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,† Marco V. Bernasconi,† Claudio Valsangiacomo,† Leen-Jan Van Doorn‡ and Jean-Claude Piffaretti†

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*, 2 1 < 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 homoplasy 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

† Istituto Cantonale Batteriosierologico, Via Ospedale 6, 6904 Lugano, Switzerland
‡ Delft Diagnostic Laboratory, R. de Graafweg 7, 2625 AD Delft, The Netherlands

**INTRODUCTION**

*Helicobacter pylori* is a microaerophilic, Gram-negative, slow-growing, spiral-shaped and flagellate microorganism. 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...
### Table 1. *H. pylori* strains and their characteristics

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

Table 1 (cont.)

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

* 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; Mukohpadhyay 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 amoxicillin (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 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.
**Methods**

**Patients.** Gastric biopsies from 356 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; Miehke 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.

**H. pylori strains.** Biopsy specimens were collected into Portagerm medium (BioMerieux) 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 days. 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-Solca 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 phylogenetic relationships. Its sequence is also 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 Hearth 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 Tag 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

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer</th>
<th>Sequence (5'–3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>atpD</td>
<td>Hp-atpD-L*</td>
<td>AGGTTTTAGGCCCGGTGG</td>
</tr>
<tr>
<td></td>
<td>Hp-atpD-R†</td>
<td>TAAGCCCTTTTATCATCACGG</td>
</tr>
<tr>
<td></td>
<td>Hp-atpD-L2†</td>
<td>GTGCAGGCGATTCATATGG</td>
</tr>
<tr>
<td>scoB</td>
<td>Hp-scoB-47a†</td>
<td>ATGAGAGAGGCTATTATCTAAGAAG</td>
</tr>
<tr>
<td></td>
<td>Hp-scoB-48a*</td>
<td>CGAATTTGAGAATCACCTA</td>
</tr>
<tr>
<td></td>
<td>Hp-scoB-48b*</td>
<td>GCCTCCTTTTCTCTCAC</td>
</tr>
<tr>
<td>glnA</td>
<td>Hp-glnA-L*†</td>
<td>CAAAACCGCACTTTATGACG</td>
</tr>
<tr>
<td></td>
<td>Hp-glnA-R*</td>
<td>GGATTTGAGAATCACCTA</td>
</tr>
<tr>
<td></td>
<td>Hp-glnA-L2*†</td>
<td>ATATGTTGAAAAATGTCG</td>
</tr>
<tr>
<td></td>
<td>Hp-glnA-R2*</td>
<td>GAATTTGAGAATCTAGGAGG</td>
</tr>
<tr>
<td>recA</td>
<td>Hp-recA-L*†</td>
<td>TCATTGAAATTTATGGGCGAG</td>
</tr>
<tr>
<td></td>
<td>Hp-recA-R*</td>
<td>AACAAGATGTTCAATTCG</td>
</tr>
<tr>
<td></td>
<td>Hp-recA-R2*</td>
<td>GCTTCTCTAAAGGCGGAG</td>
</tr>
</tbody>
</table>

*Primers used for PCR.
†Primers used for sequencing reactions.

Table 3. Characteristics of the fragments sequenced

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

Table 4. DNA polymorphism in the four housekeeping genes analysed

<table>
<thead>
<tr>
<th>Gene</th>
<th>No. of polymorphic sites (%)</th>
<th>No. of non-silent substitutions (%)</th>
<th>No. of substitutions in 3rd codon position (%)</th>
<th>% nucleotide similarity (range)</th>
<th>Mean % K_s</th>
<th>Mean % K_a</th>
<th>K_s/K_a</th>
<th>Ratio</th>
<th>H ratio*</th>
</tr>
</thead>
<tbody>
<tr>
<td>atpD</td>
<td>88 (19-8)</td>
<td>11 (7-4)</td>
<td>74 (84-1)</td>
<td>96.4 (93-100)</td>
<td>0.2</td>
<td>16.5</td>
<td>82.5</td>
<td>0.742</td>
<td></td>
</tr>
<tr>
<td>scoB</td>
<td>119 (21-3)</td>
<td>20 (10-7)</td>
<td>96 (80-7)</td>
<td>96.7 (94.1-99.8)</td>
<td>0.4</td>
<td>15.1</td>
<td>37.7</td>
<td>0.799</td>
<td></td>
</tr>
<tr>
<td>glnA</td>
<td>121 (23-7)</td>
<td>23 (13-5)</td>
<td>94 (77-7)</td>
<td>95.5 (91.8-99.6)</td>
<td>0.9</td>
<td>19.8</td>
<td>22.0</td>
<td>0.775</td>
<td></td>
</tr>
<tr>
<td>recA</td>
<td>106 (20-3)</td>
<td>10 (5-7)</td>
<td>88 (83)</td>
<td>95.3 (92.1-100)</td>
<td>1.0</td>
<td>20.3</td>
<td>20.3</td>
<td>0.778</td>
<td></td>
</tr>
</tbody>
</table>

* K_s, percentage of the mean differences between pairs of strains at nonsynonymous nucleotide positions; K_a, 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 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; DNAnastar) and aligned using Megalign (1994 release; DNAnastar). 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
DNAseq 3.0 (Rozas & Rozas, 1999) using Jukes & Cantor parameters. The homoplasy test (Maynard Smith & Smith, 1998) was performed using the HOMOPLASY 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 (LiFA) as described by Van Doorn et al. (1999b).

**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 (Ks) varied from 15.1% to 20.3%, and from 0.2% to 1% at nonsynonymous positions (Ka). 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 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 1, 2, 3 and 4).
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 ORF A 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 ORF A 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_a/K_s values were in the same range as those found for most of the housekeeping genes analysed by Achtman et al. (1999). The K_a/K_s 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 homoplasy 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 homoplasy 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 define 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.
Population genetics of *H. pylori* in S. Switzerland

**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

<table>
<thead>
<tr>
<th>Disease</th>
<th>vacA</th>
<th>cagA⁺</th>
<th>iceA</th>
<th>IS605⁺</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>s1</td>
<td>s2</td>
<td>A1</td>
<td>A2</td>
</tr>
<tr>
<td>Ulcer (n = 26)</td>
<td>17</td>
<td>9</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>Non-ulcer (n = 45)</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>29</td>
</tr>
<tr>
<td>Total (n = 71)</td>
<td>39 (55 %)</td>
<td>32 (45%)</td>
<td>36 (50.7 %)</td>
<td>42 (59.1 %)</td>
</tr>
</tbody>
</table>

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 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 iceA1 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 (homoplasy 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 recombining 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. Mizzara, P. Wanderlich, L. 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 HOMOPLAST 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


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