Microbiology (2002), **148**, 3671–3680 Printed in Great Britain

Identification of strain-specific genes located outside the plasticity zone in nine clinical isolates of *Helicobacter pylori*

Grettel Chanto,[1] Alessandra Occhialini,[1] Nathalie Gras,[1] Richard A. Alm,[2] Francis Mégraud[1] and Armelle Marais[1]†

Author for correspondence: Francis Mégraud. Tel: +33 5 56 79 59 10. Fax: +33 5 56 79 60 18. e-mail: francis.megraud@chu-bordeaux.fr

[1] Laboratoire de Bactériologie, Université Victor Segalen Bordeaux 2 and Hôpital Pellegrin, Place Amélie Raba-Léon, 33076 Bordeaux Cedex, France

[2] AstraZeneca R and D, Boston, Waltham, MA, USA

Helicobacter pylori **is a Gram-negative bacterium that is associated with the development of peptic ulcers and gastric carcinoma in humans. This species appears to be one of the most genetically variable bacteria described to date. The overall level of heterogeneity within strains of this organism was determined by comparing the genome sequences of two reference strains, J99 and 26695. The aim of this study was to measure the genetic diversity within strains of *H. pylori* by looking for strain-specific genes in nine *H. pylori* strains isolated from patients suffering from chronic gastritis (*n* = 3), duodenal ulcers (*n* = 3) or gastric cancer (*n* = 3). Seven loci that contained strain-specific genes in strains J99 and 26695 were studied. These regions were subsequently amplified from most of the clinical isolates studied and their sequences were determined. ORFs were predicted from the sequence data and were compared to sequences within the databases. The results showed that the genes flanking the ORFs specific to either strain J99 or strain 26695 were also present in a similar configuration in the genomes of the nine clinical isolates. Moreover, in most regions, ORFs homologous to those found in the corresponding loci in the two reference strains were detected. However, in 10 regions, genes similar to those located at another locus in the genome of J99 or 26695 were found. Finally, six strain-specific genes were identified in three regions of three of the *H. pylori* strains isolated from patients with duodenal ulcers (*n* = 2) and gastric cancer (*n* = 1). Of these six genes, five were putative genes and one was an orthologue of a gene encoding a transposase in *Thermotoga maritima*. However, no association with disease was found for these genes.**

Keywords: diversity, genome, pathogenicity

INTRODUCTION

Helicobacter pylori is one of the most common human pathogens and it colonizes the gastric mucosa. Infection of the gastric mucosa is associated with diverse severe gastroduodenal diseases, such as gastritis, peptic ulcers and gastric adenocarcinoma (Cover & Blaser, 1999). Among the factors influencing disease outcome, strain-dependent factors are thought to play an important role (Axon, 1999). Virulence determinants such as the presence of an intact *cag* pathogenicity island and the expression of the vacuolating cytotoxin VacA have been described (Atherton *et al.*, 1995, 1997; Censini *et al.*, 1996; Maeda *et al.*, 1998). Recent functional studies have partially elucidated the role of these factors in the virulence of *H. pylori* (Backert *et al.*, 2000; Galmiche *et al.*, 2000; Krause *et al.*, 2000; McClain *et al.*, 2000; Odenbreit *et al.*, 2000; Stein *et al.*, 2000).

The discovery of new strain-dependent factors potentially involved in the clinical outcome of infection with *H. pylori* has ensued from the genomic and post-genomic eras. The total genome sequence of two *H.*

† **Present address:** Laboratoire de Virologie, Institut de Biologie Végétale Moléculaire, Institut National de la Recherche Agronomique, Bordeaux, France.

The GenBank accession numbers for the *H. pylori* sequences reported in this paper are AF326599–AF326607 for region A, AF326608–AF326616 for region B, AF326617–AF326625 for region C, AF326626–AF326634 for region D, AF327212–AF327220 for region E, AF328909–AF328916 and AF328924 for region F, and AF32917–AF32923 for region G.

pylori strains has been available since 1999 (Alm *et al*., 1999; Tomb *et al*., 1997), and the availability of these complete genome sequences was truly the beginning of comparative genomics for *H. pylori*. The overall genome organization of the two sequenced strains of *H. pylori* differs by 10 inverted or transposed regions. Genes conserved in both strains and so-called strain-specific genes (Alm & Trust, 1999; Doig *et al*., 1999) have been identified by comparison of the gene content of the two strains. Strain-specific genes were originally defined as being present in only one of the two completely sequenced *H. pylori* genomes, although subsequent analysis has led to some of these genes being identified in other *H. pylori* isolates (Occhialini *et al*., 2000; Salama *et al*., 2000). Although the functions of the putative encoded proteins are unknown for most of the strain-specific genes (70%), these genes may play a role in the virulence capacities of *H. pylori* strains by encoding factors that contribute to a different disease outcome. Concerning their location, almost half of the strain-specific genes are clustered in a single hypervariable region, the so-called 'strain-specific plasticity zone' described by Alm *et al*. (1999). A study by Occhialini *et al*. (2000) involved the analysis of the diversity of the plasticity zone in 43 *H. pylori* strains and showed that this region appears highly mosaic in nature.

The goal of the present study was to measure the genetic diversity of *H. pylori* strains by analysing the loci that contain the J99 or 26695 strain-specific genes (65 in strain 26695 and 47 in strain J99) located outside the plasticity zone. Although these strain-specific genes were not clustered into one locus, their location did not seem to be random. Indeed, it was noted that in 17 corresponding loci both reference strains contained strain-specific genes, suggesting a limit in the flexibility of the genome to strain-specific content (Alm & Trust, 1999; Alm *et al*., 1999). Hence, we proposed the hypothesis that other *H. pylori* strains contain their own set of specific genes located in similar loci. Seven strain-specific loci among the 17 loci common to both reference strains were selected; the genetic composition of these loci in nine *H. pylori* strains isolated from patients suffering from the principal diseases caused by *H. pylori* was analysed.

METHODS

Clinical samples, bacterial strains and culture. Nine *H. pylori* strains were isolated from gastric biopsy specimens from patients living in Costa Rica who were suffering from duodenal ulcers (U) ($n=3$), gastric carcinoma (C) ($n=3$) or chronic gastritis only (g) ($n=3$). Six of these strains have been described in previous studies by Occhialini *et al*. (2000, 2001). The biopsies were ground as described previously (Marais *et al*., 1999b), before inoculation onto an in-house medium made of Wilkins–Chalgren agar (Oxoid) enriched with 10% human blood and rendered selective by the addition of antibiotics (10 mg vancomycin l^{-1}, 5 mg cefsulodin l^{-1}, 5 mg trimethoprim l^{-1} and 100 mg cycloheximide l^{-1}). The plates were incubated under microaerobic conditions at 37 °C for up to 12 days. The organisms were identified as *H. pylori* by Gram-staining, as well as by their urease, oxidase and catalase activities. Reference strains 26695 (NCTC 12455) and J99, whose genomes have been sequenced (Alm *et al*., 1999; Tomb *et al*., 1997), were included in the study. A panel of 11 *H. pylori* strains isolated from gastric cancer patients and 15 strains isolated from patients with gastritis only was also used to look at the prevalence of strain-specific genes in this organism.

In preparation for DNA extraction, the strains were subcultured on the same medium as described above for 48 h, harvested in 1 ml *Brucella* broth (BBL Microbiology Systems) and centrifuged for 15 min at 3000 ***g***; the resulting pellets were stored in sterile vials at −80 °C until use.

Total DNA extraction. The cells were resuspended in 1 ml extraction buffer [20 mM Tris/HCl (pH 8), 0·5% Tween 20] and treated with 10% SDS and proteinase K (100 μg ml^{-1}). After at least 1 h at 56 °C, the proteins were eliminated from the lysate by solvent extraction using a standard protocol (Sambrook *et al*., 1989). Nucleic acids were precipitated from the lysate in the presence of 70% ethanol and 0·3 M sodium acetate (pH 5·2) at −80 °C for 30 min. After centrifugation and washing of the DNA with 70% ethanol, it was dissolved in an appropriate volume of sterile water and stored at −20 °C. The DNA concentration was determined at 260 nm.

Amplification of strain-specific loci. Oligonucleotides used as primers to amplify strain-specific loci present in *H. pylori* DNA were designed on the basis of the published sequences of *H. pylori* strains J99 and 26695 (Alm *et al*., 1999; Tomb *et al*., 1997) [available at the *Helicobacter pylori* Genome Database web site (http://scriabin.astrazeneca-boston.com/hpylori) and the Institute for Genomic Research web site (http://www.tigr.org), respectively] and are listed in Table 1. Primers which annealed to conserved flanking genes within the sequences of J99 and 26695 were used, to allow the amplification of the intervening sequences; the sizes of the amplicons produced by these primers are shown for both reference strains in Table 1. In addition to the aforementioned primers, primers which annealed within the six strain-specific genes of interest have been described (Table 2); these were used to screen for the presence of these genes in the larger panel of strains.

Amplifications were carried out in a total volume of 50 μl containing 10 × PCR buffer [500 mM Tris/HCl (pH 9·3), 150 mM $(NH_4)_2SO_4$, 25 mM $MgCl_2$, 1% Tween 20], 25 μM dNTPs, 2·5 U of Accu*Taq* DNA polymerase (Sigma-Aldrich), 1 μM of each primer and 5 ng of template DNA. The PCRs were performed in a GeneAmp PCR System 9700 (Perkin-Elmer Applied Biosystems). The amplicons were visualized after electrophoresis was done on an agarose gel stained with ethidium bromide.

Sequencing of amplified fragments. DNA sequencing was performed by using the dideoxynucleotide chain termination method (Sanger *et al*., 1977) with the dRhodamine Termination Cycle Sequencing Kit (Perkin-Elmer). Before sequencing, the amplicons were purified by using Wizard PCR preps (Promega). The same primers as used for PCR were employed for sequencing (Table 1), as well as internal primers (not shown). According to the manufacturer's protocol, reagent mixtures containing 1–5 μl of purified PCR product were placed in the thermal cycler and cycling was carried out under the following conditions: 25 cycles at 96 °C for 10 s, 50 °C for 5 s and 60 °C for 4 min. The resulting sequences were analysed through a polyacrylamide (4·25%) urea (7 M) gel in TBE buffer [89 mM Tris/HCl (pH 8·3), 89 mM boric acid, 2 mM EDTA] at 51 °C in an ABI PRISM 377 Genetic Analyser (Perkin-Elmer). For each sample, both strands of the PCR product were sequenced.

Table 1. Oligonucleotide primers used to amplify strain-specific loci

Amplified region	Flanking ORFs (JHP no./HP no.)*	Nucleotide coordinates of the 5′ ends of the forward and reverse primers in J99 genome	Sequences of primers (5′→3′)	Expected amplicon size (bp) from	
				J99	26695
A	JHP1131/HP1208	1264459	ATGGATATTTTCCTTAATAAAGGG	825	875
	JHP1133/HP1210	1265283	CTAAAGCCAAAACCATCACAAAGG		
B	JHP317/HP334	344362	ATCAAGCTTGTAGATTTTAAGGGC	1540	1300
	JHP319/HP337	345901	TTCTACGATTTTAACCGCTTGCTC		
C	JHP725/HP788	807636	CTATAAAATGGGGTGGAACGGGCC	1746	1824
	JHP727/HP791	809381	GGGGGTTACGCTTTTAGCCTTAGC		
D	JHP1436/HP1391	1584663	TGTGGCATGAAGAGCATTTGTTGG	1222	2423
	JHP1438/HP1387	1585884	AGAGCCTATGCGGACGCTTTAGC		
E	JHP539/HP591	586689	TAAAAAAGTCCCTGCAAAAGTCGC	2189	5216
	JHP541/HP594	588877	TTTAAACGCTATCAAAATAGCGGC		
F	JHP43/HP50	50000	CTTGAAATTTTTATCCCATTCCCC	3540	6512
	JHP47/HP55	53540	GGATATTTATGAAATCGTTCCGGG		
G	JHP812/HP878	892388	AAACCTCACTAAAACCAACTGGGC	2124	1888
	JHP815/HP883	894511	CATTCAAGATGAAAACAGACCCGC		

* JHP no. and HP no. represent the names of the ORFs in the J99 and 26695 genome sequences, respectively.

Table 2. Oligonucleotide primers designed to anneal inside the six strain-specific genes of interest

Region	Flanking ORFs (JHP no./HP no.)*	Amplified ORF strain no. (size in bp)	Sequences of primer (5′→3′)	Expected amplicon size (bp)
C	JHP725/HP788	14U (282)	GCCAATTTGAGGGCGTTG	579
	JHP727/HP791		AATAAAGAAAATGAAATCAATAA	
D	JHP1436/HP1391	2C (227)	CACAACAGGTATTAATGC	628
	JHP1438/HP1387		CTCTTGTTTCCAGCCATC	
D	JHP1436/HP1391	16U-1 (52)	ACGCTCTTTAGGGATACA	105
	JHP1438/HP1387		CTGTATTGATAATGGCCTC	
D	JHP1436/HP1391	16U-2 (55)	GAATGCTATGGTAATGATGA	107
	JHP1438/HP1387		GCTGTGCTACTTCTTTTTC	
D	JHP1436/HP1391	16U-3 (102)	GAATGCTATGGTAATGATGA	183
	JHP1438/HP1387		CCATGCGACGATTATACC	
G	JHP812/HP878	2C (50)	GTGCTAATATTGTCAAGG	86
	JHP815/HP883		TCACCTTGGACCATACAC	

* JHP no. and HP no. represent the names of the ORFs in the J99 and 26695 genome sequences, respectively.

Sequence analysis and comparisons. Nucleotide sequences were analysed by using the programs SEQUENCE NAVIGATOR and AUTOASSEMBLER 2.0 (Perkin-Elmer). Predicted coding regions were defined by searching for ORFs longer than 50 codons that had a ribosome binding consensus site upstream of a potential start codon. The sequences were compared with those within the GenBank databases by using the BLAST (basic local alignment search tool) and PSI-BLAST (position-specific iterative BLAST) programs (Altshul *et al.*, 1997) at the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/). Particular motifs were identified using the PFSCAN software at the Swiss Institute for Experimental Cancer Research (ISREC; http://www-isrec.unil.ch/).

Nucleotide sequence accession numbers. DNA sequences generated in this study were deposited in the GenBank database with the following accession numbers: AF326599–AF326607 for region A; AF326608–AF326616 for region B; AF326617–AF326625 for region C; AF326626–AF326634 for region D; AF327212–AF327220 for region E; AF328909–AF328916 and AF328924 for region F; AF32917–AF32923 for region G.

RESULTS

Choice of the strain-specific loci

Of the regions of the *H. pylori* genome that contained genes defined as strain-specific, on the basis of the comparison made between the gene content of the two reference strains J99 and 26695, seven were chosen for further investigation in this study. The main criterion for the choice of these regions was based on the putative

Table 3. Strain-specific genes encoded by regions A–G in *H. pylori* reference strains J99 and 26695

Region	Strain	Strain-specific ORF	Gene	Putative function
A	J99	JHP1132	–	Unknown*
	26695	HP1209	*iceA*	Ulcer-associated gene restriction endonuclease
B	J99	JHP318	–	Hypothetical†
	26695	HP335, HP336	–	Hypothetical†
C	J99	JHP726	*hsdS*	Type I restriction enzyme (specificity subunit)
	26695	HP789	–	Unknown*
		HP790	*prrB*	Anti-codon nuclease masking agent
D	J99	JHP1437	–	Hypothetical†
	26695	HP1388, HP1389, HP1390	–	Unknown*
E	J99	JHP540	–	Hypothetical†
	26695	HP592	*res*	Type III restriction enzyme R protein
		HP593	*mod*	Adenine-specific DNA methyltransferase
F	J99	JHP44, JHP45	–	Type II DNA modification enzyme (methyltransferase)
		JHP46	–	Type II restriction enzyme
	26695	HP51	*ddeM*	Cytosine-specific DNA methyltransferase
		HP52, HP53	–	Unknown*
		HP54	–	Adenine/cytosine DNA methyltransferase
G	J99	JHP813, JHP814	–	Unknown*
	26695	HP880, HP881, HP882	–	Unknown*

* *H. pylori*-specific with no known function.

† Conserved with no known function.

functional assignments of the strain-specific genes located between two conserved flanking genes of these strain-specific regions. Most of the strain-specific genes present in the genomes of strains J99 and 26695 are also *H. pylori*-specific (58%), with no orthologues identified in the databases (Alm & Trust, 1999; Alm *et al.*, 1999). Up until now, only strain-specific outer-membrane-related genes (Alm *et al.*, 2000) have been found to have an impact on the properties of *H. pylori*.

On the basis of the above criterion, among the candidate loci of *H. pylori* with no assigned function, seven regions of the genomes of strains J99 and 26695 that contained strain-specific genes were chosen for further investigation (Table 3). All of these regions contained at least one gene with an unknown or hypothetical function, depending on whether they belonged to the '*H. pylori*-specific with no known function' group or the 'conserved with no known function' group, respectively. The latter group indicates that orthologous genes have been identified in other species, but these orthologues have no known function. Three of the regions (B, D and G) of the *H. pylori* genome identified here encode only genes that fall into the two aforementioned categories; the four remaining regions (A, C, E and F) contain genes with assigned functions and putative genes with no homologues (Table 3).

Another criterion used in choosing which regions of the *H. pylori* genome to study was the size of the polymorphisms between the conserved genes of strains J99 and 26695. Regions with similar intervening sequence sizes (A, B, C and G; Table 1) as well as those with very different sizes (D, E and F; Table 1) were chosen. Moreover, a large range of polymorphism sizes was chosen [from 825 bp (region A in J99) to 6512 bp (region F in 26695)].

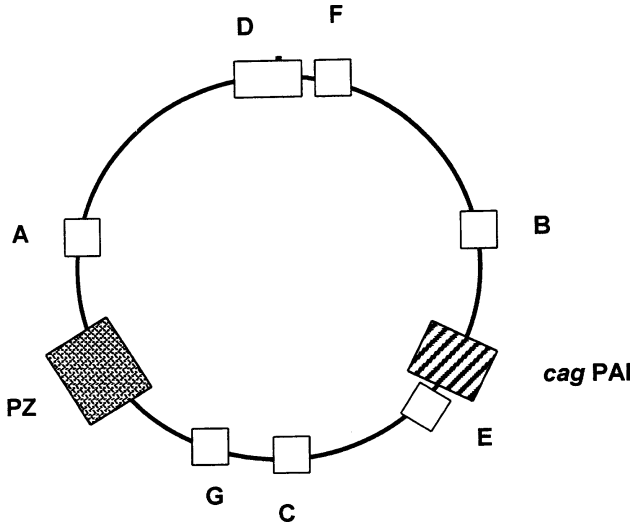

Fig. 1. Location of the seven regions studied that contain strain-specific genes in the chromosome of strain J99. The positions of the *cag* pathogenicity island (*cag* PAI) and the plasticity zone (PZ) are also shown. The vertical bar above D indicates the position of the origin of replication.

Fig. 1 shows the location of the seven strain-specific loci within the J99 genome, and their distribution around the chromosome.

Characterization of seven loci containing putative strain-specific genes

The results of this study are summarized in Table 4. Of the 63 expected amplification products that corresponded to the seven loci of the nine clinical *H. pylori* strains studied, only two were not amplified (region G from strains 15U and 38g) (Table 4). These negative results may have been due to (i) the absence of one or both of the flanking genes JHP812/HP878 and JHP815/HP883 in region G of strains 15U and 38g, (ii) the inverse orientation of the flanking genes JHP812/HP878 and JHP815/HP883 in these two strains, (iii) the non-contiguity of the genes in region G in these two strains, (iv) the absence of primer annealing due to strain-specific sequence differences in region G of these two strains or (v) the overwhelming size of the sequence between the flanking genes of region G. The amplification of the flanking genes JHP812/HP878 and JHP815/HP883 from the DNA of strains 15U and 38g (data not shown) confirmed that these two genes were conserved in these strains, but the other four possibilities detailed above still exist to explain the failure of amplification of any possible intervening sequence.

All other amplifications involving the seven loci of the nine clinical isolates yielded single fragments, suggesting that the two flanking genes of each of the seven regions were conserved not only in the J99 and 26695 genomes, but also in the genomes of the nine clinical isolates studied here. The orientation of these flanking genes was also conserved, allowing amplifications to be done using the forward and reverse primers that were designed on the basis of the J99 and 26695 whole-genome sequences (Table 1).

The most-conserved regions of the *H. pylori* genome, in terms of the size of the amplified fragment, were regions A and E (Table 4). With respect to region E, the nine clinical isolates resembled strain J99 more than they resembled strain 26695. The ORF present in region E shared significant homology with ORF JHP540, which is located between the same flanking genes in the J99 genome as region E is in the nine clinical strains (Tables 3 and 4). Despite the similar sizes of the amplified fragments from the nine *H. pylori* clinical strains, region A does not encode the same proteins in all of these strains. Indeed, from the amplification products of four of the strains (2C, 15U, 35g and 38g), ORFs similar to the IceA1 protein of strain 26695 were predicted (Table 4). Although no ORF was predicted in strains 4C, 9C, 14U and 29g that was similar to *iceA1*, sequence similarity was detected between the region A sequences of these strains and this gene. The lack of predicted ORFs for region A of these four strains seemed to be due to the accumulation of mutations in this region that led to the creation of stop codons. Region A of strain 16U was found to encode an ORF homologous to ORF JHP1132 of J99. In strain J99, this ORF encodes the IceA2 protein. In contrast to region E, region A generally seemed to be more closely related to the corresponding locus in strain 26695 than the one in strain J99.

Even though the size of the amplified fragments varied extensively (859–1556 bp; Table 4), region B was highly conserved in the nine clinical isolates. Indeed, this region was either homologous to the JHP318 gene of strain J99, or similar to the JHP1024 gene of strain J99, a paralogue of JHP318 (Table 4).

With respect to region F, three groups of strains were distinguished depending on the composition of this region (Table 4). The first region F group of strains was composed of the three gastritis strains and strain 2C, in which this region was the largest (5000 bp). Several ORFs were predicted from the sequence of region F in these four strains. The first two ORFs had similarity with ORFs JHP46 and JHP45 of strain J99; the other predicted ORFs resembled chimeric ORFs of genes found in strains 26695 (ORFs HP52 and HP51) and J99 (ORF JHP44). Indeed, ORF1 (361 codons) of strains 2C, 29g, 35g and 38g was found to be homologous to JHP44 of strain J99 in the NH_2 part (first 72 codons) and to HP51 of strain 26695 in the remaining part (289 codons). The same situation was observed for ORF2 (408 codons), whose first 292 codons were homologous to HP52 of 26695 (90% identity) and whose remaining 116 codons were homologous to JHP44 of strain J99 (85% identity). The second region F group of strains (4C, 9C and 14U) was related to strain J99. The size of region F and the three homologous ORFs encoded by these strains were similar to those found in strain J99. Finally, the third group of region F strains comprised strains 15U and 16U. These strains had a deleted form of region F compared to that of strain J99. Indeed, region F of 15U and 16U contained only one ORF, which was similar to ORF JHP44 of strain J99, instead of the three J99 ORFs JHP44, JPH45 and JHP46.

Regions C, D and G were of particular interest to us. In some strains these regions were found to contain genes defined as strain-specific due to their absence from the genomes of strains J99 and 26695 and their presence in only one of our nine clinical isolates. Region D of strains 2C and 16U contained strain-specific ORFs (Table 4). Region D of the seven other strains contained an ORF similar to ORF JHP1437 of strain J99 (region D). Of the strain-specific loci studied here, regions C and G presented the greatest diversity – five combinations for region C and six for region G (Table 3). Moreover, both regions contained strain-specific genes in strains 14U and 2C. In the other seven clinical strains, regions C and G contained either ORFs homologous to those expected in the reference strains J99 or 26695 or ORFs encoded by the J99 or 26695 genomes but present in another locus (e.g. ORF JHP1044 in region G of strain 35g; Table 4). After examining the composition of region C in more detail, it was noted that the genes present at this locus belonged to the same paralogous gene family, i.e. the '*ghp* type I restriction enzyme, specificity subunit' family. The gene homologous to HP848 (*hsdS_2*) found in region C of strains 35g and 9C was a paralogue of the HP790 (*hsdS_5*) gene contained in strains 2C, 4C, 15U and 38g, and reference strain 26695 (Table 4); these genes displayed 56% identity in their amino-acid

Table 4. Characterization of seven strain-specific loci in nine *Helicobacter pylori* strains isolated from patients suffering from gastric carcinoma, duodenal ulcers and chronic gastritis

For each strain the size of the amplified fragment is indicated in bp; ORFs similar to HP (26695) ORFs or JHP (J99) ORFs are indicated and their sizes (bp) are indicated as subscript text; the size and percentage identity of each ORF in amino acids are indicated in parentheses.

Region	Gastric carcinoma strain			Duodenal ulcer strain			Chronic gastritis strain		
	2C	4C	9C	14U	15U	16U	29g	35g	38g
A	816 $HP1209^{172}$ (228, 88 %)	895 No ORF	865 No ORF	864 No ORF	797 $HP1209^{172}$ (228, 82 %)	766 $JHP1132^{59}$ (59, 94 %)	786 No ORF	818 $HP1209^{172}$ (228, 90 %)	804 $HP1209^{172}$ (228, 80 %)
B	1157 $JHP318^{251}$ (215, 67 %)	1556 No ORF	1280 $JHP318^{251}$ (213, 63 %)	1385 $JHP318^{251}$ (131, 71 %)	1208 No ORF	859 $JHP318^{251}$ (46, 51 %)	1324 $JHP318^{251}$ (216, 62 %)	1548 No ORF	1176 $JHP1024^{290}$ (122, 56 %)
C	1581 $HP790^{431}$ (409, 60 %)	1785 $HP790^{431}$ (212–175, 94 %, 80 %)*	1745 $HP848^{298}$ (413, 88 %)	2677 $JHP1422^{624}$ (370, 38 %) **Sp. ORF**† (**282**)‡	1763 $HP790^{431}$ (449, 82 %)	967 $JHP785^{207}$ (200, 51 %)	1758 $JHP726^{454}$ (424, 69 %)	1653 $HP848^{298}$ (436, 82 %)	1816 $HP790^{431}$ (435, 69 %)
D	2211 **Sp. ORF** (**227**)	1144 $JHP1437^{256}$ (253, 80 %)	1359 $JHP1437^{256}$ (310, 68 %)	1387 $JHP1437^{256}$ (325, 74 %)	1145 $JHP1437^{256}$ (255, 87 %)	1000 **Sp. ORF** (**52**) **Sp. ORF** (**55**) **Sp. ORF** (**102**)‡	1173 $JHP1437^{256}$ (125–115, 81 %, 88 %)*	1349 $JHP1437^{256}$ (336, 69 %)	1350 $JHP1437^{256}$ (327, 69 %)
E	2098 $JHP540^{591}$ (591, 93 %)	2113 $JHP540^{591}$ (591, 92 %)	2113 $JHP540^{591}$ (591, 94 %)	2117 $JHP540^{591}$ (127–431, 97 %, 92 %)*	2133 $JHP540^{591}$ (591, 93 %)	2139 $JHP540^{591}$ (591, 94 %)	2146 $JHP540^{591}$ (591, 93 %)	2106 $JHP540^{591}$ (592, 90 %)	2119 $JHP540^{591}$ (421–180, 89 %, 96 %)*
F	5173 $JHP44^{185}$/$HP51^{355}$ (361, 35 %, 98 %) $HP52^{330}$/$JHP44^{185}$ (408, 90 %, 85 %) $JHP45^{343}$ (343, 94 %) $JHP46^{260}$ (261, 96 %)	3427 $JHP44^{185}$ (186, 82 %) $JHP45^{343}$ (343, 94 %) $JHP46^{260}$ (261, 97 %)	3508 $JHP44^{185}$ (185, 82 %) $JHP45^{343}$ (343, 94 %) $JHP46^{260}$ (261, 96 %)	3370 $JHP44^{185}$ (185, 78 %) $JHP45^{343}$ (315, 93 %) $JHP46^{260}$ (229, 89 %)	1048 $JHP44^{185}$ (185, 83 %)	1003 $JHP44^{185}$ (185, 80 %)	5245 $JHP44^{185}$/$HP51^{355}$ (361, 82 %, 97 %) $HP52^{330}$/$JHP44^{185}$ (408, 83 %, 35 %) $JHP45^{343}$ (343, 94 %) $JHP46^{260}$ (261, 97 %)	5157 $JHP44^{185}$/$HP51^{355}$ (361, 35 %, 98 %) $HP52^{330}$/$JHP44^{185}$ (408, 90 %, 85 %) $JHP45^{343}$ (343, 94 %) $JHP46^{260}$ (261, 97 %)	5190 $JHP44^{185}$/$HP51^{355}$ (107–234, 36 %, 96 %)* $HP52^{330}$/$JHP44^{185}$ (411, 86 %, 86 %) $JHP45^{343}$ (343, 94 %) $JHP46^{260}$ (261, 96 %)
G	1838 **Sp. ORF** (**50**) $JHP814^{117}$ (53, 42 %) $JHP440^{912}$ (57, 98 %)	1917 $HP488$/$HP1116^{957}$ (59, 47 %) $JHP1044^{1154}$ (104, 75 %) $JHP814^{117}$ (110, 76 %)	3793 JHP1044 (126, 64 %) $JHP813^{155}$ (113, 85 %)	2052 $JHP1044^{1154}$ (115, 76 %) $JHP813^{155}$ (147, 92 %)‡	NA	1490 $JHP1044^{1154}$(125, 73 %)	1715 $JHP813^{155}$ (72, 71 %)	2358 $JHP1044^{1154}$(145, 88 %)	NA

NA, No amplification.
* Two ORFs due to frameshift mutations.
† **Sp. ORF**, strain-specific ORF.
‡ Overlapping ORFs.

Table 5. Distribution of the six newly identified *H. pylori* strain-specific genes in a panel of 26 strains isolated in Costa Rica

The number of positive strains and the percentage of positive strains are shown.

Region	Amplified ORF strain no. (size in bp)	Gastric cancer (no. strains tested = 11)	Gastritis only (no. strains tested = 15)
C	14U (282)	1 (9%)	4 (26·6%)
D	2C (227)	8 (72·7%)	10 (66·6%)
D	16U-1 (52)	0	0
D	16U-2 (55)	5 (45·4%)	4 (26·6%)
D	16U-3 (102)	4 (36·3%)	1 (6·6%)
G	2C (50)	9 (81·8%)	10 (66·6%)

sequences. Moreover, the HP848 gene of strain 26695 corresponded to the JHP785 (*hsdS*_2) gene of strain J99, which was also found in region C of strain 16U (92·6% identity). Finally, the ORF homologous to JHP1422 of strain J99 predicted in region C of strain 14U (*hsdS*_3a) was a paralogue of ORF JHP785 of J99 (22% identity). The same observation could be made for the ORFs present in region G.

Characterization of the six newly identified strain-specific genes of *H. pylori*

The characterization of seven strain-specific loci (detailed above) of the nine clinical isolates of *H. pylori* studied here allowed the discovery of six strain-specific genes that were not present in reference strains J99 and 26695. Only three of the seven strain-specific regions contained strain-specific genes – regions C, D and G. Not all of the clinical strains contained these genes. Region C of strain 14U encoded a strain-specific ORF of 282 codons. Strain 2C contained two strain-specific ORFs in regions D and G of 227 and 50 codons, respectively. Finally, region D of strain 16U presented three strain-specific ORFs with sizes of 52, 55 and 102 codons.

The six strain-specific ORFs identified in this study shared no similarity among themselves, even when found in the same locus in different strains, i.e. region D in strains 2C and 16U. A comparison of the amino-acid sequences of the six newly identified strain-specific ORFs with the sequences contained within the databases showed that five of the six ORFs had no orthologue and hence should be classified as *H. pylori*-specific – the majority of the strain-specific genes found in the genomes of J99 and 26695 already have this classification (Doig *et al*., 1999). The ORF found in region D of strain 2C showed weak similarity (E-value of 10^{-8}) with a transposase of *Thermotoga maritima* (32% similarity, 23% identity). When searching for particular motifs in this ORF, a slight similarity was detected with an *N*-glycosylation site. Slight similarities with protein kinase C phosphorylation sites were also found in the ORFs in region C of strain 14U, in region G of strain 2C and in the ORFs of 52 and 102 codons in region D of strain 16U (data not shown).

Finally, screening for the presence of the six newly identified strain-specific ORFs in a panel of strains from the same geographical origin was performed, but no association with disease outcome was found for these genes (Table 5).

DISCUSSION

Comparative genomics of the two available *H. pylori* genome sequences has revealed that the great majority of metabolic and biosynthetic functions are conserved in reference strains J99 and 26695 (Alm *et al*., 1999; Doig *et al*., 1999; Marais *et al*., 1999a). However, around 6% of the genes in these strains are defined as strain-specific, because they are absent in one of the two genomes. Approximately half of these strain-specific genes are clustered into the so-called plasticity zone of *H. pylori*, described by Alm *et al*. (1999). A study was carried out to analyse the composition of the plasticity zone in a collection of *H. pylori* strains from diverse clinical origins (Occhialini *et al*., 2000). The results of the study by Occhialini *et al*. (2000) showed that the plasticity zone is highly mosaic and should be considered a genomic island, rather than a pathogenicity island per se. In this study, the strain-specific genes located outside the plasticity zone of *H. pylori* were investigated.

The observation that 17 strain-specific loci were common in the two reference strains J99 and 26695 (Alm *et al*., 1999) led us to develop a classical approach for the study of these loci in nine clinical isolates of *H. pylori*. This approach consisted of (i) amplification of the loci containing the strain-specific genes and (ii) the subsequent identification of the genes contained in the amplified fragments by sequencing and comparison of the resulting gene sequences with those contained in the databases. The results partially verified the hypothesis of a similar location of strain-specific genes in different strains of *H. pylori*, in that six new strain-specific genes not present in the genomes of the two reference strains were identified. Five are putative genes and, hence, are specific to *H. pylori*, like the majority of the strain-specific genes of J99 and 26695 (Alm *et al*., 1999). It

should be noted that three predicted ORFs, two in region D of strain 16U and one in region G of strain 2C, are very small in size (52, 55 and 50 codons, respectively) (Table 4) and, therefore, may not be genes. Nevertheless, we found consensus ribosome-binding sites upstream of initiation codons in these small predicted ORFs. The remaining strain-specific gene identified in region D of strain 2C showed homology with a transposase of *T. maritima*. This finding of a gene involved in DNA exchange and which may promote genetic diversity in *H. pylori* is not surprising. Indeed, many of the strain-specific genes of the two *H. pylori* reference strains belong to putative restriction–modification systems (10%) and 4% share similarities with genes encoding transposases (Salama *et al.*, 2000). Nevertheless, Kong *et al.* (2000) found that $<30\%$ of the potential type II restriction–modification systems in *H. pylori* J99 were fully functional. Another *H. pylori* strain, J166, has been shown to contain 18 specific genes when compared by subtractive hybridization to strain 26695, seven of which show homology to restriction–modification systems (Akopyants *et al.*, 1998). Kersulyte *et al.* (2000) identified a transposable element called IS*607* in *H. pylori*, located on a fragment present in only certain strains of this organism, which was also discovered by subtractive hybridization. Strain-specific genes involved in such systems have been identified in other bacterial species, e.g. *Klebsiella pneumoniae* (Lai *et al.*, 2000), *Neisseria meningitidis* (Bart *et al.*, 2000; Claus *et al.*, 2000) and *Aeromonas hydrophila* (Zhang *et al.*, 2000). Using representational difference analysis, Bart *et al.* (2000) and Claus *et al.* (2000) showed that restriction–modification systems were specifically present in lineage III meningococci. Suppression subtractive hybridization was used to identify genetic differences between virulent and avirulent strains of *A. hydrophila* isolated from diseased fish (Zhang *et al.*, 2000). Among the 69 genomic regions present only in the virulent strain of *A. hydrophila*, two-thirds encoded genes specific to *A. hydrophila* and one ORF belonged to a type II restriction–modification system. Using the same methodology as Akopyants *et al.* (1998), Lai *et al.* (2000) identified genes specifically present in a virulent strain of *K. pneumoniae*; among the 25 subtracted DNA clones, one encoded the transposase of Tn*3926*.

Besides the identification of the six new strain-specific genes in the *H. pylori* clinical isolates, we detected the presence of ORFs homologous to those found in either J99 or 26695 in the same loci. These results confirm that the gene order is highly conserved among isolates of *H. pylori* (Alm *et al.*, 1999; Bereswill *et al.*, 2000; Doig *et al.*, 1999), despite the extreme genetic diversity displayed by this bacterium, as shown by studies on genetic variability and population structure (Achtman *et al.*, 1999; Suerbaum, 2000; Suerbaum *et al.*, 1998). Overall, the nine clinical isolates are more closely related to strain J99 than to strain 26695, especially with regard to the plasticity zone (Occhialini *et al.*, 2000). However, in region A, eight of the nine strains contained a DNA fragment homologous to that present in 26695, i.e. the *iceA1* gene (Table 4). The remaining strain (16U) harboured the unrelated gene *iceA2*, found in J99. As in the study by Figueiredo *et al.* (2000), who analysed the *iceA* locus in 321 *H. pylori* strains from 24 different countries, we confirmed the presence of these two gene families (i.e. *iceA* and *iceA2*) at this locus. Figueiredo *et al.* (2000) found that the majority of strains (14/19) did not encode the full-length homologue of *Nla*III, a restriction endonuclease from *Neisseria lactamica* (Morgan *et al.*, 1996; Peek *et al.*, 1998). In our study, four of the eight strains studied contained an ORF of 228 codons that potentially encodes a full-length IceA1 protein (Table 4). Nevertheless, the association between the presence of *iceA1*-positive strains and the development of peptic ulcers, as described by Peek *et al.* (1998) and van Doorn *et al.* (1998), was not verified in our study. A recent study by Solcà *et al.* (2001) also showed that the *iceA1* allele was more frequent than the *iceA2* allele in *H. pylori* (59% versus 41%).

The results of the study by Figueiredo *et al.* (2000) suggested that the organization of the *iceA2* locus is very complex, with the presence of a variable number of tandem repeats (VNTRs) of an 8 bp sequence in the intergenic region upstream of the initiation codon of the IceA2 ORF. Moreover, *iceA2* was shown to encode proteins of various sizes, consisting of two conserved domains of 14 and 10 aa in length and a variable number of a 35 aa cassette, which was made up of domains of 13, 16 and 6 aa in length. This classification allowed Figueiredo *et al.* (2000) to distinguish five *iceA2* variants. Therefore, the *iceA2* variant present in strain 16U should be defined as being of the *iceA2B* form, as it could encode a protein of 59 residues that includes the 14 and 10 aa cassettes, flanking three internal peptide domains of 13, 16 and 6 aa, respectively. Only one VNTR was located in the intergenic region between *iceA2* and JHP1133/HP1210. The same proportion of *iceA2*-positive strains (from Costa Rica) was found in this study as in the study by Figueiredo *et al.* (2000) (5/34 strains). A relationship between the cassette structure of *iceA2* and expression was shown by Peek *et al.* (1998). *In vitro* expression of *iceA2* in strain 16U was confirmed by RT-PCR (data not shown). Neither the role of *iceA2* in *H. pylori* nor the relevance of the conserved genetic organization of this gene is understood, as yet.

Six of the nine clinical isolates of *H. pylori* included in our study were among the 43 strains whose plasticity zones have been analysed and for whom the compositions of the *cag* pathogenicity island have been determined (Occhialini *et al.*, 2000, 2001). Therefore, we attempted to find a correlation between the organization of the plasticity zone and the *cag* pathogenicity island, and between the organization of the plasticity zone and the strain-specific loci. All nine of the clinical strains studied here were found to contain an intact *cag* pathogenicity island (Occhialini *et al.*, 2000). Four patterns for the plasticity zone were distinguished among the nine strains – A1, A2, B1 and B2 (Occhialini *et al.*, 2001). No association was found between any one of these plasticity-zone groups and the composition of

the strain-specific loci, which is consistent with the high level of DNA diversity seen within strains of *H. pylori*.

Finally, the identification of new strain-specific genes in our study supports the idea that *H. pylori* strains contain other strain-specific genes that are not present in the J99 and 26695 sequences (Salama *et al*., 2000). Indeed, the study by Salama *et al*. (2000) was conducted to characterize the genetic diversity of *H. pylori* by examining the genomic content of 15 clinical isolates of this organism, using a whole-genome *H. pylori* DNA micro-array. These authors found that at least 12–18 % of the genome of each strain was composed of strain-specific genes that were not present in all of the strains surveyed (i.e. they lay outside of the 'core' set of genes). Micro-array technology is a particularly powerful tool for quantifying differential levels of expression of each gene for cells grown under different conditions (Nierman *et al*., 2000); however, for genetic variability studies, the experimental system itself leads to an underestimation of the number of strain-specific genes, as a micro-array contains only genes present in sequenced genomes. Alternative strategies for the identification of new strain-specific genes are promising, such as subtractive hybridization (Akopyants *et al*., 1998; Kersulyte *et al*., 2000; Lai *et al*., 2000; Zhang *et al*., 2000) or the classical methodology used in this study, which was made possible by the previous identification of candidate loci.

Although the discovery of the strain-specific genes described in this study adds to our knowledge of the *H. pylori* genome, none of these genes seems to be clinically relevant, based on the small survey performed here. The inclusion of these newly identified genes on *H. pylori* DNA micro-arrays will confirm their distribution and a functional approach to identifying their specific functions will contribute to assessing their role in *H. pylori*.

ACKNOWLEDGEMENTS

G. Chanto was supported by a research grant from the Service Culturel et de Coopération Scientifique de l'Ambassade de France à San José, Costa Rica. This work was supported by the Association pour la Recherche contre le Cancer (ARC) and the Conseil Régional d'Aquataine. We would like to thank Kathryn Mayo (Laboratoire de Bactériologie, Université Victor Segalen Bordeaux 2) for English corrections.

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.

Akopyants, N. S., Fradkov, A., Diatchenko, L., Hill, J. E., Siebert, P. D., Lukyanov, S. A., Sverdlov, E. D. & Berg, D. E. (1998). PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. *Proc Natl Acad Sci U S A* **95**, 13108–13113.

Alm, R. A. & Trust, T. J. (1999). Analysis of the genetic diversity of *Helicobacter pylori*: the tale of two genomes. *J Mol Med* **77**, 834–846.

Alm, R. A., Ling, L. S. L., 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.

Alm, R. A., Bina, J., Andrews, B. M., Doig, P., Hancock, R. E. W. & Trust, T. J. (2000). Comparative genomics of *Helicobacter pylori*: analysis of the outer membrane protein families. *Infect Immun* **68**, 4155–4168.

Altshul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Atherton, J. C., Cao, P., Peek, R. M., Jr, Tummuru, M. K. R., 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.

Atherton, J. C., Peek, R. M., Jr, Tham, K. T., Cover, T. L. & Blaser, M. J. (1997). Clinical and pathological importance of heterogeneity in *vacA*, the vacuolating cytotoxin gene of *Helicobacter pylori*. *Gastroenterology* **112**, 92–99.

Axon, A. T. R. (1999). Are all helicobacters equal? Mechanisms of gastroduodenal pathology and their clinical implications. *Gut* **45**, 11–14.

Backert, S., Ziska, E., Brinkmann, V., Zimny-Arndt, U., Fauconnier, A., Jungblut, P. R., Naumann, M. & Meyer, T. F. (2000). Translocation of the *Helicobacter pylori* CagA protein in gastric epithelial cells by a type IV secretion apparatus. *Cell Microbiol* **2**, 155–164.

Bart, A., Dankert, J. & van der Ende, A. (2000). Representational difference analysis of *Neisseria meningitidis* identifies sequences that are specific for the hyper-virulent lineage III clone. *FEMS Microbiol Lett* **188**, 111–114.

Bereswill, S., Schonenberger, R., Thies, C., Stahler, F., Strobel, S., Pfefferle, P., Wille, L. & Kist, M. (2000). New approaches for genotyping of *Helicobacter pylori* based an amplification of polymorphisms in intergenic DNA regions and at the insertion site of the *cag* pathogenicity island. *Med Microbiol Immunol* **189**, 105–113.

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 U S A* **93**, 14648–14653.

Claus, H., Friedrich, A., Frosch, M. & Vogel, U. (2000). Differential distribution of novel restriction–modification systems in clonal lineages of *Neisseria meningitidis*. *J Bacteriol* **182**, 1296–1303.

Cover, T. L. & Blaser, M. J. (1999). *Helicobacter pylori* factors associated with disease. *Gastroenterology* **117**, 257–261.

Doig, P., de Jonge, B. L., Alm, R. A. & 10 other authors (1999). *Helicobacter pylori* physiology predicted from genomic comparison of two strains. *Microbiol Mol Biol Rev* **63**, 675–707.

Figueiredo, C., Quint, W. G. V., Sanna, R. & 7 other authors (2000). Genetic organization and heterogeneity of the *iceA* locus of *Helicobacter pylori*. *Gene* **246**, 59–68.

Galmiche, A., Rassow, J., Doye, A. & 9 other authors (2000). The N-terminal 34 kDa fragment of *Helicobacter pylori* vacuolating cytotoxin targets mitochondria and induces cytochrome *c* release. *EMBO J* **19**, 6361–6370.

Kersulyte, D., Mukhopadhyay, A. K., Shirai, M., Nakazawa, T. & Berg, D. E. (2000). Functional organization and insertion specificity of IS607, a chimeric element of *Helicobacter pylori*. *J Bacteriol* **182**, 5300–5308.

Kong, H., Lin, L. F., Porter, N., Stickel, S., Byrd, D., Posfai, J. &

Roberts, R. J. (2000). Functional analysis of putative restriction–modification system genes in the *Helicobacter pylori* J99 genome. *Nucleic Acids Res* **28**, 3216–3223.

Krause, S., Barcena, M., Pansegrau, W., Lurz, R., Carazo, J. M. & Lanka, E. (2000). Sequence-related protein export NTPases encoded by the conjugative transfer region of RP4 and by the *cag* pathogenicity island of *Helicobacter pylori* share similar hexameric ring structures. *Proc Natl Acad Sci U S A* **97**, 3067–3072.

Lai, Y. C., Yang, S. L., Peng, H. L. & Chang, H. Y. (2000). Identification of genes present specifically in a virulent strain of *Klebsiella pneumoniae*. *Infect Immun* **68**, 7149–7151.

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.

Marais, A., Mendz, G. L., Hazell, S. L. & Mégraud, F. (1999a). Metabolism and genetics of *Helicobacter pylori*: the genome era. *Microbiol Mol Biol Rev* **63**, 642–674.

Marais, A., Monteiro, L., Occhialini, A., Pina, M., Lamouliatte, H. & Mégraud, F. (1999b). Direct detection of *Helicobacter pylori* resistance to macrolides by a polymerase chain reaction/DNA enzyme immunoassay in gastric biopsy specimens. *Gut* **44**, 463–467.

McClain, M. S., Schraw, W., Ricci, V., Boquet, P. & Cover, T. L. (2000). Acid activation of *Helicobacter pylori* vacuolating cytotoxin (VacA) results in toxin internalization by eukaryotic cells. *Mol Microbiol* **37**, 433–442.

Morgan, R. D., Camp, R. R., Wilson, G. G. & Xu, S. Y. (1996). Molecular cloning and expression of *Nla*III restriction–modification system in *Escherichia coli*. *Gene* **183**, 215–218.

Nierman, W. C., Eisen, J. A., Fleischmann, R. D. & Fraser, C. M. (2000). Genome data: what do we learn? *Curr Opin Struct Biol* **10**, 343–348.

Occhialini, A., Marais, A., Alm, R., Garcia, F., Sierra, R. & Mégraud, F. (2000). Distribution of open reading frames of plasticity region of strain J99 in *Helicobacter pylori* strains isolated from gastric carcinoma and gastritis patients in Costa Rica. *Infect Immun* **68**, 6240–6249.

Occhialini, A., Marais, A., Urdaci, M., Sierra, R., Munoz, N., Covacci, A. & Mégraud, F. (2001). Composition and gene expression of the *cag* pathogenicity island in *Helicobacter pylori* strains isolated from gastric carcinoma and gastritis patients in Costa Rica. *Infect Immun* **69**, 1902–1908.

Odenbreit, S., Puls, J., Sedlmaier, B., Gerland, E., Fischer, W. & Haas, R. (2000). Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. *Science* **287**, 1497–1500.

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.

Salama, N., Guillemin, K., McDaniel, T. K., Sherlock, G., Tompkins, L. & Falkow, S. (2000). A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc Natl Acad Sci U S A* **97**, 14668–14673.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463–5467.

Solcà, N. M., Bernasconi, M. V., Valsangiacomo, C., Van Doorn, L. J. & Piffaretti, J. C. (2001). 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. *Microbiology* **147**, 1693–1707.

Stein, M., Rappuoli, R. & Covacci, A. (2000). Tyrosine phosphorylation of the *Helicobacter pylori* CagA antigen after *cag*-driven host cell translocation. *Proc Natl Acad Sci U S A* **97**, 1263–1268.

Suerbaum, S. (2000). Genetic variability within *Helicobacter pylori*. *Int J Med Microbiol* **290**, 175–181.

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 U S A* **95**, 12619–12624.

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.

van Doorn, L.-J., Figueiredo, C., Sanna, R., Plaisier, A., Schneeberger, P., de Boer, W. & Quint, W. (1998). Clinical relevance of the *cagA*, *vacA*, and *iceA* status of *Helicobacter pylori*. *Gastroenterology* **115**, 58–66.

Zhang, Y. L., Ong, C. T. & Leung, K. Y. (2000). Molecular analysis of genetic differences between virulent and avirulent strains of *Aeromonas hydrophila* isolated from diseased fish. *Microbiology* **146**, 999–1009.

Received 1 February 2002; revised 17 May 2002; accepted 17 July 2002.