Population genetics of *Helicobacter pylori* in the southern part of Switzerland analysed by sequencing of four housekeeping genes (*atpD*, *glnA*, *scoB* and *recA*), and by *vacA*, *cagA*, *iceA* and IS605 genotyping

Nadia Maggi Solcà,¹ Marco V. Bernasconi,¹ Claudio Valsangiacomo,¹ Leen-Jan Van Doorn² and Jean-Claude Piffaretti¹

Author for correspondence: Jean-Claude Piffaretti. Tel: +41 91 923 25 22. Fax: +41 91 922 09 93.
e-mail: jean-claude.piffaretti@ti.ch

¹ Istituto Cantonale Batteriosierologico, Via Ospedale 6, 6904 Lugano, Switzerland

² Delft Diagnostic Laboratory, R. de Graafweg 7, 2625 AD Delft, The Netherlands

The population biology of 78 *Helicobacter pylori* strains (71 from Swiss Italian, 4 from East Asian and 3 from South African patients) was investigated by sequence analysis of four housekeeping genes: *atpD*, *scoB*, *glnA* and *recA*. The *vacA* genotype, the presence of *cagA* and IS605, the *iceA* allelic type, and the resistance to metronidazole, clarithromycin and amoxycillin were determined. A high percentage of DNA polymorphic sites (19.8% for *atpD*, 21.3% for *scoB*, 23.7% for *glnA* and 20.3% for *recA*) was found. The phylogenetic trees based on the nucleotide sequences of the four gene fragments showed different topologies and were incongruent. The virulence-associated markers were distributed over the dendrograms and no association was found with phylogenetic clusters or clinical manifestations (chronic gastritis, gastric or duodenal ulcer, MALT lymphoma). Moreover, the *H* ratios (calculated with the homoplasmy test) ranged from 0.742 to 0.799, depending on the gene fragment examined. All these observations suggest that *H. pylori* exists as a recombinant population. The clustering of the strains according to their geographical origin (USA/Europe, East Asia, South Africa) that has recently been demonstrated elsewhere could only be confirmed for the East Asian *vacA* s1c strains. In contrast, the South African strains clustered together only in the *atpD* tree. Presumably, recombination at the different loci has masked the evolutionary relationship among the strains.

Keywords: *H. pylori*, housekeeping genes, antibiotic resistance

INTRODUCTION

Helicobacter pylori is a microaerophilic, Gram-negative, slow-growing, spiral-shaped and flagellate micro-organism. Its ecological niche is the human gastric mucosa, where it can survive in an acid environment thanks to a powerful urease activity (Blaser, 1997). The bacterium, discovered in 1983 by Robin Warren and Barry Marshall, has been associated with various gastropathologies such as chronic atrophic gastritis,

peptic ulcer, gastric cancer and mucosa-associated lymphoid tissue lymphoma (MALT lymphoma) (Blaser *et al.*, 1995; Roggero *et al.*, 1995). Once acquired, *H. pylori* usually persists for life, unless eradicated by antimicrobial therapy. Even though the prevalence of infection may be very high (70–90% in developing countries, 25–50% in developed countries), most humans infected with *H. pylori* are asymptomatic and only a few patients develop peptic ulcer or gastric cancer (Pounder, 1995). Host genetic predisposition and local environmental factors, together with bacterial genotypes, may play an important role in the development of disease.

Since *H. pylori* is involved in gastric pathology, researchers have tried to identify specific virulence

Abbreviation: MALT lymphoma, gastric mucosa-associated lymphoid tissue lymphoma.

The GenBank accession numbers for the sequences reported in this paper are AY004351–AY004662

Table 1. *H. pylori* strains and their characteristics

Strain*	Clinical manifestation†	MTZ‡	CLA‡	<i>vacA</i> genotype	Presence of <i>cagA</i>	<i>iceA</i> allelic type	Presence of IS605
100A	CG	R	S	s1a-m1	Yes	A1	No
103A	CG	s	S	s2-m2a	Yes	A2	Yes
104C	CG	S	S	s1a-m2a	No	A1	No
106A	CG	S	S	s1a-m1	Yes	A1	No
10A2	CG	R	S	s1b-m1	Yes	A1	No
110A	CG	S	S	s2-m2a	Yes	A1	No
111A	CG	S	S	s1b-m1	Yes	A1	No
122A	DU	R	S	s1b-m2a	No	A2	No
123C	CG	S	S	s2-m2a	No	A2	No
124A	DU	R	S	s1b-m2a	No	A2	No
125A2	CG	S	S	s2-m2a	No	A1	No
128A	DU	R	R	s1a-m1	No	A1	Yes
136C2M2	CG	R	S	s2-m2a	No	A1	No
149C	DU	S	S	s2-m2a	Yes	A2	No
151A2	CG	S	S	s1a-m2a	Yes	A2	No
160A	DU	S	S	s1a-m2a	Yes	A2	No
163C	GU	R	R	s1a-m2a	No	A2	Yes
165C2	DU	R	R	s1b-m2a	Yes	A2	No
168A2	CG	S	S	s1b-m1	Yes	A1	No
177A	CG	S	S	s1b-m1	Yes	A2	No
186A	MALT	S	S	s1b-m1	Yes	A2	No
18A	CG	R	R	s1a-m1	Yes	A2	Yes
192A	DU	S	S	s1a-m1	Yes	A2	Yes
22A	CG	S	S	s1a-m1	Yes	A1	No
24C	CG	S	S	s2-m2a	Yes	A2	No
25A	DU	R	S	s1b-m2a	No	A1	No
27A	CG	S	S	s2-m2a	No	A2	No
28A	CG	S	S	s2-m2a	No	A2	No
293CG	CG	S	S	s1a-m1	No	A1	No
296A	DU	S	S	s2-m2a	No	A2	No
29A	CG	S	S	s2-m2a	No	A1	No
2A	CG	R	S	s1b-m1	Yes	A1	No
303A	DU	S	S	s1a-m1	Yes	A1	No
304A	DU	R	R	s1a-m1	No	A1	No
314C	DU	S	S	s2-m2a	No	A2	No
318A2	DU	R	S	s1a-m1	Yes	A2	No
31A	CG	R	S	s2-m2a	No	A1	No
322A	CG	R	R	s2a-m1	Yes	A1	Yes
32A	DU	S	S	s1a-m1	Yes	A1	No
330C	D	S	S	s2-m2a	No	A2	No
337C	CG	R	R	s1b-m2a	Yes	A1	No
335A	GU	S	S	s1a-m2a	Yes	A2	No
340A	CG	R	S	s2-m2a	No	A1	No
342C2	CG	R	R	s2-m2a	No	A1	No
345A2	CG	R	R	s1b-m1	Yes	A2	No
348C	CG	S	R	s1a-m2a	Yes	A1	Yes
356A1-2	D	S	S	s2-m2a	No	A2	No
35C	DU	S	S	s1a-m2a	Yes	A1	Yes
40A	CG	S	S	s2-m2a	No	A1	No
44A	GU	S	S	s1a-m2a	Yes	A1	No
47A	CG	R	S	s1b-m2a	Yes	A2	Yes
48A	CG	S	S	s1b-m1	Yes	A1	Yes
4C	DU	S	S	s2-m2a	No	A1	No

Table 1 (cont.)

Strain*	Clinical manifestation†	MTZ‡	CLA‡	<i>vacA</i> genotype	Presence of <i>cagA</i>	<i>iceA</i> allelic type	Presence of IS605
56A	DU	R	S	s2-m2a	No	A2	No
58A	D	R	S	s2-m2a	No	A1	No
61A	D	S	S	s2-m2a	No	A1	No
64A	CG	R	S	s1b-m2a	Yes	A1	No
65A	CG	R	S	s2-m2a	No	A2	Yes
70A	CG	S	R	s2-m2a	No	A1	No
73C	CG	R	R	s2-m2a	No	A1	No
7A	DU	S	S	s2-m2a	No	A1	No
81A	CG	R	S	s2-m2a	No	A1	No
83A2	MALT	S	S	s2-m2a	No	A1	No
84A	CG	R	S	s1b-m1	Yes	A2	Yes
86C	CG	S	S	s1a-m2a	Yes	A1	Yes
87A2	DU	R	S	s2-m2a	No	A1	No
88C	GU	S	S	s1b-m1	Yes	A1	No
95-1	DU	S	S	s1b-m2a	Yes	A1	No
97A	DU	S	S	s2-m2a	No	A1	No
98A	DU	S	S	s2-m2a	No	A2	No
99A2	CG	S	S	s1b-m2a	Yes	A2	Yes
CC28 (S. Africa)	U	U	U	s1a/s1b	Yes	A2	No
CC35 (S. Africa)	U	U	U	U	U	A1	No
CC33 (S. Africa)	U	U	U	U	U	A2	Yes
HK9728 (Hong Kong)	U	U	U	s1c-m1	Yes	A1	No
HPK28 (Japan)	U	U	U	s1c-m1	No	A1	No
HPK76 (Japan)	U	U	U	s1c-m1	No	A2	No
HPK154 (East Asia)	U	U	U	s2-m2a	No	A1	No

* Unless otherwise stated, all strains came from Swiss Italian patients. The strains are labelled A or C, according to their origin, antrum or corpus, respectively.

† CG, chronic gastritis; D, dyspepsia; DU, duodenal ulcer; GU, gastric ulcer; MALT, gastric MALT lymphoma.

‡ CLA, clarithromycin; MTZ, metronidazole; R, resistant; S, susceptible; U, unknown.

factors or markers associated with different clinical manifestations of the infection. This led to the discovery of the *VacA* cytotoxin (which induces vacuolation in eukaryotic cells), the high-molecular-mass antigen *CagA* and the *iceA* gene, which is induced by contact with the epithelium (Censini *et al.*, 1996; Cover *et al.*, 1994; Peek *et al.*, 1998). The *vacA* gene exhibits various signal sequence types (e.g. s1a, s1b, s1c and s2) and middle region types (e.g. m1a, m1b, m1c, m2a and m2b) (Atherton *et al.*, 1995; Pan *et al.*, 1998; Mukhopadhyay *et al.*, 2000; Van Doorn *et al.*, 1998a, c). The *cagA* gene is a marker for the presence of a pathogenicity island that may be present or absent (Censini *et al.*, 1996; Van der Ende *et al.*, 1998; Van Doorn *et al.*, 1998a, c). The *iceA* gene has two allelic forms, either *iceA1* or *iceA2* (Van Doorn *et al.*, 1998c; Figueiredo *et al.*, 2000). Furthermore, *H. pylori* strains may also be phenotypically distinguished according to their resistance to metronidazole, clarithromycin and amoxycillin (the three antibiotics used in different combinations in anti-*H. pylori* therapy). Finally, another factor differentiating *H. pylori* strains is the presence of the transposon-like IS605 element. Because of their role in promoting DNA rearrangements, transposable elements in *H. pylori* may

explain part of the diversity in the genome organization encountered in this species (Hook-Nikanne *et al.*, 1998).

In many populations of bacterial pathogens, particular clones are responsible for severe syndromes or for epidemic outbreaks (Musser, 1996; Piffaretti *et al.*, 1989; Selander *et al.*, 1986). Among the different techniques used, DNA sequencing of appropriate targets in the genome is the most powerful tool to discriminate between different strains or species. Automated DNA sequencing of housekeeping genes has extended the use of these techniques to genotyping and phylogenetic studies (Busse *et al.*, 1996; Maiden *et al.*, 1998). For instance, genes such as *atpD* (Christensen & Olsen, 1998), *recA* (Eisen, 1995), *hbb* (Valsangiacomo *et al.*, 1997) and *glnA* (Kumada *et al.*, 1993), have been widely used for population genetic studies. The population structure of *H. pylori* has been investigated with MLEE (Go *et al.*, 1996) and with other methods based on DNA sequence analysis (Achtman *et al.*, 1999; Akopyanz *et al.*, 1992a, b; Forbes *et al.*, 1995; Gibson *et al.*, 1998; Salaun *et al.*, 1998; Suerbaum *et al.*, 1998; Tee *et al.*, 1992; Van Doorn *et al.*, 1999a, b; Yamaoka *et al.*, 1998), and the results suggested a panmictic nature for *H.*

pylori. However, apart from MLEE, most of these methods considered the genomic diversity in pathogenicity-associated genes, e.g. *vacA* and *cagA* (Salaun *et al.*, 1998; Suerbaum *et al.*, 1998; Van Doorn *et al.*, 1999a, b; Yamaoka *et al.*, 1998), or in colonization factors such as the urease genes (Akopyanz *et al.*, 1992a; Salaun *et al.*, 1998) and the *flaA* and *flaB* genes (Forbes *et al.*, 1995; Salaun *et al.*, 1998; Suerbaum *et al.*, 1998), while other studies analysed a limited number of strains (Achtman *et al.*, 1999; Gibson *et al.*, 1998; Salaun *et al.*, 1998). In the present study, the population structure of *H. pylori* has been investigated by comparative sequence analysis of four housekeeping genes (*atpD*, *scoB*, *glnA* and *recA*) and on a relatively large number of isolates.

With the aim of better characterizing our population of *H. pylori*, we also determined the *vacA* and the *iceA* status, and the presence of *cagA* and IS605. In addition, we tested all strains for their susceptibility to metronidazole, clarithromycin and amoxicillin.

Most of the strains used in this study (71/78) were isolated from patients living in the southern part of Switzerland with various gastropathologies. The prevalence of infection in this area, together with the northern part of Italy, is the highest found in Europe (Doglioni *et al.*, 1992; EUROGAST Study Group, 1993). These regions are also of particular interest because of a significant incidence of gastric MALT lymphoma and gastric cancer. Finally, we looked at the geographical distribution of *H. pylori* genotypes that has been recently discovered (Achtman *et al.*, 1999; Campbell *et al.*, 1997; Miehle *et al.*, 1996; Van der Ende *et al.*, 1998; Van Doorn *et al.*, 1999a, b). For this reason, three strains from South Africa and four from East Asia were also included in this study.

METHODS

Patients. Gastric biopsies from 356 patients living in the southern part of Switzerland were tested for *H. pylori* by culture. For each patient, gastroenterologists were asked to complete a questionnaire on the gastroduodenal disease, age, sex, family gastric pathology history and previous treatment of *H. pylori* infections. The endoscopic diagnosis of the disease was confirmed by histopathological examination of the gastric biopsy performed at the Istituto Cantonale di Patologia (Locarno, Switzerland). *H. pylori* strains could be cultured from 178 (50%) patients. Of these 178 isolates, 71 were randomly selected for the phylogenetic analysis, including strains from 41 (57.7%) patients with chronic gastritis, 4 (5.6%) with dyspepsia, 20 (28.8%) with duodenal ulcer, 4 (5.6%) with gastric ulcer and 2 (1.4%) with gastric MALT lymphoma. The Ethical Committee of the Cantone Ticino approved this work and informed consent was obtained from all patients.

***H. pylori* strains.** Biopsy specimens were collected into Portagerm medium (BioMérieux) and processed in our laboratory within 4 h of gastroduodenoscopy. *H. pylori* strains were isolated by streaking gastric biopsies (either from the antrum or from the corpus) onto Brain Heart Infusion Agar (Oxoid) supplemented with 5% sheep blood and Vitox (Oxoid) and Columbia Agar supplemented with 5% sheep

blood and Skirrow's supplement (Oxoid). Plates were incubated at 37 °C in 5% O₂/10% CO₂/85% N₂ for up to 7 d. Isolates with typical colony morphology, Gram stain and biochemical tests positive for urease, catalase and oxidase were harvested in 25% peptone-glycerol and stored at -70 °C. A total of 142 strains were collected and were tested for antibiotic susceptibility, the results of which have been published elsewhere (Maggi-Solcà *et al.*, 2000). Of these strains, 71 were randomly chosen for the phylogenetic analysis. S. Suerbaum (Würzburg, Germany) and M. J. Blaser (Nashville, TN, USA) kindly provided DNA of 3 South African strains and DNA of 4 East Asian strains, respectively. The characteristics of the strains are reported in Table 1.

Choice of the four housekeeping genes. We considered the following housekeeping genes: *atpD*, *scoB*, *glnA* and *recA*. The beta subunit of the F₁-F₀ ATPase, encoded in *H. pylori* by the *atpD* gene, is one of the five subunits of the F₁ catalytic portion of the enzyme, and it has the most conserved primary structure of all the F₁-F₀ subunits (McGowan *et al.*, 1997). This gene has already been used to analyse bacterial phylogeny, generating trees congruent with those obtained with the 16S rDNA (Amann *et al.*, 1988; Christensen & Olsen, 1998). The succinyl-CoA:acetoacetate CoA-transferase (SCOT)-encoding gene (*sco*) of *H. pylori* was discovered by Corthésy-Theulaz *et al.* (1997). In contrast to the monomeric eukaryotic homologues, the *H. pylori* enzyme consists of two subunits, A and B. Both subunits were present in all strains tested. For the phylogenetic analysis we chose *scoB*, encoding the subunit B. The glutamine-synthetase-encoding *glnA* gene is well-conserved (Garner *et al.*, 1998) and has also been employed for phylogenetic studies (Kumada *et al.*, 1993). The RecA protein is involved in many essential biological processes, like homologous recombination, recombinational DNA repair, and activation of the SOS system. The primary sequence of *recA* is very well conserved among eubacterial species and it is also generally considered as an appropriate target for phylogenetic studies (Eisen, 1995).

DNA preparation, PCR and sequencing. *H. pylori* strains were subcultured for 3 d on fresh Brain Heart Infusion Agar supplemented with 5% sheep blood and Vitox. The cells were collected from the plates and DNA extraction and purification were performed in a single step using a commercial ion-exchange resin (InstaGene matrix; Bio-Rad), according to the manufacturer's instructions. Specific primers (Table 2) were used to amplify the four target genes (i.e. *atpD*, *scoB*, *glnA* and *recA*). A 2 µl portion of DNA extract was used for the PCR in a total reaction volume of 50 µl. The reaction mixture contained 5 µl PCR buffer (Roche Molecular Biochemicals), each deoxynucleoside triphosphate at a concentration of 200 µM, the appropriate primers, each at a concentration of 0.5 µM, and 1 U *Taq* DNA polymerase (Roche Molecular Biochemicals). The thermal profile used for the amplification of *atpD* was 2 min at 94 °C, followed by 35 cycles consisting of 94 °C for 1 min, 53 °C for 1 min and 72 °C for 1.5 min; for the amplification of the *scoB* gene fragment the annealing temperature was 50 °C, and for both *glnA* and *recA* the annealing temperature was 52 °C. The PCR amplicons were used for cycle sequencing after purification with the Qiaquick PCR purification Kit (Qiagen). Cycle sequencing reactions were performed with the dRhodamine Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer Biosystems) and with an automated DNA sequencer (ABI 310, Perkin Elmer). For the fragments with nucleotide substitutions resulting in an amino acid change, the DNA sequence of both strands was determined. The sequencing primers are listed in Table 2.

Table 2. Primers for PCR and sequencing experiments

Gene	Primer	Sequence (5'–3')
<i>atpD</i>	Hp-atpD-L*	AGGT [†] TTTAGGCCCGGTGG
	Hp-atpD-R* [†]	TAAGCCCTTTTCATCACGG
	Hp-atpD-L2 [†]	GTGCGAGCGATTGCTATGG
<i>scoB</i>	Hp-scoB-47a* [†]	ATGAGAGAGGCTAATCATTA [†] AAAAG
	Hp-scoB-48a*	CGAATTTGAAGTGCACCTATA
	Hp-scoB-48b*	GCTTCTGTTTTTCTCTCAC
<i>glnA</i>	Hp-glnA-L* [†]	CAAAACCGCCACTTTTCATGCC
	Hp-glnA-R*	GGATAAATTGTTTCGCTAAAGAC
	Hp-glnA-L2* [†]	ATATGTGGTTAAAATGGTCG
	Hp-glnA-R2*	GAATTTAAGAGACTGATAGGC
<i>recA</i>	Hp-recA-L* [†]	TCATTGAAATTTATGGGCCAG
	Hp-recA-R*	AAACATGATGTCAAATTCGG
	Hp-recA-R2*	GCTTCTCTAAAGGGCGGAG

* Primers used for PCR.

[†] Primers used for sequencing reactions.**Table 3.** Characteristics of the fragments sequenced

Gene	Accession no.	Gene length (bp)	Coordinates of sequenced fragment	Length of sequenced fragment (bp)
<i>atpD</i>	AF004014	1407	231–675	444
<i>scoB</i>	AJ000086	621	30–588	558
<i>glnA</i>	AF053357	1446	825–1035	510
<i>recA</i>	Z35478	1044	234–756	522

Table 4. DNA polymorphism in the four housekeeping genes analysed

Gene	No. of polymorphic sites (%)	No. of non-silent substitutions (%)	No. of substitutions in 3rd codon position (%)	% nucleotide similarity (range)	Mean % K_A^*	Mean % K_S^*	Ratio K_S/K_A	<i>H</i> ratio*
<i>atpD</i>	88 (19.8)	11 (7.4)	74 (84.1)	96.4 (93–100)	0.2	16.5	82.5	0.742
<i>scoB</i>	119 (21.3)	20 (10.7)	96 (80.7)	96.7 (94.1–99.8)	0.4	15.1	37.7	0.799
<i>glnA</i>	121 (23.7)	23 (13.5)	94 (77.7)	95.5 (91.8–99.6)	0.9	19.8	22	0.775
<i>recA</i>	106 (20.3)	10 (5.7)	88 (83)	95.3 (92.1–100)	1.0	20.3	20.3	0.778

* K_A , percentage of the mean differences between pairs of strains at nonsynonymous nucleotide positions; K_S , percentage of the mean differences between pairs of strains at synonymous nucleotide positions; the homoplasy ratio (*H* ratio) is a number whose expected value is 0 if the population is clonal and 1.0 if it is in linkage equilibrium.

Analysis of the nucleotide sequences. The sequences of the four housekeeping genes, *atpD*, *scoB*, *glnA* and *recA*, were handled and stored with the help of the Lasergene program EditSeq (1994 release; DNASTar) and aligned using Megalign (1994 release; DNASTar). Detailed information on the sequenced genes is reported in Table 3. The sequences were analysed using the neighbour-joining method (Saitou & Nei,

1987), with Kimura two-parameter distances as implemented in MEGA (Molecular Evolutionary Genetics Analysis 1.01, Kumar *et al.*, 1993). The reliability of internal branches was assessed by bootstrapping (Felsenstein, 1988), with 500 pseudo-replicates. Percentages of the mean differences between pairs of strains at synonymous nucleotide positions (K_S) and nonsynonymous positions (K_A) were calculated with

Fig. 1. Neighbour-joining tree with Kimura two-parameter distances based on *atpD* sequences. Bootstrap values from 500 pseudo-replicates are shown when they exceeded 30 %. The strains are labelled with a or c, according to their origin, antrum or corpus, respectively. The clinical manifestation, the *vacA* type, the presence of *cagA* and *IS605* are shown in

DNASP 3.0 (Rozas & Rozas, 1999) using Jukes & Cantor parameters. The homoplasy test (Maynard Smith & Smith, 1998) was performed using the HOMOPHY program (Suerbaum *et al.*, 1998) to test the importance of recombination (Achtman *et al.*, 1999).

***vacA*, *cagA* and *iceA* genotyping.** Typing of the *vacA* signal sequence and middle region of the *iceA* allele and detection of the *cagA* gene of the Swiss Italian strains were performed with a single-step procedure based on PCR and reverse hybridization (LiPA) as described by Van Doorn *et al.* (1998b).

IS605 detection. All 78 strains were screened for the presence of IS605 by PCR amplification of the two open reading frames (A and B) using published primers (Hook-Nikanne *et al.*, 1998). The PCR program was 2 min at 94 °C, followed by 35 cycles consisting of 94 °C for 1 min, 52 °C for 1 min and 72 °C for 1.5 min. Reaction conditions were similar to those described above.

Statistical analysis. Data were analysed by the chi-squared test.

RESULTS

Nucleotide sequence analysis

The nucleotide sequences of the fragments originating from the *atpD*, *scoB*, *glnA* and *recA* genes were determined for 78 *H. pylori* isolates of our collection (see Methods and Table 1). Among these sequences, 73 allelic variants were found for *atpD*, 78 for *scoB*, 78 for *glnA*, and 77 for *recA*. No deletions or insertions were found in any of the analysed gene fragments. The number of polymorphic sites (Table 4) was similar for all the four gene fragments (range 19.8–23.7%) and the substitutions were randomly distributed over the DNA sequences. The percentages of the mean differences between pairs of strains at synonymous nucleotide positions (K_S) varied from 15.1% to 20.3%, and from 0.2% to 1% at nonsynonymous positions (K_A). Most of the nucleotide replacements (77.7%–84.1%) were at the third codon position. According to the amino acid substitution frequencies, *atpD* and *recA* were the most conserved genes. In *atpD*, the region encoding the nucleotide-binding site (nucleotide positions 453–465) was highly conserved (data not shown). The amino acid substitutions in *recA* were mostly concentrated in a specific region arbitrarily designated as 'group II defining region' (see below). For *scoB* and *glnA*, there were no particular regions with a higher level of mutations. The sets of sequences were tested by the homoplasy test, which measures the importance of recombination (Maynard Smith & Smith, 1998). The resulting *H* ratios (Table 4) ranged from 0.742 to 0.799, indicating frequent recombination for all four gene fragments examined (*H* is a number whose expected

value is 0 if the population is clonal and 1.0 if it is in linkage equilibrium).

Phylogenetic analyses

Phylogenetic trees constructed from sequence analyses of the four housekeeping genes, *atpD*, *scoB*, *glnA* and *recA*, of the 78 *H. pylori* strains are presented in Figs 1, 2, 3 and 4. Comparison of these phylogenetic trees provided further evidence for the high genomic diversity found at the nucleotide level. The topologies of the trees are very different and most of the internal nodes are supported by low bootstrap values. In general, no significant clustering of strains was found for the *atpD*, *scoB* and *glnA* genes (Figs 1, 2 and 3). In contrast, the tree generated with the *recA* gene was clearly divided into two groups (designated I and II) (Fig. 4).

Cluster analysis of strains from other countries

The distribution of the East Asian and of the South African strains was variable and depended on the molecular marker used (*atpD*, *scoB*, *glnA* or *recA* gene). The three East Asian strains (HK9728, HPK28, HPK76) with the s1c *vacA* genotype, a newly discovered *vacA* type which is mainly found in the Asian population (Van Doorn *et al.*, 1999a, b) were always grouped together in the four trees (Figs 1, 2, 3 and 4). However, only in the *scoB* and *glnA* trees (Figs 2 and 3) did they form a well-supported monophyletic group (bootstrap values 68% and 97%, respectively). In contrast, in the *atpD* (Fig. 1) and *recA* (Fig. 4) trees, they were paraphyletic with the inclusion of other strains, namely strain 111A for *atpD* and 125A2 for *recA*. Both strains had been isolated from Swiss Italian patients and were *vacA* type s1b and s2, respectively. The East Asian strain HPK154, which is *vacA* s2-m2, never clustered with the s1c Asian strains in any tree. The South African strains (CC28, CC33 and CC35) formed a well-supported cluster (75% bootstrap value) only in the *atpD* tree (Fig. 1). For the *scoB* gene (Fig. 2), the three South African strains were distributed throughout the tree, whereas in the *glnA* tree (Fig. 3), only strains CC33 and CC35 clustered together (bootstrap value 100%). In the *recA* tree (Fig. 4), even though all the South African strains belonged to 'group II', only strains CC35 and CC28 clustered together (bootstrap value 88%).

vacA, *cagA* and *iceA* genotyping

This analysis was only performed for the 71 Swiss Italian strains (Table 1). The *vacA* genotype and the presence of *cagA* for each strain are also indicated in parentheses beside the taxon name in the four trees (Figs

parentheses. CG, chronic gastritis; GU, gastric ulcer; DU duodenal ulcer; D, dyspepsia; MALT, gastric MALT lymphoma. Strains for which the *cagA* and IS605 status is not mentioned lack these sequences. East Asian strains with the s1c-m1 *vacA* genotype are marked by black circles and that with the s2-m2a *vacA* genotype by a black triangle; South African strains are indicated by black squares.

Fig. 2. Neighbour-joining tree with Kimura two-parameter distances based on *scoB* sequences. See Fig. 1 for more details.

1, 2, 3 and 4). The *vacA* s1 type was slightly more common than s2 (54.9% vs 45%), with the s1a and s1b subtypes nearly equally distributed (28.2% vs 26.8%). Among the *vacA* middle region types, m2a was more prevalent than m1 (69% vs 31%). The s1c and the m2b alleles were not found in our collection of Swiss Italian strains. The following *vacA* type combinations were detected: s1a-m1 (17%), s1a-m2a (11%), s1b-m1 (14%), s1b-m2a (13%) and s2-m2a (45%). The combination s2-m1 was not found. The *cagA* gene was present in 50% of the strains and the *iceA1* allele was found more frequently than *iceA2* (59% vs 41%).

There were no significant associations between the *vacA*, *cagA*, *iceA* status and particular clinical outcomes of the infection (Table 5). Apart from *vacA* s1, which was strongly associated with the presence of *cagA* ($P < 0.001$), there was no correlation between any of the other virulence-associated markers.

IS605

IS605 ORFA and ORFB sequences were identified in the following 12 strains: 86C, 348C, 48A, 47A, 65A, 128A, 99A2, 84A, 192A, 163C, 18A, 35C (Table 1). However, in strain 322A, only the ORFA sequence and in strain 103A, only the ORFB sequence could be detected. According to Hook-Nikanne *et al.* (1998), we considered these two strains as positive for IS605. The presence of IS605 was not correlated with specific clinical manifestations, but was associated with *vacA* s1 ($P = 0.009$) and with the presence of *cagA* ($P = 0.02$), as reported earlier (Hook-Nikanne *et al.*, 1998).

Antibiotic resistance

Antibiotic susceptibility was tested for the 71 Swiss Italian strains (Table 1). No association was found between antibiotic resistance and other genetic markers investigated in this study.

DISCUSSION

DNA polymorphism and recombination

The amount of polymorphic sites we found in the genes analysed was higher in *H. pylori* (range 19.8–23.7%) than in other pathogenic bacteria (e.g. 0.01% in *Mycobacterium tuberculosis* complex, 6.18% in *Neisseria meningitidis*, 11.77% in *Escherichia coli*) (Sreevatsan *et al.*, 1997). However, the percentage of DNA polymorphism was similar to those found by Suerbaum *et al.* (1998) for *vacA*, *flaA* and *flaB* sequences of *H. pylori*. The number of allelic variants was very high: nearly every isolate contained a unique gene sequence (Table 4). The mean K_S values (measuring the rate of synonymous changes) were higher than the K_A values (non-synonymous changes), supporting previous studies (Alm *et al.*, 1999; Suerbaum *et al.*, 1998).

Moreover, the K_S/K_A values were in the same range as those found for most of the housekeeping genes analysed by Achtman *et al.* (1999). The K_S/K_A ratio was lower in virulence-associated genes (i.e. *vacA* and *cagA*) than in sequences encoding housekeeping enzymes (Achtman *et al.*, 1999; Suerbaum *et al.*, 1998), i.e. virulence genes presented a higher percentage of nonsynonymous changes. The importance of recombination was estimated by calculating the H ratios (Table 4), using the homoplasmy test. The values obtained were similar to those previously reported for other genes (Achtman *et al.*, 1999; Suerbaum *et al.*, 1998) and clearly indicated frequent recombination. The homoplasmy test is particularly appropriate when the sequences are rather similar, differing by 1–5% of nucleotides (Maynard Smith & Smith, 1998). However, this test has been successfully used also for sequences with much higher polymorphisms (Achtman *et al.*, 1999; Suerbaum *et al.*, 1998), as is the case for *H. pylori*.

The question why the genome of *H. pylori* has so much polymorphism still has to be clarified. Apparently *H. pylori* does not have a full DNA repair system (Tomb *et al.*, 1997); this, together with a high recombination frequency, could offer a possible explanation. Further studies on the *H. pylori* repair system are required to understand this observation better.

Topology of the trees

In general, phylogenetic analyses of the four housekeeping genes revealed no significant clustering of strains, and most of the internal nodes in the four trees were characterized by extremely low bootstrap values. The observation that the grouping of the 78 strains, and their genetic distances, were clearly different in the four dendrograms (Figs 1, 2, 3 and 4) again suggests the existence of frequent recombination events, thus supporting the hypothesis of a panmictic structure of the *H. pylori* population.

Only the *recA*-based tree showed a subdivision into two groups, indicating the existence of two distinct allelic variants of this gene (Fig. 4). The two groups resulted from specific mutations at positions 309 (C → T; His → Tyr), 363–365 (GAT → AGC; Asp → Ser), 368 (A → G, silent) and 374–375 (GC → AG; Gln → Glu). These mutations were linked and defined these two allelic forms of *recA*. The existence of those groups is supported by the finding that sequences belonging to groups I and II had already been previously independently reported by Schmitt *et al.* (1995) and Thompson & Blaser (1995), respectively.

Clustering according to geographical origin

The analyses of the four housekeeping genes were unable to resolve the geographical relationships among the *H. pylori* strains examined, with the exception of the East

Fig. 3. Neighbour-joining tree with Kimura two-parameter distances based on *glnA* sequences. See Fig. 1 for more details.

Fig. 4. Neighbour-joining tree with Kimura two-parameter distances based on *recA* sequences. See Fig. 1 for more details.

Table 5. Relationship of the different markers with the type of disease

Disease	<i>vacA</i>		<i>cagA</i> ⁺	<i>iceA</i>		IS605 ⁺
	s1	s2		A1	A2	
Ulcer (<i>n</i> = 26)	17	9	12	13	13	4
Non-ulcer (<i>n</i> = 45)	22	23	24	29	16	10
Total (<i>n</i> = 71)	39 (55 %)	32 (45 %)	36 (50.7 %)	42 (59.1 %)	29 (40.9 %)	14 (19.7 %)

Asian strains in the *scoB* and *glnA* trees (Figs 2 and 3). Recently, Van Doorn *et al.* (1999a) investigated the worldwide distribution of the *vacA* alleles and found a gradient: in Northern Europe the s1a genotype prevailed, whereas in France and Italy the s1a and s1b genotypes were nearly equally present and in Spain and in Portugal the s1b type was highly prevalent. The Swiss Italian strains apparently show the same distribution of *vacA* s-types as in France and Italy. The East Asian and the South African populations show particular features. In East Asia nearly all the *H. pylori* isolates are *cagA*⁺ and *vacA* s1c (Maeda *et al.*, 1998; Van Doorn *et al.*, 1999a). These genotypes have been associated with ulcer disease and, in fact, the incidence of ulcer and gastric cancer in East Asia is the highest in the world (Maeda *et al.*, 1998). In South Africa mostly the s1b *vacA* type is found (Letley *et al.*, 1999) and, in spite of a high prevalence of *H. pylori* infection, the incidence of gastric cancer is low (this is called the 'African enigma'; Holcombe, 1992). Furthermore, Suerbaum *et al.* (1998) reported a more conserved *H. pylori* population among South Africans. Concerning the South African strains analysed in the present study, only the *vacA* type of strain CC28 was known and it was unusual because it contained a hybrid of *vacA* s1a and s1b. This strain was previously included in a study by Achtman *et al.* (1999) dealing with the population genetics and geographical diversity of 20 *H. pylori* strains from different parts of the world. In this study, strain CC28 belonged to the weakly clonal group called 'clone 2', which included also one strain from Gambia, one strain from the USA and one strain from Guatemala.

vacA, *cagA*, *iceA*

Previous studies have indicated that the presence of the *vacA* s1, *cagA*⁺ and *iceA*1 genotypes is associated with a severe manifestation of the infection (Blaser *et al.*, 1995; Peek *et al.*, 1998; Van Doorn *et al.*, 1998a, c). We could not confirm any of these associations (Table 5). In addition, other authors recently failed to find a linkage between specific *vacA* and *cagA* genotypes and the severity of the disease (Go & Graham, 1996; Go *et al.*, 1998; Maeda *et al.*, 1998; Yamaoka *et al.*, 1999). Two factors could have influenced the results. One is the problem of the reliability of the clinical data: for

instance, it cannot be excluded that a patient with gastritis at the time of endoscopy has experienced ulcer disease in the past. The second factor is related to the fact that we analysed only one strain from each patient, although colonization with multiple strains is possible (Taylor *et al.*, 1995). Additional analyses of a collection of isolates originating from other diseased people should clarify the existence of this association between particular traits and virulence.

IS605 and antibiotic resistance

The presence of IS elements is usually associated with genome rearrangements. Specific genome rearrangements may provide a selective advantage to some strains. Nevertheless, the presence of IS605 was not correlated with specific strain clusters on the dendrograms (Figs 1, 2, 3 and 4). The antibiotic-resistant phenotype, which also results from a selective process, was also not associated with particular groups of strains.

Conclusions

Various authors (Achtman *et al.*, 1999; Go *et al.*, 1996; Salaun *et al.*, 1998; Suerbaum *et al.*, 1998) have suggested the non-clonal nature of *H. pylori*. Our study, based on the sequences of four housekeeping genes and involving a considerable number of strains originating from a limited geographical region (South Switzerland), confirms and reinforces this finding. The hypothesis of a recombining structure in our *H. pylori* population is based on the following points: (i) the level of DNA polymorphism found was high; (ii) the topology of the trees generated from four housekeeping genes was different for each genetic marker; (iii) the distribution of the strains according to their geographical origin was different in the four trees; (iv) no association between the distribution of the strains on the dendrograms and any of the characteristics of the strains (clinical manifestation, virulence markers, antibiotic resistance and IS605) was found; (v) finally, the *H* ratios measuring the importance of recombination (homoplasmy test) ranged from 0.742 to 0.799. The extensive recombination structure can be partially explained by the natural competence of *H. pylori* (Alm *et al.*, 1999; Hofreuter *et al.*, 1998). In a recombinant population, clonal groups

are difficult to identify. It is therefore surprising that distribution of *H. pylori* strains according to the ethnic origin of the host (even though not homogeneous) could still be recognized (Achtman *et al.*, 1999; Campbell *et al.*, 1997; Suerbaum *et al.*, 1998; Van Doorn *et al.*, 1999a, b). It is likely that, besides frequent recombination events, host characteristics as well as environmental conditions might influence the selection process, leading for instance to parallel or convergent evolution. Geographical clusters might also be in part preserved by geographical barriers to some extent hindering global, worldwide recombination among strains. The *H. pylori* population structure deserves further investigations, for instance by analysing more strains from different geographical origins and by collecting more information on the host's predisposition.

ACKNOWLEDGEMENTS

We are grateful to all the gastroenterologists who participated in this work: V. Saglini, B. Miazza, P. Wunderlich, I. Fumagalli and A. Bonetti. We thank also E. Pedrinis of the Istituto Cantonale di Patologia in Locarno for the histological examinations. For providing the DNA of the East Asian strains and of the South African strains, we are grateful to M. J. Blaser (Nashville, TN, USA) and to S. Suerbaum (Würzburg, Germany), respectively. For technical assistance we thank R. Sanna and A. Verschuuren (Delft, The Netherlands). Many thanks also to E. Roggero and E. Zucca for helpful discussions and to P. Kuhnert (Bern, Switzerland) for reading the manuscript. We also thank M. Achtman (Berlin, Germany) for providing the HOMOPLASMY program. This research was supported by grant 31-45914.95 from the Swiss National Science Foundation, by the Helmut Horten Foundation, and by Astra Pharmaceutica (Dietikon, Switzerland).

REFERENCES

- Achtman, M., Azuma, T., Berg, D. E. & 7 other authors (1999). Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol Microbiol* **32**, 459–470.
- Akopyanz, N., Bukanov, N. O., Westblom, T. U. & Berg, D. E. (1992a). PCR-based RFLP analysis of DNA sequence diversity in the gastric pathogen *Helicobacter pylori*. *Nucleic Acids Res* **20**, 6221–6225.
- Akopyanz, N., Bukanov, N. O., Westblom, T. U., Kresovich, S. & Berg, D. E. (1992b). DNA diversity among clinical isolates of *Helicobacter pylori* detected by PCR-based RAPD fingerprinting. *Nucleic Acids Res* **20**, 5137–5142.
- Alm, R. A., Ling, L. S., Moir, D. T. & 20 other authors (1999). Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **397**, 176–180.
- Amann, R., Ludwig, W. & Schleifer, K. H. (1988). Beta-subunit of ATP-synthase: a useful marker for studying the phylogenetic relationship of eubacteria. *J Gen Microbiol* **134**, 2815–2821.
- Atherton, J. C., Cao, P., Peek, R. M., Jr, Tummuru, M. K., Blaser, M. J. & Cover, T. L. (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.
- Blaser, M. J. (1997). Ecology of *Helicobacter pylori* in the human stomach. *J Clin Invest* **100**, 759–762.
- Blaser, M. J., Perez-Perez, G. I., Kleanthous, H., Cover, T. L., Peek, R. M., Chyou, P. H., Stemmermann, G. N. & Nomura, A. (1995). Infection with *Helicobacter pylori* strains possessing *cagA* is associated with an increased risk of developing adenocarcinoma of the stomach. *Cancer Res* **55**, 2111–2115.
- Busse, H. J., Denner, E. B. & Lubitz, W. (1996). Classification and identification of bacteria: current approaches to an old problem. Overview of methods used in bacterial systematics. *J Biotechnol* **47**, 3–38.
- Campbell, S., Fraser, A., Holliss, B., Schmid, J. & O'Toole, P. W. (1997). Evidence for ethnic tropism of *Helicobacter pylori*. *Infect Immun* **65**, 3708–3712.
- Censini, S., Lange, C., Xiang, Z., Crabtree, J. E., Ghiara, P., Borodovsky, M., Rappuoli, R. & Covacci, A. (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.
- Christensen, H. & Olsen, J. E. (1998). Phylogenetic relationships of *Salmonella* based on DNA sequence comparison of *atpD* encoding the beta subunit of ATP synthase. *FEMS Microbiol Lett* **161**, 89–96.
- Corthésy-Theulaz, I. E., Bergonzelli, G. E., Henry, H., Bachmann, D., Schorderet, D. F., Blum, A. L. & Ornston, L. N. (1997). Cloning and characterization of *Helicobacter pylori* succinyl CoA:acetoacetate CoA-transferase, a novel prokaryotic member of the CoA-transferase family. *J Biol Chem* **272**, 25659–25667.
- Cover, T. L., Tummuru, M. K., Cao, P., Thompson, S. A. & Blaser, M. J. (1994). Divergence of genetic sequences for the vacuolating cytotoxin among *Helicobacter pylori* strains. *J Biol Chem* **269**, 10566–10573.
- Dogliani, C., Wotherspoon, A. C., Moschini, A., de Boni, M. & Isaacson, P. G. (1992). High incidence of primary gastric lymphoma in northeastern Italy. *Lancet* **339**, 834–835.
- Eisen, J. A. (1995). The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. *J Mol Evol* **41**, 1105–1123.
- EUROGAST Study Group (1993). An international association between *Helicobacter pylori* infection and gastric cancer. *Lancet* **341**, 1359–1362.
- Felsenstein, J. (1988). Phylogenies from molecular sequences: inference and reliability. *Annu Rev Genet* **22**, 521–565.
- Figueiredo, C., Quint, W. G., Sanna, R. & 7 other authors (2000). Genetic organization and heterogeneity of the *iceA* locus of *Helicobacter pylori*. *Gene* **246**, 59–68.
- Forbes, K. J., Fang, Z. & Pennington, T. H. (1995). Allelic variation in the *Helicobacter pylori* flagellin genes *flaA* and *flaB*: its consequences for strain typing schemes and population structure. *Epidemiol Infect* **114**, 257–266.
- Garner, R. M., Fulkerson, J., Jr & Mobley, H. L. (1998). *Helicobacter pylori* glutamine synthetase lacks features associated with transcriptional and posttranslational regulation. *Infect Immun* **66**, 1839–1847.
- Gibson, J. R., Slater, E., Xerry, J., Tompkins, D. S. & Owen, R. J. (1998). Use of an amplified-fragment length polymorphism technique to fingerprint and differentiate isolates of *Helicobacter pylori*. *J Clin Microbiol* **36**, 2580–2585.
- Go, M. F. & Graham, D. Y. (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.
- Go, M. F., Kapur, V., Graham, D. Y. & Musser, J. M. (1996).

- Population genetic analysis of *Helicobacter pylori* by multilocus enzyme electrophoresis: extensive allelic diversity and recombinational population structure. *J Bacteriol* 178, 3934–3938.
- Go, M. F., Cissell, L. & Graham, D. Y. (1998). Failure to confirm association of *vacA* gene mosaicism with duodenal ulcer disease. *Scand J Gastroenterol* 33, 132–136.
- Hofreuter, D., Odenbreit, S., Henke, G. & Haas, R. (1998). Natural competence for DNA transformation in *Helicobacter pylori*: identification and genetic characterization of the *comB* locus. *Mol Microbiol* 28, 1027–1038.
- Holcombe, C. (1992). *Helicobacter pylori*: the African enigma. *Gut* 33, 429–431.
- Hook-Nikanne, J., Berg, D. E., Peek, R. M., Jr, Kersulyte, D., Tummuru, M. K. & Blaser, M. J. (1998). DNA sequence conservation and diversity in transposable element IS605 of *Helicobacter pylori*. *Helicobacter* 3, 79–85.
- Kumada, Y., Benson, D. R., Hillemann, D., Hosted, T. J., Rochefort, D. A., Thompson, C. J., Wohlleben, W. & Tatenos, Y. (1993). Evolution of the glutamine synthetase gene, one of the oldest existing and functioning genes. *Proc Natl Acad Sci USA* 90, 3009–3013.
- Kumar, S., Tamura, K. & Nei, M. (1993). MEGA: Molecular Evolutionary Genetics Analysis, version 1.01. University Park, PA: Pennsylvania State University.
- Letley, D. P., Lastovica, A., Louw, J. A., Hawkey, C. J. & Atherton, J. C. (1999). Allelic diversity of the *Helicobacter pylori* vacuolating cytotoxin gene in South Africa: rarity of the *vacA* s1a genotype and natural occurrence of an s2/m1 allele. *J Clin Microbiol* 37, 1203–1205.
- McGowan, C. C., Cover, T. L. & Blaser, M. J. (1997). Analysis of F₁F₀-ATPase from *Helicobacter pylori*. *Infect Immun* 65, 2640–2657.
- Maeda, S., Ogura, K., Yoshida, H., Kanai, F., Ikenoue, T., Kato, N., Shiratori, Y. & Omata, M. (1998). Major virulence factors, VacA and CagA, are commonly positive in *Helicobacter pylori* isolates in Japan. *Gut* 42, 338–343.
- Maggi-Solcà, N., Valsangiacomo, C. & Piffaretti, J.-C. (2000). Prevalence of *Helicobacter pylori* resistant strains in the Southern part of Switzerland. *Clin Microbiol Infect* 6, 38–40.
- Maiden, M. C., Bygraves, J. A., Feil, E. & 10 other authors (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 95, 3140–3145.
- Maynard Smith, J. & Smith, N. H. (1998). Detecting recombination from gene trees. *Mol Biol Evol* 15, 590–599.
- Miehke, S., Kibler, K., Kim, J. G., Figura, N., Small, S. M., Graham, D. Y. & Go, M. F. (1996). Allelic variation in the *cagA* gene of *Helicobacter pylori* obtained from Korea compared to the United States. *Am J Gastroenterol* 91, 1322–1325.
- Mukhopadhyay, A. K., Kersulyte, D., Jeong, J. Y. & 9 other authors (2000). Distinctiveness of genotypes of *Helicobacter pylori* in Calcutta, India. *J Bacteriol* 182, 3219–3227.
- Musser, J. M. (1996). Molecular population genetic analyses of emerged bacterial pathogens: selected insights. *Emerg Infect Dis* 2, 1–17.
- Pan, Z. J., Berg, D. E., Van der Hulst, R. W., Su, W. W., Raudonikienė, A., Xiao, S. D., Dankert, J., Tytgat, G. N. & Van der Ende, A. (1998). Prevalence of vacuolating cytotoxin production and distribution of distinct *vacA* alleles in *Helicobacter pylori* from China. *J Infect Dis* 178, 220–226.
- Peek, R. M., Jr, Thompson, S. A., Donahue, J. P., Tham, K. T., Atherton, J. C., Blaser, M. J. & Miller, G. G. (1998). Adherence to gastric epithelial cells induces expression of a *Helicobacter pylori* gene, *iceA*, that is associated with clinical outcome. *Proc Assoc Am Physicians* 110, 531–544.
- Piffaretti, J.-C., Kressebuch, H., Aeschbacher, M., Bille, J., Bannerman, E., Musser, J. M., Selander, R. K. & Rocourt, J. (1989). Genetic characterization of the bacterium *Listeria monocytogenes* causing epidemic disease. *Proc Natl Acad Sci USA* 86, 3818–3822.
- Pounder, R. E. (1995). The prevalence of *Helicobacter pylori* in different countries. *Aliment. Pharmacol Ther* 9 (suppl 2), 33–44.
- Roggero, E., Zucca, E., Pinotti, G. & 7 other authors (1995). Eradication of *Helicobacter pylori* infection in primary low-grade gastric lymphoma of mucosa-associated lymphoid tissue. *Ann Intern Med* 122, 767–769.
- Rozas, J. & Rozas, R. (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15, 174–174.
- Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4, 406–425.
- Salaun, L., Audibert, C., Le Lay, G., Burucoa, C., Fauchere, J. L. & Picard, B. (1998). Panmictic structure of *Helicobacter pylori* demonstrated by the comparative study of six genetic markers. *FEMS Microbiol Lett* 161, 231–239.
- Schmitt, W., Odenbreit, S., Heuermann, D. & Haas, R. (1995). Cloning of the *Helicobacter pylori* *recA* gene and functional characterization of its product. *Mol Gen Genet* 248, 563–572.
- Selander, R. K., Caugant, D. A., Ochman, H., Musser, J. M., Gilmour, M. N. & Whittam, T. S. (1986). Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl Environ Microbiol* 51, 873–884.
- Sreevatsan, S., Pan, X., Stockbauer, K. E., Connell, N. D., Kreiswirth, B. N., Whittam, T. S. & Musser, J. M. (1997). Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci USA* 94, 9869–9874.
- Suerbaum, S., Smith, J. M., Bapumia, K., Morelli, G., Smith, N. H., Kunstmann, E., Dyrek, I. & Achtman, M. (1998). Free recombination within *Helicobacter pylori*. *Proc Natl Acad Sci USA* 95, 12619–12624.
- Taylor, N. S., Fox, J. G., Akopyants, N. S. & 7 other authors (1995). Long-term colonization with single and multiple strains of *Helicobacter pylori* assessed by DNA fingerprinting. *J Clin Microbiol* 33, 918–923.
- Tee, W., Lambert, J., Smallwood, R., Schembri, M., Ross, B. C. & Dwyer, B. (1992). Ribotyping of *Helicobacter pylori* from clinical specimens. *J Clin Microbiol* 30, 1562–1567.
- Thompson, S. A. & Blaser, M. J. (1995). Isolation of the *Helicobacter pylori* *recA* gene and involvement of the *recA* region in resistance to low pH. *Infect Immun* 63, 2185–2193.
- Tomb, J. F., White, O., Kerlavage, A. R. & 39 other authors (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388, 539–547.
- Valsangiacomo, C., Balmelli, T. & Piffaretti, J. C. (1997). A phylogenetic analysis of *Borrelia burgdorferi sensu lato* based on sequence information from the *hbb* gene, coding for a histone-like protein. *Int J Syst Bacteriol* 47, 1–10.
- Van der Ende, A., Pan, Z. J., Bart, A., van der Hulst, R. W., Feller, M., Xiao, S. D., Tytgat, G. N. & Dankert, J. (1998). *cagA*-positive *Helicobacter pylori* populations in China and The Netherlands are distinct. *Infect Immun* 66, 1822–1826.
- Van Doorn, L. J., Figueiredo, C., Sanna, R., Plaisier, A., Schneeberger, P., de Boer, W. & Quint, W. G. (1998a). Clinical

relevance of the *cagA*, *vacA*, and *iceA* status of *Helicobacter pylori*. *Gastroenterology* **115**, 58–66.

Van Doorn, L. J., Figueiredo, C., Sanna, R., Pena, S., Midolo, P., Ng, E. K., Atherton, J. C., Blaser, M. J. & Quint, W. G. (1998b). Expanding allelic diversity of *Helicobacter pylori vacA*. *J Clin Microbiol* **36**, 2597–2603.

Van Doorn, L. J., Figueiredo, C., Rossau, R., Jannes, G., van Asbroek, M., Sousa, J. C., Carneiro, F. & Quint, W. G. (1998c). Typing of *Helicobacter pylori vacA* gene and detection of *cagA* gene by PCR and reverse hybridization. *J Clin Microbiol* **36**, 1271–1276.

Van Doorn, L. J., Figueiredo, C., Mégraud, F. & 14 other authors (1999a). Geographic distribution of *vacA* allelic types of *Helicobacter pylori*. *Gastroenterology* **116**, 823–830.

Van Doorn, L. J., Figueiredo, C., Sanna, R., Blaser, M. J. & Quint, W. G. (1999b). Distinct variants of *Helicobacter pylori cagA* are associated with *vacA* subtypes. *J Clin Microbiol* **37**, 2306–2311.

Yamaoka, Y., Kodama, T., Kita, M., Imanishi, J., Kashima, K. & Graham, D. Y. (1998). Relationship of *vacA* genotypes of *Helicobacter pylori* to *cagA* status, cytotoxin production, and clinical outcome. *Helicobacter* **3**, 241–253.

Yamaoka, Y., Kodama, T., Gutierrez, O., Kim, J. G., Kashima, K. & Graham, D. Y. (1999). Relationship between *Helicobacter pylori iceA*, *cagA*, and *vacA* status and clinical outcome: studies in four different countries. *J Clin Microbiol* **37**, 2274–2279.

Received 14 July 2000; revised 25 January 2001; accepted 13 February 2001.