacteriol. **81:**509–517.
25. **Miehlke, S., K. Kibler, J. G. Kim, N. Figura, S. M. Small, D. Y. Graham, and M. F. Go.** 1996. Allelic variation in the *cagA* gene of *Helicobacter pylori* obtained from Korea compared to the United States. Am. J. Gastroenterol. **91:**1322–1325.
26. **Ogura, K., F. Kanai, S. Maeda, H. Yoshida, M. Ogura, K. H. Lan, K. Hirota, T. Kawabe, Y. Shiratori, and M. Omata.** 1997. High prevalence of cytotoxin positive *Helicobacter pylori* in patients unrelated to the presence of peptic ulcers in Japan. Gut **41:**463–468.
27. **Pan, Z. J., R. W. van der Hulst, M. Feller, S. D. Xiao, G. N. Tytgat, J. Dankert, and A. van der Ende.** 1997. Equally high prevalences of infection with *cagA*-positive *Helicobacter pylori* in Chinese patients with peptic ulcer disease and those with chronic gastritis-associated dyspepsia. J. Clin. Microbiol. **35:**1344–1347.
28. **Peek, R. M., G. G. Miller, K. T. Tham, G. Perez-Perez, X. Zhao, J. C. Atherton, and M. J. Blaser.** 1995. Heightened inflammatory response and cytokine expression in vivo to *cagA*+ *Helicobacter pylori* strains. Lab. Invest. **73:**760–770.
29. **Perez-Perez, G. I., N. Bhat, J. Gaensbauer, A. Fraser, D. N. Taylor, E. J. Kuipers, L. Zhang, W. C. You, and M. J. Blaser.** 1997. Country-specific constancy by age in *cagA*+ proportion of *Helicobacter pylori* infections. Int. J. Cancer **72:**453–456.
30. **Rudi, J., C. Kolb, M. Maiwald, I. Zuna, A. von Herbay, P. R. Galle, and W. Stremmel.** 1997. Serum antibodies against *Helicobacter pylori* proteins VacA and CagA are associated with increased risk for gastric adenocarcinoma. Dig. Dis. Sci. **42:**1652–1659.
31. **Sharp, P. M.** 1991. Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*: codon usage, map position, and concerted evolution. J. Mol. Evol. **33:**23–33.
32. **Stuyver, L., R. Rossau, A. Wyseur, M. Duhamel, B. Vandenborght, H. van Heuverswyn, and G. Maertens.** 1993. Typing of HCV isolates and characterization of new (sub)types using a line probe assay. J. Gen. Virol. **74:**1093–1102.
33. **Suerbaum, S., J. M. Smith, K. Bapumia, G. Morelli, N. H. Smith, E. Kunstmann, I. Dyrek, and M. Achtman.** 1998. Free recombination within *Helicobacter pylori*. Proc. Natl. Acad. Sci. USA **95:**12619–12624.
34. **Takata, T., S. Fujimoto, K. Anzai, T. Shirotani, M. Okada, Y. Sawae, and J. Ono.** 1998. Analysis of the expression of *cagA* and *vacA* and the vacuolating activity in 167 *Helicobacter pylori* isolates from patients with either peptic ulcers or non-ulcer dyspepsia. Am. J. Gastroenterol. **93:**30–34.
35. **Taylor, N. S., J. G. Fox, N. S. Akopyants, D. E. Berg, N. Thompson, B. Shames, L. Yan, E. Fontham, F. Janney, F. M. Hunter, and P. Correa.** 1995. Long-term colonization with single and multiple strains of *Helicobacter pylori* assessed by DNA fingerprinting. J. Clin. Microbiol. **33:**918–923.
36. **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. Wattey, 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.
37. **Tummuru, M. K., T. L. Cover, and M. J. Blaser.** 1993. Cloning and expression of a high-molecular-mass major antigen of *Helicobacter pylori*: evidence of linkage to cytotoxin production. Infect. Immun. **61:**1799–1809.
38. **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.
39. **Van de Peer, Y., and R. De Wachter.** 1994. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comput. Appl. Biosci. **10:**569–570.
40. **van der Ende, A., Z. J. Pan, A. Bart, R. W. van der Hulst, M. Feller, S. D. Xiao, G. N. J. Tytgat, and J. Dankert.** 1998. *cagA*-positive *Helicobacter pylori* populations in China and The Netherlands are distinct. Infect. Immun. **66:**1822–1826.
41. **van der Ende, A., E. A. Rauws, M. Feller, C. J. Mulder, G. N. Tytgat, and J. Dankert.** 1996. Heterogeneous *Helicobacter pylori* isolates from members of a family with a history of peptic ulcer disease. Gastroenterology **111:**638–647.
42. **van Doorn, L. J., C. Figueiredo, R. Rossau, G. Jannes, M. van Asbroeck, J. C. Sousa, F. Carneiro, and W. Quint.** 1998. Typing of the *Helicobacter pylori vacA* gene and detection of the *cagA* gene by PCR and reverse hybridization. J. Clin. Microbiol. **36:**1271–1276.
43. **van Doorn, L. J., C. Figueiredo, R. Sanna, A. S. Pena, P. Midolo, E. K. Ng, J. C. Atherton, M. J. Blaser, and W. Quint.** 1998. Expanding allelic diversity of *Helicobacter pylori vacA*. J. Clin. Microbiol. **36:**2597–2603.
44. **van Doorn, L. J., C. Figueiredo, R. Sanna, A. Plaisier, P. Schneeberger, W. A. de Boer, and W. Quint.** 1998. Clinical relevance of the *cagA*, *vacA*, and *iceA* status of *Helicobacter pylori*. Gastroenterology **115:**58–66.
45. **van Doorn, L. J., W. G. V. Quint, P. M. Schneeberger, G. N. J. Tytgat, and W. A. de Boer.** 1997. Association between *vacA* and *cagA* status of *Helicobacter pylori* and the efficacy of a 1-day quadruple therapy. Lancet **350:**71–72.
46. **Vicari, J. J., R. M. Peek, G. W. Falk, J. R. Goldblum, K. A. Easley, J. Schnell, G. I. Perez-Perez, S. A. Halter, T. W. Rice, M. J. Blaser, and J. E. Richter.** 1998. The seroprevalence of *cagA*-positive *Helicobacter pylori* strains in the spectrum of gastroesophageal reflux disease. Gastroenterology **115:**50–57.
47. **Webb, P. M., M. C. Yu, D. Forman, B. Henderson, D. G. Newell, J. M. Yuan, Y. T. Gao, and R. K. Ross.** 1996. An apparent lack of association between *Helicobacter pylori* infection and risk of gastric cancer in China. Int. J. Cancer **67:**603–607.
48. **Yamaoka, Y., M. Kita, T. Kodama, N. Sawai, and J. Imanishi.** 1996. Helicobacter pylori *cagA* gene and expression of cytokine messenger RNA in gastric mucosa. Gastroenterology **110:**1744–1752.