Expanding Allelic Diversity of Helicobacter pylori vacA

LEEN-JAN VAN DOORN,1,* CEU 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. QUINT1

Delft Diagnostic Laboratory, 2625 AD Delft,1 and Free University Hospital, Department of Gastroenterology, Amsterdam,8 The Netherlands; IPATIMUP and Medical Faculty, University of Porto, Porto, Portugal2; Monash Medical Center, Victoria, Australia4; Prince of Wales Hospital, Hong Kong, China6; Division of Infectious Diseases, Vanderbilt University, Nashville, Tennessee1; and Division of Gastroenterology and Institute of Infections and Immunity, University Hospital, Nottingham, United Kingdom6

Received 27 February 1998/Returned for modification 28 May 1998/Accepted 11 June 1998

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 (m1a, 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. Amplifiers 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 cagA genes have been identified (7, 8, 43). The cytotoxin-associated gene (cagA), which is not present in most strains, is associated with more severe clinical outcomes (6, 22, 27). 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.

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 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 of diverse 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 = 11), Portgual (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% O2, 10% CO2, 85% N2). 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 x 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
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.

The presence of vacA was detected with primers cagAF and cagAR, yielding an amplimer of one of the expected sizes (176 bp for s1 and 203 bp for s2) was generated from each of the 98 isolates, and the PCR amplicon was analyzed by PCR and direct sequencing of the amplimers.

### 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 amplicons. 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 amplicons were sequenced, including 67 type s1 fragments and 11 type s2 fragments. Phylogenetic analy-

**TABLE 1.** PCR primers for vacA and cagA used in this study

<table>
<thead>
<tr>
<th>Gene and segment</th>
<th>Primer designation (polarity)</th>
<th>Primer sequence (5’–3’)</th>
<th>Positions</th>
</tr>
</thead>
<tbody>
<tr>
<td>vacA S</td>
<td>VA1F (+)</td>
<td>Bio-ATGGAATAACACAAACACACACAC</td>
<td>1–21†</td>
</tr>
<tr>
<td>vacA S</td>
<td>VA1XR (–)</td>
<td>Bio-CCTGARACCGCTCTAACG</td>
<td>157–176†</td>
</tr>
<tr>
<td>vacA M</td>
<td>HPMGF (+)</td>
<td>CACAGCCACTTTCAATAAAGA</td>
<td>1443–1465†</td>
</tr>
<tr>
<td>vacA M</td>
<td>HPMGR (–)</td>
<td>CGTCAAAAATAATCCAGGG</td>
<td>1824–1843†</td>
</tr>
<tr>
<td>vacA M</td>
<td>MF1.1 (+)</td>
<td>Bio-GTGGATGCCCATACG</td>
<td>1495–1514†</td>
</tr>
<tr>
<td>vacA M</td>
<td>MF1.2 (+)</td>
<td>Bio-GTGGATGCCCATACG</td>
<td>1490–1499†</td>
</tr>
<tr>
<td>vacA M</td>
<td>MF1.3 (+)</td>
<td>Bio-GTGGATGCCCATACG</td>
<td>1495–1514†</td>
</tr>
<tr>
<td>vacA M</td>
<td>MF1.4 (+)</td>
<td>Bio-GCCAGGGCTATACGGTCA</td>
<td>1495–1514†</td>
</tr>
<tr>
<td>vacA M</td>
<td>MR1 (–)</td>
<td>Bio-RTGAAGTGTGATATGGAC</td>
<td>1582–1601†</td>
</tr>
<tr>
<td>vacA M</td>
<td>MR2 (–)</td>
<td>Bio-ACTCCCTAATGTGGAGGATTC</td>
<td>178–199†</td>
</tr>
</tbody>
</table>

† R = G or A, Y = C or T, M = A or C, Bio = biotin at the 5’ end of the primer.

The primer was from Atherton et al. (4).

† Positions corresponding to the vacA s1-m1 open reading frame of *H. pylori* 60190 (GenBank accession no. U05676).

† Positions corresponding to the vacA s2-m2 open reading frame of *H. pylori* T30a (GenBank accession no. U29401).

† Positions corresponding to the cagA open reading frame of *H. pylori* ATCC 53726 (GenBank accession no. L11714).

2000 VAN DOORN ET AL. J. CLIN. MICROBIOL.

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

<table>
<thead>
<tr>
<th>Gene-segment and probe designation</th>
<th>Probe sequence (5’–3’)</th>
<th>Positions</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>vacA-S</td>
<td>GGACGTRRRTGCTACGCTAC</td>
<td>61–80</td>
<td>s1</td>
</tr>
<tr>
<td>P21S1</td>
<td>GGCTTAAGTTGCAGRHG</td>
<td>52–72</td>
<td>s1</td>
</tr>
<tr>
<td>P21b</td>
<td>GGCGGTGGATAGCGTGC</td>
<td>61–80</td>
<td>s1</td>
</tr>
<tr>
<td>P21b</td>
<td>GTTTTACAGGAGGCTG</td>
<td>52–72</td>
<td>s1</td>
</tr>
<tr>
<td>P31</td>
<td>GGCTTAGGTAGGGATAC</td>
<td>26–45</td>
<td>s1c</td>
</tr>
<tr>
<td>P41</td>
<td>GCCTTATGGGCTGATAGT</td>
<td>17–36</td>
<td>s1c</td>
</tr>
<tr>
<td>P152</td>
<td>GCTAAYAGCCTAAAYGATGTC</td>
<td>88–107</td>
<td>s2</td>
</tr>
<tr>
<td>P2S2</td>
<td>GCATCCCATACAGCGGAGG</td>
<td>103–122</td>
<td>s2</td>
</tr>
</tbody>
</table>

† R = G or A, Y = C or T, M = A or C, Bio = biotin at the 5’ end of the primer.

† Positions according to the vacA s1-m1 open reading frame of *H. pylori* 60190 (GenBank accession no. U05676).

† Positions according to the vacA s2-m2 open reading frame of *H. pylori* T30a (GenBank accession no. U29401).

† Positions according to the cagA open reading frame of *H. pylori* ATCC 53726 (GenBank accession no. L11714).

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 5′-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 0.01 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 amplicons. 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 amplicons 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

<table>
<thead>
<tr>
<th>Gene-segment and probe designation</th>
<th>Probe sequence (5’–3’)</th>
<th>Positions</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>vacA-S</td>
<td>GGACGTRRRTGCTACGCTAC</td>
<td>61–80</td>
<td>s1</td>
</tr>
<tr>
<td>P21S1</td>
<td>GGCTTAAGTTGCAGRHG</td>
<td>52–72</td>
<td>s1</td>
</tr>
<tr>
<td>P21b</td>
<td>GGCGGTGGATAGCGTGC</td>
<td>61–80</td>
<td>s1</td>
</tr>
<tr>
<td>P21b</td>
<td>GTTTTACAGGAGGCTG</td>
<td>52–72</td>
<td>s1</td>
</tr>
<tr>
<td>P31</td>
<td>GGCTTAGGTAGGGATAC</td>
<td>26–45</td>
<td>s1c</td>
</tr>
<tr>
<td>P41</td>
<td>GCCTTATGGGCTGATAGT</td>
<td>17–36</td>
<td>s1c</td>
</tr>
<tr>
<td>P152</td>
<td>GCTAAYAGCCTAAAYGATGTC</td>
<td>88–107</td>
<td>s2</td>
</tr>
<tr>
<td>P2S2</td>
<td>GCATCCCATACAGCGGAGG</td>
<td>103–122</td>
<td>s2</td>
</tr>
</tbody>
</table>

† R = G or A, Y = C or T, M = A or C, Bio = biotin at the 5’ end of the primer.

† Positions according to the vacA s1-m1 open reading frame of *H. pylori* 60190 (GenBank accession no. U05676).

† Positions according to the vacA s2-m2 open reading frame of *H. pylori* T30a (GenBank accession no. U29401).

† No reference sequences of m1b are available, but the position numbers correspond to the positions in m2a sequences.

† Positions corresponding to the cagA open reading frame of *H. pylori* ATCC 53726 (GenBank accession no. L11714).
Analysis 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
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. pylori* 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.

### Table 3. Nucleotide and amino acid sequence homologies between different vacA s-region genotypes

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Transitions: transversions(^a)</th>
<th>Nonsynonymous: synonymous(^b)</th>
<th>% Similarity of nucleotide (amino acid) sequences(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>s1a (n = 16)</td>
<td>3.56</td>
<td>0.059</td>
<td>96.0 ± 2.5 (97.8 ± 1.9)</td>
</tr>
<tr>
<td>s1b (n = 30)</td>
<td>2.51</td>
<td>0.114</td>
<td>86.9 ± 2.5 (80.4 ± 2.3)</td>
</tr>
<tr>
<td>s1c (n = 21)</td>
<td>9.72</td>
<td>0.021</td>
<td>89.9 ± 2.3 (93.1 ± 1.4)</td>
</tr>
<tr>
<td>s2 (n = 12)</td>
<td>14.08</td>
<td>0.079</td>
<td>61.0 ± 1.4 (56.2 ± 2.7)</td>
</tr>
</tbody>
</table>

\(^{a}\) 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.
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 *s* *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 *vacA* nucleotide substitution rates (transitions to transversions and nonsynonymous to synonymous substitutions) 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

**DISCUSSION**

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

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Ratio</th>
<th>% Similarity of nucleotide (amino acid) sequences&lt;sup&gt;a&lt;/sup&gt;</th>
<th>m1</th>
<th>m2a</th>
<th>m2b</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1 (n = 38)</td>
<td>2.47</td>
<td>0.284</td>
<td>95.2 ± 1.1 (91.9 ± 3.1)</td>
<td>57.7 ± 0.5 (50.0 ± 2.9)</td>
<td>62.9 ± 0.6 (49.9 ± 3.9)</td>
</tr>
<tr>
<td>m2a (n = 41)</td>
<td>1.36</td>
<td>0.294</td>
<td>96.4 ± 1.2 (94.6 ± 3.1)</td>
<td>82.5 ± 0.7 (72.9 ± 6.8)</td>
<td>95.7 ± 2.1 (95.3 ± 2.5)</td>
</tr>
<tr>
<td>m2b (n = 3)</td>
<td>2.13</td>
<td>0.284</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Percentages are based on pairwise comparisons of sequences of the corresponding subtypes and are given as means ± standard deviations.

<sup>b</sup>The values represent the ratio of transitions to transversions, based on pairwise comparisons of all sequences belonging to the same subtype.

<sup>c</sup>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. 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.

ACKNOWLEDGMENTS

C. Figueiredo is supported by PRAXIS XXI (project 1/2.1/SAU/1356/95). This work was supported in part by grant R01 DK 53707 from the National Institutes of Health.

REFERENCES

16. Felsenstein, J. 1993. PHYML, Phylogeny Inference Package, version 3.5g (computer program). J. Felsenstein, University of Washington, Seattle. (Distributed by the author.)
40. van Doorn, L. J. Unpublished observations.