Expanding Allelic Diversity of *Helicobacter pylori vacA*

LEEN-JAN VAN DOORN,[1]* CÉU FIGUEIREDO,[1,2] RICARDO SANNA,[1] SALVADOR PENA,[3]
PETER MIDOLO,[4] ENDERS K. W. NG,[5] JOHN C. ATHERTON,[6]
MARTIN J. BLASER,[7] AND WIM G. V. QUINT[1]

*Delft Diagnostic Laboratory, 2625 AD Delft,[1] and Free University Hospital, Department of Gastroenterology,
Amsterdam,[3] The Netherlands; IPATIMUP and Medical Faculty, University of Porto, Porto, Portugal[2];
Monash Medical Center, Victoria, Australia[4]; Prince of Wales Hospital, Hong Kong, China[5]; Division of
Infectious Diseases, Vanderbilt University, Nashville, Tennessee[7]; and Division of Gastroenterology
and Institute of Infections and Immunity, University Hospital, Nottingham, United Kingdom[6]*

The diversity of the gene encoding the vacuolating cytotoxin (*vacA*) of *Helicobacter pylori* was analyzed in 98 isolates obtained from different geographic locations. The studies focused on variation in the previously defined s and m regions of *vacA*, as determined by PCR and direct sequencing. Phylogenetic analysis revealed the existence of four distinct types of s-region alleles: aside from the previously described s1a, s1b, and s2 allelic types, a novel subtype, designated s1c, was found. Subtype s1c was observed exclusively in isolates from East Asia and appears to be the major s1 allele in that part of the world. Three different allelic forms (m1, m2a, and m2b) were detected in the m region. On the basis of sequence alignments, universal PCR primers that allow effective amplification of the s and m regions from *H. pylori* isolates from all over the world were defined. Amplimers were subsequently analyzed by reverse hybridization onto a line probe assay (LiPA) that allows the simultaneous and highly specific hybridization of the different *vacA* s- and m-region alleles and tests for the presence of the cytotoxin-associated gene (*cagA*). This PCR-LiPA method permits rapid analysis of the *vacA* and *cagA* status of *H. pylori* strains for clinical and epidemiological studies and will facilitate identification of any further variations.

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 (9). The use of DNA fingerprinting techniques has revealed substantial genetic heterogeneity among different isolates (1, 25, 40), and the total variation of *H. pylori* is greater than that in other bacteria that have been studied (18). During the past decade, the products of several *H. pylori* genes that play a role in disease have been identified (7, 8, 43). The cytotoxin-associated gene (*cagA*), which is not present in every *H. pylori* strain (12, 15, 37), is a marker for a genomic pathogenicity (*cag*) island of about 40 kbp (11) whose presence is associated with more severe clinical outcomes (6, 22, 27). The *cag* island contains genes encoding proteins that enhance the virulence of the strain, for example, by inducing cytokine production by the host (2, 11, 38).

A cytotoxin that damages epithelial cells by inducing the formation of vacuoles is encoded by *vacA* (10, 13, 14, 16, 23, 28, 30, 36). Although *vacA* is present in all *H. pylori* strains, it contains at least two variable parts (4). The s region (encoding the signal peptide) exists as s1 or s2 allelic types. Among type s1 strains, subtypes s1a and s1b have been identified. The m (middle) region occurs as m1 or m2 allelic types. There is a close association between the presence of *cagA* and *vacA* type s1, because most s1 strains are *cagA* positive (4). The particular *vacA* s and m genotype is a marker of the pathogenicity of an individual strain, since in vitro production of the cytotoxin, in vivo epithelial damage, and peptic ulcer disease are all related to the *vacA* genotype (4, 5).

Analysis of total genomic DNA or polymorphic regions usually shows extremely large variations among isolates. However, different DNA segments of the bacterial genome are not uniform in their variability. The heterogeneity of single-copy genes encoding enzymes is usually restricted due to functional constraints on the encoded proteins, and mutations tend to be synonymous. Virulence-associated genes may be subjected to the selection of variants with altered biological specificity. This may lead to phenotypically distinct variants which can be distinguished at the genetic level. In the present study, the variability of the s and m regions of the cytotoxin-encoding *vacA* gene was investigated among *H. pylori* strains from various geographic locations. On the basis of sequence data, a PCR-based assay for analysis of *vacA* allelic variation that permits rapid genotyping of *H. pylori* strains from diverse parts of the world was developed.

MATERIALS AND METHODS

***H. pylori* isolates.** A total of 98 *H. pylori* isolates were obtained from different locations around the world, including Australia ($n = 6$), Costa Rica ($n = 4$), China ($n = 6$), Hong Kong ($n = 21$), Japan ($n = 5$), The Netherlands ($n = 12$), Peru ($n = 6$), Portugal ($n = 25$), Thailand ($n = 5$), the United States ($n = 6$), and New Zealand (where strains were obtained from Maoris, who are native inhabitants of New Zealand of Polynesian origin; $n = 2$). Each isolate was obtained from a different patient who underwent gastroscopy and gastric biopsy, usually due to symptoms of dyspepsia. After the 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, the 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, 80% N_2). The *H. pylori* cells were harvested from the plates by suspension in 2 ml of a sterile 0.9% NaCl solution and were pelleted by centrifugation at $10,000 \times g$ 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 (SDS)–0.1 mg of proteinase K per ml and were incubated

* 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 *vacA* and *cagA* used in this study

Gene and segment	Primer designation (polarity)	Primer sequence (5′→3′)[a]	Positions
vacA S	VA1F (+)[b]	Bio-ATGGAAATACAACAAACACAC	1–21[c]
vacA S	VA1XR (−)	Bio-CCTGARACCGTTCCTACAGC	157–176[c] 184–203[d]
vacA M	HPMGF (+)	CACAGCCACTTTCAATAACGA	1443–1463[c] 1419–1439[d]
vacA M	HPMGR (−)	CGTCAAAATAATTCCAAGGG	1824–1843[c] 1875–1894[d]
vacA M	MF1.1 (+)	Bio-GTGGATGCCCATACGGCTAA	1495–1514[c] 1470–1490[d]
vacA M	MF1.2 (+)	Bio-GTGGATGCTCATACAGCTWA	1495–1514[c] 1470–1490[d]
vacA M	MF1.3 (+)	Bio-GTGGATGCCCATACGATCAA	1495–1514[c] 1470–1490[d]
vacA M	MF1.4 (+)	Bio-GCGAGCGCTCATACGGTCAA	1495–1514[c] 1470–1490[d]
vacA M	MR1 (−)	Bio-RTGAGCTTGTTGATATTGAC	1582–1601[c] 1633–1652[d]
cagA	cagAF (+)	Bio-TTGACCAACAACCACAAACCG AAG	17–40[e]
cagA	cagAR (−)	Bio-CTTCCCTTAATTGCGAGATTCC	178–199[e]

[a] R = G or A, W = A or C, Bio = biotin at the 5′ end of the primer.
[b] The primer was from Atherton et al. (4).
[c] Positions corresponding to the *vacA* s1a-m1 open reading frame of *H. pylori* 60190 (GenBank accession no. U05676).
[d] Positions corresponding to the *vacA* s2-m2a open reading frame of *H. pylori* Tx30a (GenBank accession no. U29401).
[e] Positions corresponding to the *cagA* open reading frame of *H. pylori* ATCC 53726 (GenBank accession no. L11714).

for 2 to 4 h at 55°C. Proteinase K was inactivated by incubation at 95°C for 10 min. The lysate was diluted 1/100 in sterile water and was directly used for PCR.

PCRs. All primers used in this study are presented in Table 1. To amplify the *vacA* s region, primers VA1F and VA1xR were used, resulting in a fragment of 176 bp for type s1 variants and a fragment of 203 bp for type s2 variants. Primers HPMGF and HPMGR were used to amplify part of the *vacA* m region and yielded fragments of 401 and 476 bp for type m1 and type m2, respectively. These fragments were used for sequence analysis. Subsequently, on the basis of these m region sequences, novel primers that amplify a smaller part of the m region and that have higher sensitivities than HPMGF and HPMGR were designed. This new m-region PCR comprised a mixture of forward primers MF1.1, MF1.2, MF1.3, and MF1.4 and reverse primer MR1 and resulted in the amplification of fragments of 107 and 182 bp for m1 and m2 strains, respectively.

The presence of *cagA* was detected with primers cagAF and cagAR, yielding a product of 183 bp. All PCR mixtures consisted of 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl$_2$, deoxynucleotides at concentrations of 200 μM each, 25 pmol 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 the culture lysates was used in each reaction mixture. The mixture was covered with mineral oil to prevent evaporation. The PCR program comprised 9 min of predenaturation at 94°C, followed by 40 cycles of 30 s at 94°C, 45 s at 50°C, and 45 s at 72°C and a final incubation at 72°C for 5 min. PCR products were inspected by electrophoresis on 2% agarose gels.

Sequence analysis. The PCR products were sequenced with the Thermo-Sequenase cycle sequencing kit (Amersham), by using Cy-labelled primers, followed by electrophoresis on an ALF-express automatic sequencer (Pharmacia Biotech). Sequences were analyzed with PCGene software (Intelligenetics Inc.) and the Clustal program. Phylogenetic analyses were performed with various modules of the Phylogeny Inference Package (Phylip), version 3.5c. (17). Pairwise sequence comparisons were performed with the program DNADIST by using the Kimura two-parameter setting, and matrices of sequence distances were produced; phylogenetic relationships were further analyzed by the NEIGH-

BOR-JOINING method. The program Treecon (39) was used to create a graphic output. Substitution rates were calculated with Windows Easy Tree software (version 1.31; developed by J. Dopazo, 1997 [available at http://www.tdi.es]). Classification of m-region sequences into types and subtypes was based on the frequency distribution of nucleotide sequence distances, as calculated from the distance matrices, as has been described for the classification of hepatitis C virus types and subtypes (31). Three distinct levels of sequence variability are recognized. The major groupings of sequence variants (m1 or m2) are designated types, whereas the more closely related groups are termed subtypes (m2a or m2b). Sequences belonging to the same subtype show even more limited heterogeneity and are designated isolates.

Reverse hybridization LiPA. PCR products were analyzed by reverse hybridization by a line probe assay (LiPA) (32). This assay consists of a nitrocellulose strip that contains dT-tailed oligonucleotide probes (Table 2) immobilized as parallel lines. For each strain, 10 μl of each PCR product (containing biotin at the 5′ end of each primer) was denatured by the addition of an equal amount of 400 mM NaOH–10 mM EDTA in a plastic trough. After 5 min, 1 ml of pre-warmed hybridization solution (2× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate], 50 mM Tris-HCl [pH 7.5], 0.1% SDS) was added, and a LiPA strip was submerged and incubated in a shaking water bath at 50 ± 0.5°C for 1 h. The strips were washed with 1 ml of 2× SSC–0.1% SDS for 30 min at 50°C. Subsequently, the strips were rinsed three times in phosphate buffer, and conjugate (streptavidin-alkaline phosphatase) was added. After incubation at room temperature for 30 min, the strips were rinsed again and 4-nitroblue tetrazolium chloride and 5-bromo-4-chloro-3-indolylphosphate substrate was added. Hybrids are visible as purple probe lines. Interpretation of the hybridization patterns was performed visually.

RESULTS

***vacA* S region.** The sequence heterogeneity of the *vacA* s and m regions in *H. pylori* isolates obtained from different locations in Europe, North America, South America, Oceania, and Asia was analyzed by PCR and direct sequencing of the amplimers. A PCR amplimer of one of the expected sizes (176 bp for s1 and 203 bp for s2) was generated from each of the 98 isolates, indicating the broad specificity of the primers. For analysis of the s region, 78 PCR amplimers were sequenced, including 67 type s1 fragments and 11 type s2 fragments. Phylogenetic anal-

TABLE 2. Allele-specific probes for *vacA* and *cagA* used for reverse hybridization

Gene-segment and probe designation	Probe sequence (5′→3′)[a]	Positions	Specificity
vacA-S			
P1S1	GGAGCRTTRGTCAGCATCAC	61–80[b]	s1a
P22S1a	GCTTTAGTAGGAGCRTTRGTC	52–72[b]	s1a
P1S1b	GGAGCGTTGATTAGYKCCAT	61–80[b]	s1b
P2S1b	GTTTTAGCAGGAGCGTTGA	52–72[b]	s1b
P3s1	GGGYTATTGGTYAGCATCAC	26–45[b]	s1c
P4s1	GCTTTAGTRGGGYTATTGGT	17–36[b]	s1c
P1S2	GCTAAYACGCCAAAYGATCC	88–107[c]	s2
P2S2	GATCCCATACACAGCGAGAG	103–122[c]	s2
vacA-M			
P1M1	TTGATACGGGTAATGGTGG	1526–1544[b]	m1
P2M1	GGGTAATGGTGGTTTCAACA	1533–1552[b]	m1
P1M2a	ACGAATTTAAGAGTGAATGGC	1522–1542[c]	m2a
P2M2a	AGAGCGATAACGGGCTAAACA	1577–1597[c]	m2a
P2M2b	AGGGTAGAAATGGTATCGACA	1577–1597[d]	m2b
cagA			
cagApro1	GTTGATAACGCTGTCGCTTC	94–113[e]	*cagA*
cagApro2	TAATCTTCARGTRGCTTTTCTT	69–90[e]	*cagA*

[a] R = G or A, Y = C or T, and K = G or T.
[b] Positions according to the *vacA* s1a-m1 open reading frame of *H. pylori* 60190 (GenBank accession no. U05676).
[c] Positions according to the *vacA* s2-m2 open reading frame of *H. pylori* Tx30a (GenBank accession no. U29401).
[d] No reference sequences of m1b are available, but the position numbers correspond to the positions in m2a sequences.
[e] Positions corresponding to the *cagA* open reading frame of *H. pylori* ATCC 53726 (GenBank accession no. L11714).
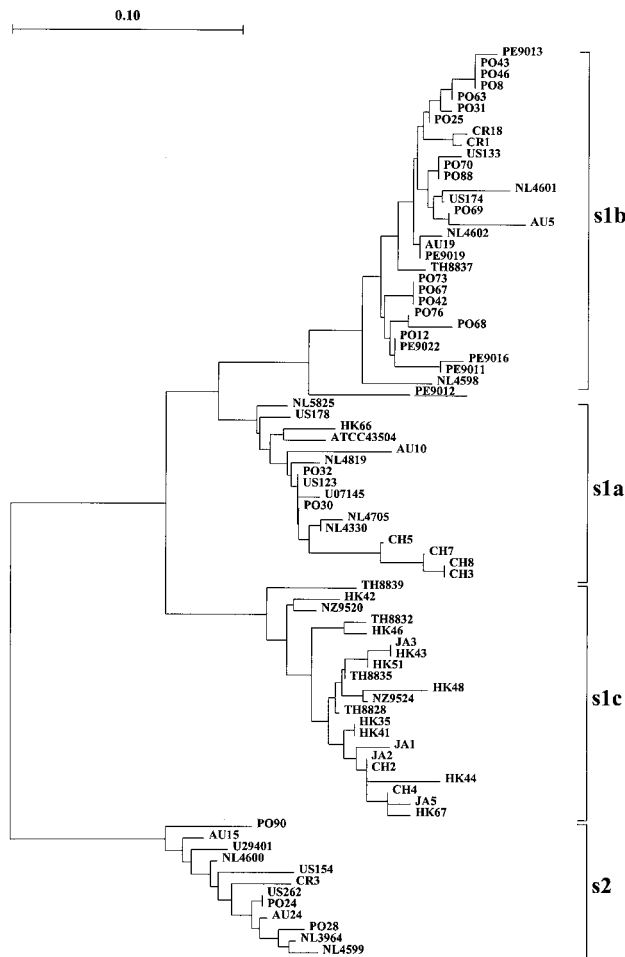
FIG. 1. Phylogenetic tree of *vacA* s-region nucleotide sequences, including two reference sequences (U07145 and U29401). Clusters of subtypes s1a, s1b, s1c, and s2 are indicated. The sequence from PE9012, containing recombinant s1a-s1b sequences, is underlined. Letters in the names of the isolates indicate the country or location of origin: Australia (AU), Costa Rica (CR), China (CH), Hong Kong (HK), Japan (JA), The Netherlands (NL), New Zealand (NZ), Peru (PE), Portugal (PO), Thailand (TH), and the United States (US). A reference bar is shown for molecular distance.

ysis showed that the type s2 sequences form a homogeneous cluster (Fig. 1). There were no substantial differences among type s2 sequences obtained from strains of different geographic origins. Among type s1 sequences, three distinct subtypes were

clearly distinguished, corresponding to the previously described subtypes s1a, s1b (4), and a new subtype, designated s1c. Type s1c shows consistent variation at several loci compared to types s1a and s1b and can be considered a specific recombinant of s1a and s1b (Fig. 2). The deduced amino acid sequences also revealed consistent differences among the three different s1 subtypes (Fig. 3). We confirm a previous report (4) that there are five amino acid differences between subtypes s1a and s1b (i.e., Ala18Val, Val20Ala, Val24Ile, Ile26Ala/Ser, and Thr27Ile). Subtype s1c differs from subtype s1a at two positions (Ala22Leu and Gln30Lys) and from s1b at seven positions (Ala18Val, Val20Ala, Ala22Leu, Ile24Val, Ala/Ser26Ile, Ile27Thr, and Glu/Gln30Lys). s1c strains contain two specific amino acids (Leu22 and Lys30) not found in other strains of the other subtypes and were isolated almost exclusively from persons from East Asia (Fig. 1). Within the s1a and s1b subtypes, there was no correlation with the geographic origins of the *H. pylori* isolates.

Pairwise comparisons of all sequences showed that the nucleotide sequences within each subtype are more than 95% conserved. Transitions were more frequent than transversions, and the majority of nucleotide substitutions were synonymous. The deduced amino acid sequences of all subtypes are more than 94% conserved (Table 3). The *vacA* s region sequence from PE9012 is located at an intermediate branch in the phylogenetic tree between the s1a and s1b clusters (Fig. 1) and can be considered a recombinant of s1a and s1b (Fig. 2). The PCR product from PE9012 hybridized strongly to probe P2s1b but very weakly to probe P1s1b (data not shown), and this result is in complete agreement with the sequence analysis results.

vacA **m region.** For analysis of the m region, 82 PCR amplimers, including 38 type m1 fragments and 44 type m2 fragments, comprising 288 and 362 bp, respectively, were examined by sequence analysis. Phylogenetic analysis (Fig. 4) indicated a clear separation between m1 and m2 sequences (sequence distances, as calculated with DNADIST by the Kimura two-parameter method, were 0.333 ± 0.026), largely because the m2 sequences contain an additional 75 nucleotides that are not present in m1 sequences. The sizes of the amplimers of two isolates from Hong Kong (isolates HK41 and HK46) and one isolate from China (isolate CH4) were the same as those for other m2 strains, but the amplimers differed markedly from those of other m2 sequences, as shown by their position on a separate branch. Molecular distances between the 362-bp nucleotide sequences from these three isolates and the other 41 m2 sequences were 0.198 ± 0.012, whereas among the latter group the distances were only 0.036 ± 0.017. Since these sequence distance distributions were not overlapping, the three


```
                36
      s2        |
      U29401  CCCTATTATTTCTCTCGCTTTAGTGGGGGTGTTAATGGGCACCGAACTGGGGGCTAACACGCCAAATGATCCCATACACAGCGAGAGTCGCGCTTTTTTC
      NL3964  ---------C---------C-------------------T--------A----------------C-----------------------------C-----T
      s1a
      U07145  ----C-GG---------------A--A-CA---G-CA---T-AC--.................---..--CA-AGT---G-..............---C------
      CH3     ----C-GG-------T--------A--A-CA---GG-CA---T-AC--.................---..--CA-AGT---G-..............----C----T
      AU10    ---------C-------------A--A-CA---GG-CA---T-AC--.................---..--CA-AGT---G-..............---C------
      rec s1a/s1b
      PE9012  ----C-GG-------G-T-----CA--A-C---G--TA---T-AC--.................--A-...-CA-AGT---G-..............---C------
      s1b
      PO76    ----T-AG---------T-----CA--A-C---G--TA--T--AT--.................---..--CA-AGT---G-..............---C------
      PO70    ----T-AG---------T-----CA--A-C---G--TA-TG--AT--.................---..--GAGAGT---G-..............---C------
      s1c
      HK43    ----T-AG------------------T-A--GG-CA---T-AC--.................--A-...-AA-AGT---G-.............T--C--C--T
      JA3     ----T-AG------------------T-A--GG-CA---T-AC--.................--A-...-AA-AGT---G-.............T--C--C--T
```
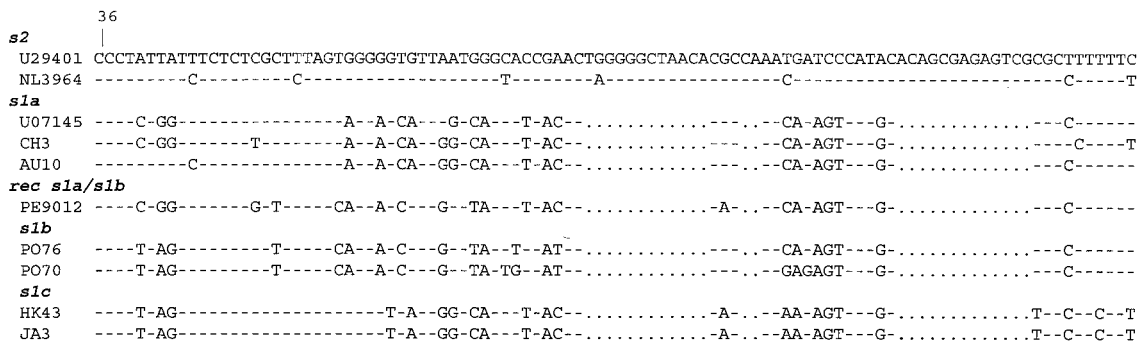
FIG. 2. Alignment of partial nucleotide sequences of the *vacA* s region, showing differences between s2, s1a, s1b, and s1c subtypes. The recombinant s1a-s1b sequence from PE9012 is also shown. Each of the 100-bp sequences presented here starts at position 36 of the *vacA* open reading frame. To permit alignment between type s1 and s2 sequences, a dot indicates the absence of a nucleotide in the s1 variants. Identical nucleotides are indicated by a hyphen.

TABLE 3. Nucleotide and amino acid sequence homologies between different *vacA* s-region genotypes

Genotype	Ratio		% Similarity of nucleotide (amino acid) sequences[a]			
	Transitions: transversions[b]	Nonsynonymous: synonymous[c]	s1a	s1b	s1c	s2
s1a (*n* = 16)	3.56	0.059	96.0 ± 2.5 (97.8 ± 1.9)	86.9 ± 2.5 (80.4 ± 2.3)	89.9 ± 2.3 (93.1 ± 1.4)	61.0 ± 1.4 (56.2 ± 2.7)
s1b (*n* = 30)	2.51	0.114		96.0 ± 1.3 (94.0 ± 2.8)	82.6 ± 2.4 (78.7 ± 3.5)	60.2 ± 1.7 (49.1 ± 3.2)
s1c (*n* = 21)	9.72	0.021			95.9 ± 2.0 (96.1 ± 1.4)	61.8 ± 0.9 (53.4 ± 2.2)
s2 (*n* = 12)	14.08	0.079				96.5 ± 1.5 (95.4 ± 2.6)

[a] Percentages are based on pairwise comparisons of sequences of the corresponding subtypes and are given as means ± standard deviations.
[b] The values represent the ratio of transitions to transversions, based on pairwise comparisons of all sequences belonging to the same subtype.
[c] The values represent the ratio of nonsynonymous to synonymous substitutions, based on pairwise comparisons of all sequences belonging to the same subtype.

sequences were classified into a separate subtype, designated m2b. All other m2 sequences were named m2a. The three m2b strains had been typed s1c. Within the m2a cluster, there was no apparent association with the geographic origins of the strains.

Within the group of 38 type m1 sequences, sequence distances were 0.059 ± 0.031. The sequences obtained from nine Hong Kong isolates and two Japanese isolates formed a separate phylogenetic cluster, with sequence distances of 0.018 ± 0.009, suggesting the presence of specific m1 variation in these isolates (data for only seven isolates are presented in Fig. 4). However, the molecular distances between these 11 sequences and the remaining 27 m1 sequences were only 0.090 ± 0.012. Since the sequence distance distributions are overlapping, the sequence variation is not sufficient to warrant the definition of two separate subtypes within m1. Pairwise comparisons of all sequences showed that the nucleotide sequences within each subtype are more than 95% conserved. Transitions were more frequent than transversions, and the majority of nucleotide substitutions were synonymous. The deduced amino acid sequences are more than 91% conserved (Table 4).

On the basis of the sequence data generated, a set of general PCR primers was developed for universal amplification of the m region. To ensure amplification from all m-region variants, a mix of four forward primers (primers VAMF1.1 to VAMF1.4) was used (Table 1). The target sequence for the reverse primer MR1 appeared to be highly conserved. These PCR primers allowed specific amplification of m-region sequences from all 98 investigated isolates.

Reverse hybridization. To develop a rapid, sensitive, and specific method that could be used for the genotyping of *H. py-*

lori strains for clinical purposes, we used the sequence data to design a reverse hybridization LiPA. Oligonucleotide probes specific for s1a, s1b, s1c, m1, m2a, and m2b (as indicated in Table 2) were deduced from the sequences of the s and m regions. All probes were individually optimized to obtain a high degree of specificity under the defined reverse hybridization conditions. Initial studies used well-characterized strains, and there was agreement between LiPA and genotyping data


```
s2      1        10        20        30        40        50
        |        |         |         |         |         |
U29401  MEIQQTHRKINRPIISLALVGVLMGTELGANTPNDPIHSESRAFFTTVIIPAIVG
NL3964  ------------------------------------------------------
s1a
U07145           -LV------A-VSI.....-QQSHA....--------------
CH3              -LV------A-VSI.....-QQSHA....--------------
AU10             ---------A-VSI.....-QQSHA....-------F------
rec s1a/s1b
PE9012           -LV--V-A-A-ISI......-QQSHA....--------------
s1b
PO76             -LV--V-A-A-ISS.....I-QQSHA....--------------
PO70             -LV--V-A-A-ISA.....I-QESHA....--------------
s1c
HK43             -LV------L-VSI.....--QKSHA....--------------
JA3              -LV------L-VSI.....--QKSHA....--------------

s1       1        10        20        30        40
         |        |         |         |         |
```
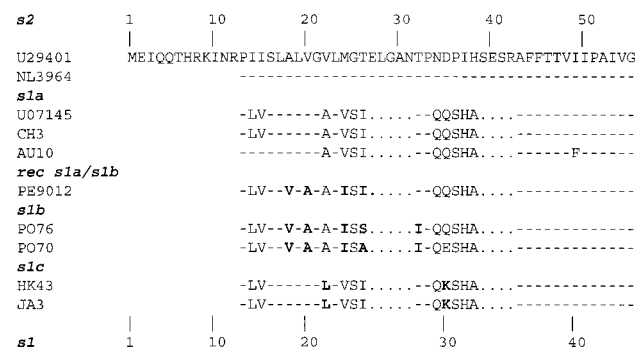
FIG. 3. Alignment of partial amino acid sequences of the *vacA* s region. Consistent differences between s1a, s1b, and s1c subtypes are indicated in bold-face. To obtain proper alignment between type s1 and s2 sequences, a dot indicates the absence of an amino acid in s1 variants. Identical amino acids are indicated by a hyphen. The sequences U07145 and U29401 are also presented as references for s1a and s2, respectively.
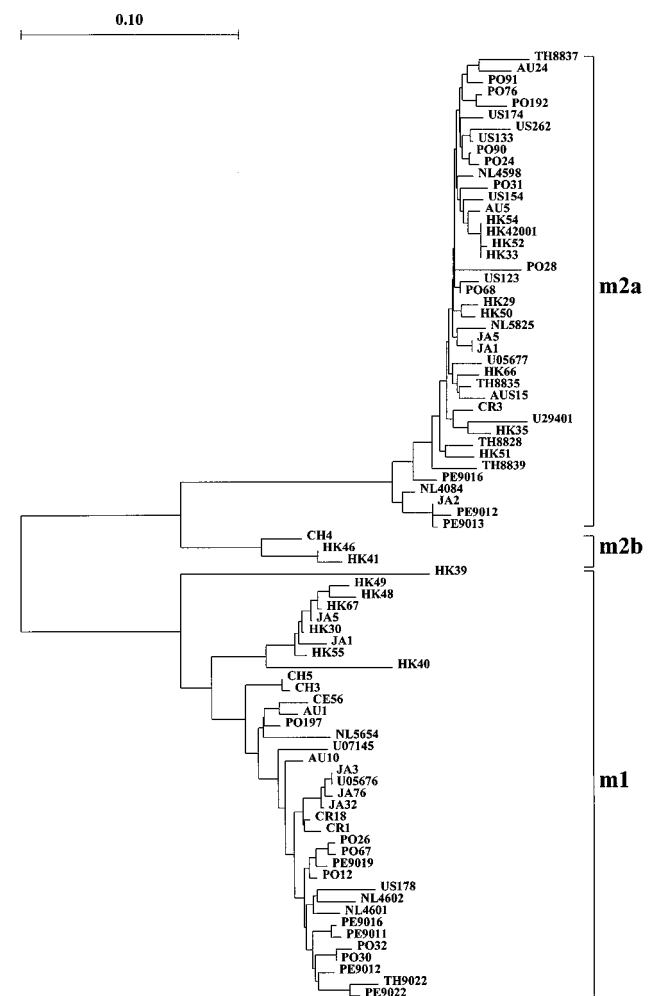
FIG. 4. Phylogenetic tree of *vacA* m-region sequences, including four reference sequences (U05676, U05677, U07145, and U29401). Clusters of type m1 and subtype m2a and m2b are indicated. The origins of the isolates are defined in the legend to Fig. 1.

TABLE 4. Nucleotide and amino acid sequence homologies between different *vacA* m-region genotypes

Genotype	Ratio		% Similarity of nucleotide (amino acid) sequences[a]		
	Transitions:transversions[b]	Nonsynonymous:synonymous[c]	m1	m2a	m2b
m1 (*n* = 38)	2.47	0.284	95.2 ± 1.1 (91.9 ± 3.1)	57.7 ± 0.5 (50.0 ± 2.9)	62.9 ± 0.6 (49.9 ± 3.9)
m2a (*n* = 41)	1.36	0.294		96.4 ± 1.2 (94.6 ± 3.1)	82.5 ± 0.7 (72.9 ± 6.8)
m2b (*n* = 3)	2.13	0.284			95.7 ± 2.1 (95.3 ± 2.5)

[a] Percentages are based on pairwise comparisons of sequences of the corresponding subtypes and are given as means ± standard deviations.
[b] The values represent the ratio of transitions to transversions, based on pairwise comparisons of all sequences belonging to the same subtype.
[c] The values represent the ratio of nonsynonymous to synonymous substitutions, based on pair-wise comparisons of all sequences belonging to the same subtype.

(Fig. 5). The results of subsequent genotyping based on the hybridization patterns of the LiPA were completely concordant with the results of genotyping based on sequence analysis of the *vacA* s region (78 strains) and m region (82 strains).

DISCUSSION

Knowledge of the existence of different *H. pylori* genotypes may become clinically important since strains containing *cagA* are more likely to cause more severe disease than strains that lack *cagA* (7, 15, 33), 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 (42). Similarly, adhesion of *H. pylori* to the human gastric epithelium, mediated by histo-blood group antigens, also appeared to be related to distinct allelic variants of the bacterial *babA* gene (19).

Although *H. pylori* is cosmopolitan, with prevalences ranging from approximately 30% in developed countries to more than 80% in the developing world, little is known about the geographic distributions of specific *H. pylori* strains (26, 29, 34). Differences in the prevalence of particular *H. pylori* genotypes could play a role in the incidence of *H. pylori*-associated diseases in certain localities.

Analysis of *H. pylori* isolates from diverse geographic locations permitted a comprehensive description of the s and m regions of *vacA*. The recognition of two main s-region types, s1 and s2, with both the size and amino acid sequences of isolates from around the world being highly conserved, confirmed the findings of Atherton et al. (4), who studied U.S. isolates only.

However, in addition to the s1a and s1b subtypes identified previously, a third subtype (subtype s1c) was defined in the present study. In a study based on subtype-specific PCR without hybridization or sequence analysis, Ito et al. (20) found that s1a was the predominant type in Japan. Our present data indicate that PCR analysis with primers specific for subtypes s1a and s1b, as described earlier (4), would identify all s1c isolates as subtype s1a. The present study, as well as a more extensive study of a worldwide collection of strains (data not shown), indicates that s1c is the most prevalent subtype in isolates from East Asia. Interestingly, the one Dutch patient colonized with an s1c *H. pylori* strain (as determined by LiPA) had a Chinese mother and was born in Indonesia. Whether s1c strains are phenotypically different from s1a and s1b strains remains to be determined.

Despite the very high overall genetic heterogeneity of *H. pylori* (18), the distinct s- and m-region variants of *vacA* appear to be well conserved (>95% at the nucleotide level), consistent with the functional constraints of the encoded cytotoxin. The nucleotide substitution rates (transitions to transversions and nonsynonymous to synonymous substitution rates) observed in the s region indicate a high degree of sequence conservation, as would be expected for a signal sequence. Subtype s1c sequences were especially highly conserved. The selective pressure in the m region appears to be less than that in the s region, as shown by higher nonsynonymous to synonymous substitution rates. The substitution rates of m1 and m2 variants were very similar. On the basis of phylogenetic analyses, the number of recombinants between the distinct subtypes appears
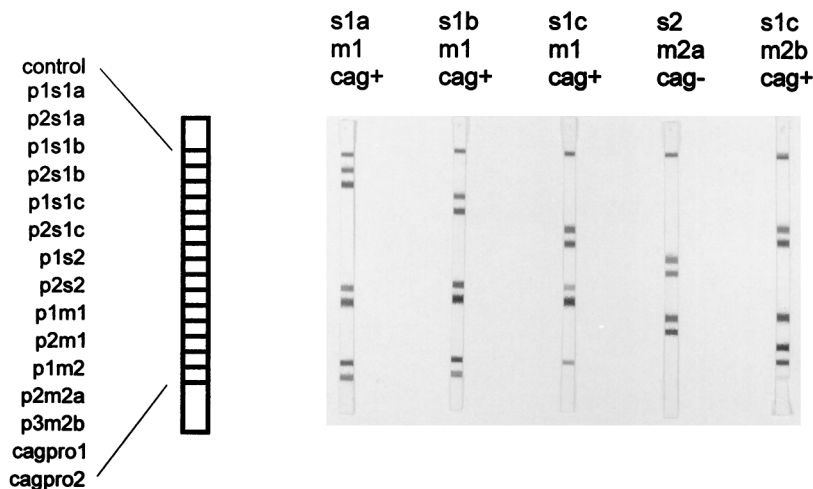

FIG. 5. Use of the reverse hybridization LiPA to determine the *vacA* and *cagA* genotypes of *H. pylori* strains whose sequences were known. The strains (genotypes) are as follows: AU10 (s1a, m1, *cagA* positive), PO12 (s1b, m1, *cagA* positive), HK44 (s1c, m1, *cagA* positive), PO24 (s2, m2a, *cagA* negative), and HK46 (s1c, m2b, *cagA* positive). The positions of the specific probes, as listed in Table 2, are indicated.

to be limited. However, PCR products were not cloned but were directly sequenced. Therefore, it is also possible that the recombinant s1a-s1b sequence represents consensus sequences of PCR products from mixed strains. Although type s2 strains produce little or no active cytotoxin in vitro (4, 16), phylogenetic analysis revealed that type s2 sequences are as closely related to one another as sequences from the toxin-producing type s1 strains. This observation suggests the existence of selective pressure on type s2 alleles as well, implying that the s2 gene product has a functional role in vivo.

Among the m-region sequences, both type m1 and m2 sequences were well conserved. There was no obvious association between sequence variants and the geographic origins of the isolates. Although type m1 sequences from East Asian isolates differed from that of the main m1 variant, this difference was insufficient to permit molecular classification as a separate subtype. Although the amplimer size for the distinct m2 variant (m2b) found in three East Asian isolates was the same as that for m2a, subtype m2b represents a separate branch in the phylogenetic tree and can be clearly distinguished from subtype m2a.

Earlier PCR analysis of *vacA* (4) focused on a more downstream m region. Several isolates from Japan failed to yield a PCR product when the original m-region primers were used (24), suggesting sequence heterogeneity in this part of *vacA*. The fact that the novel m-region primers used in this study, as deduced from the alignment of 86 sequences, permitted efficient amplification of m-region sequences from all tested strains indicates their broad applicability for analysis of *H. pylori* isolates from various geographic locations.

The PCR-reverse hybridization method reported here offers a tool for epidemiological and clinical studies and permits the rapid identification of the various *vacA* allelic types and the detection of *cagA*. Hybridization analysis is more reliable than visual inspection of PCR products on agarose gels and is required to specifically discriminate among defined subtypes. PCR can be performed not only with cultured isolates but also directly with gastric biopsy specimens (41), circumventing expensive and time-consuming culture procedures. Some biopsy specimens contain multiple strains (21, 35), and the reverse hybridization assay is a very sensitive method for the simultaneous detection of multiple genotypes.

In conclusion, analysis of a worldwide collection of *H. pylori* strains showed that the distinct *vacA* alleles are well conserved. Extension of previous analyses revealed that substantial s- and m-region heterogeneity exists, particularly among strains from East Asia. Further subtyping of *H. pylori* strains by these methods should facilitate insights into the evolution, epidemiology, and clinical importance of these organisms.

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.** 1996. Role of *vacA* and the *cagA* locus of *Helicobacter pylori* in human disease. Aliment. Pharmacol. Ther. **10**(Suppl. 1)**:**73–77.
9. **Blaser, M. J.** 1997. Ecology of *Helicobacter pylori* in the human stomach. J. Clin. Invest. **100:**759–762.
10. **Bukanov, N. O., and D. E. Berg.** 1994. Ordered cosmid library and high-resolution physical-genetic map of *Helicobacter pylori* strain NCTC11638. Mol. Microbiol. **11:**509–523.
11. **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.
12. **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.
13. **Cover, T. L.** 1996. The vacuolating cytotoxin of *Helicobacter pylori*. Mol. Microbiol. **20:**241–246.
14. **Cover, T. L., and M. J. Blaser.** 1992. Purification and characterization of the vacuolating toxin from *Helicobacter pylori*. J. Biol. Chem. **267:**10570–10575.
15. **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.
16. **Cover, T. L., M. K. Tummuru, P. Cao, S. A. Thompson, and M. J. Blaser.** 1994. Divergence of genetic sequences for the vacuolating cytotoxin among *Helicobacter pylori* strains. J. Biol. Chem. **269:**10566–10573.
17. **Felsenstein, J.** 1993. PHYLIP, Phylogeny Inference Package, version 3.5c (computer program). J. Felsenstein, University of Washington, Seattle. (Distributed by the author.).
18. **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.
19. **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.
20. **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.
21. **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.
22. **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.
23. **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.
24. **Maeda, S., K. Ogura, M. Ishitobi, F. Kanai, H. Yoshida, S. Ota, Y. Shiratory, and M. Omata.** 1996. Diversity of *Helicobacter pylori vacA* gene in Japanese strains—high activity type s1 is dominant in Japan. Gastroenterology **100:**A182.
25. **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.
26. **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.
27. **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.
28. **Phadnis, S. H., D. Ilver, L. Janzon, S. Normark, and T. U. Westblom.** 1994. Pathological significance and molecular characterization of the vacuolating toxin gene of *Helicobacter pylori*. Infect. Immun. **62:**1557–1565.
29. **Schlemper, R. J., S. D. van der Werf, J. P. Vandenbroucke, I. Biemond, and C. B. Lamers.** 1996. Risk factors of peptic ulcer disease: difference impact of

Helicobacter pylori in Dutch and Japanese populations? J. Gastroenterol. Hepatol. **11:**825–831.

30. **Schmitt, W., and R. Haas.** 1994. Genetic analysis of the *Helicobacter pylori* vacuolating cytotoxin: structural similarities with the IgA protease type of exported protein. Mol. Microbiol. **12:**307–319.

31. **Simmonds, P., E. C. Holmes, T. A. Cha, S. W. Chan, F. McOmish, B. Irvine, E. Beall, P. L. Yap, J. Kolberg, and M. S. Urdea.** 1993. Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. J. Gen. Virol. **74:**2391–2399.

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. **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 isolates from patients with either peptic ulcers or non-ulcer dyspepsia. Am. J. Gastroenterol. **93:**30–34.

34. **Talley, N. J., and H. Xia.** 1996. Ethnicity as a risk factor for *Helicobacter pylori* infection and gastric cancer: environment, genetics, or both? Aust. N. Z. J. Med. **26:**628–631.

35. **Taylor, N. S., J. G. Fox, N. S. Akopyantz, 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. **Telford, J. L., P. Ghiara, M. Dell'Orco, M. Comanducci, D. Burroni, M.** Bugnoli, M. F. Tecce, S. Censini, A. Covacci, and Z. Xiang. 1994. Gene structure of the *Helicobacter pylori* cytotoxin and evidence of its key role in gastric disease. J. Exp. Med. **179:**1653–1658.

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., 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.

41. **van Doorn, L. J.** Unpublished observations.

42. **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.

43. **Xiang, Z., S. Censini, P. F. Bayeli, J. L. Telford, N. Figura, R. Rappuoli, and A. Covacci.** 1995. Analysis of expression of *cagA* and *vacA* virulence factors in 43 strains of *Helicobacter pylori* reveals that clinical isolates can be divided into two major types and that *cagA* is not necessary for expression of the vacuolating cytotoxin. Infect. Immun. **63:**94–98.