

Genomic Methylation: a Tool for Typing *Helicobacter pylori* Isolates^{∇†}

Filipa F. Vale¹ and Jorge M. B. Vitor^{2*}

Faculdade de Engenharia, Universidade Católica Portuguesa, Estrada Octávio Pato, 2635-631 Rio de Mouro, Portugal,¹ and Centro de Estudos de Ciências Farmacêuticas, Faculdade de Farmácia da Universidade de Lisboa, Av. Prof. Gama Pinto, 1649-019 Lisboa, Portugal²

Received 25 January 2007/Accepted 1 May 2007

The genome sequences of three *Helicobacter pylori* strains revealed an abundant number of putative restriction and modification (R-M) systems within a small genome (1.60 to 1.67 Mb). Each R-M system includes an endonuclease that cleaves a specific DNA sequence and a DNA methyltransferase that methylates either adenosine or cytosine within the same DNA sequence. These are believed to be a defense mechanism, protecting bacteria from foreign DNA. They have been classified as selfish genetic elements; in some instances it has been shown that they are not easily lost from their host cell. Possibly because of this phenomenon, the *H. pylori* genome is very rich in R-M systems, with considerable variation in potential recognition sequences. For this reason the protective aspect of the methyltransferase gene has been proposed as a tool for typing *H. pylori* isolates. We studied the expression of *H. pylori* methyltransferases by digesting the genomic DNAs of 50 strains with 31 restriction endonucleases. We conclude that methyltransferase diversity is sufficiently high to enable the use of the genomic methylation status as a typing tool. The stability of methyltransferase expression was assessed by comparing the methylation status of genomic DNAs from strains that were isolated either from the same patient at different times or from different stomach locations (antrum and corpus). We found a group of five methyltransferases common to all tested strains. These five may be characteristic of the genetic pool analyzed, and their biological role may be important in the host/bacterium interaction.

Microorganisms with spiral morphology in human and animal stomach have been known for over a century, but for many years, the stomach environment was considered too harsh for bacteria, and those observations were neglected (11, 19). After the isolation and cultivation of *Helicobacter pylori* from a human gastric biopsy specimen, achieved by Barry Marshall and Robin Warren in 1982, the scientific community finally accepted the idea that there are bacteria living in the human stomach (24). Soon after this important discovery, similar microorganisms were isolated from the stomachs of various wild and domestic mammals, such as cheetahs, dogs, pigs, and cats. Marshall and Warren discovered that several human gastric pathologies are induced by *H. pylori* (gastritis, or ulcers of the duodenum or of the stomach) and that the bacterium is associated with inflammation of the stomach mucous membrane. Later, several epidemiological studies found a strong correlation between gastric cancer and *H. pylori*. Since 1994, the World Health Organization has also classified *H. pylori* as a group 1 carcinogen (15, 41).

More than half of the human population is infected with *H. pylori*. In North America, as well as in northern and western Europe, 5 to 15% of children and 10 to 60% of adults are infected with *H. pylori*. The prevalence is much higher elsewhere (3). Although only a small percentage develop disease, this is a very large number of patients.

For these reasons, *H. pylori* research has been of high priority in human medical research over the last 2 decades, with

more than 23,500 papers listed in the PubMed database. Despite the intensive work on this microorganism, there are many unanswered questions about its basic biology. There are also some controversies: these include the mode of transmission of *H. pylori*, whether the human stomach is the only reservoir, and whether to classify it as a human pathogen or a commensal (7, 8).

In 1997 the first *H. pylori* genome was sequenced. The sequenced strain, NTCC 26695, was isolated from a patient with superficial gastritis. Remarkably, the genome had an extraordinarily large number of putative restriction-and-modification (R-M) systems: 26 putative methyltransferases (MTases) were identified (22, 38, 44). Two years later, a second strain, J99, obtained from a patient with duodenal ulcer, was sequenced. It was possible for the first time to compare two microbial genomes (1). Again, a large number of R-M systems were found. The 26 putative R-M systems of strains J99 are very similar to the 26 R-M systems of strain NTCC 26695: their genomic positions are very close, and a large number of these systems have more than 85% sequence similarity. *H. pylori* strains were believed to exhibit a large degree of genomic and allelic diversity, but it was found that the overall genomic organization, gene order, and predicted proteomes of the two strains were quite similar (1, 12). Recently, a third *H. pylori* strain, HPAG1, isolated from a patient with chronic atrophic gastritis, was sequenced, and 30 putative R-M systems were identified (31, 38). Definitively, the presence of large numbers of R-M systems is characteristic of *H. pylori*.

An R-M system is defined by the association of at least two genes: one codes for a restriction endonuclease (REase) that recognizes a specific DNA sequence and cuts both strands; the other gene codes for a cognate MTase that methylates the same DNA sequence, thus protecting it from being cleaved by the companion REase (39). According to the selfish gene the-

* Corresponding author. Mailing address: Faculdade de Farmácia de Lisboa, Av. Prof. Gama Pinto, 1649-019 Lisboa, Portugal. Phone: (351) 217946400, ext. 280 or 201. Fax: (351) 217946470. E-mail: jvitor@ff.ul.pt.

† Supplemental material for this article may be found at <http://aem.asm.org/>.

∇ Published ahead of print on 4 May 2007.

ory, once a type II R-M system is acquired by a bacterial genome, the deletion or inactivation of the type II MTase implies cell death (17, 18, 20, 28). This is because the cell loses the ability to protectively methylate all recognition sites in the newly synthesized chromosome, while unmethylated sites are cut by the residual REase still present in the bacterial cytoplasm. Due to the pressure of the REase on the expression of the cognate MTase, type II R-M systems tend to be maintained after acquisition. This last observation and the diversity of R-M systems in *H. pylori* led to the suggestion that the MTase from R-M systems could be used to type *H. pylori* strains (48). Typing methods are useful for understanding the natural history, epidemiology, mode of transmission, reservoir, and clinical implications of bacterial infection. However the typing methods used so far for *H. pylori* have not been very successful because of its genomic and allelic diversity. Indeed, no single molecular method has yet emerged as an internationally accepted basis for typing. A typing method for *H. pylori* will be useful for understanding many of the unclear and controversial items listed above, for appreciating population genetic structure, and for understanding the evolution of the microorganism (7, 8, 33).

In this study, we evaluated the diversity of MTase expression among 50 *H. pylori* strains isolated from Portuguese patients, and we assessed the use of genomic methylation status for typing *H. pylori* isolates. MTase expression is variable enough to be useful in tracing epidemics but stable enough to enable identification of a suitable set of restriction enzymes for the test.

MATERIALS AND METHODS

***H. pylori* strains.** The present study was performed on a collection of *H. pylori* strains containing 1,021 isolates from Portuguese patients obtained between 1990 and 2001 and maintained by the Portuguese National Institute of Health (INSA) in Lisbon.

We used a first subset of 21 strains, X1, randomly chosen (P1 to P21; see Table S1 in the supplemental material). The second subset of strains, X2, had 28 strains, isolated from 11 patients after biopsies performed with a gap of months to years (usually before and after antibiotic therapy) (P22.x to P32.x; see Table S2 in the supplemental material) or from biopsies performed on the same day from stomach antrum and corpus (patients A, B, and K). Some of these strains had already been typed by randomly amplified polymorphic DNA (RAPD), PCR-restriction fragment length polymorphism (RFLP) (*ureA-ureB*), and pulsed-field gel electrophoresis (32). This set allowed us to compare the utility of our method with that of existing methods. One strain, P33, was isolated from an asymptomatic patient (wife of patient E). A final reference strain, CCUG 15818, was isolated from a 65-year-old Australian male with a duodenal ulcer in 1983.

In total, 51 strains were used in this study, but only 32 were considered independently: all strains from Table S1 and the first strain isolated from each of the 11 patients with repeated observations (P22.1, . . . , P32.1; see Table S2 in the supplemental material).

The Kolmogorov-Smirnov (two-sided) test and the Welch-Satterthwaite test were used to confirm that both samples (X1 and X2) belonged to the same population, so that they could be analyzed together.

***H. pylori* culture.** The biomass from *H. pylori* strains was obtained from strains cultured on *H. pylori* selective agar (Wilkins-Chalgren agar supplemented with 10% horse blood, vancomycin [10 mg liter⁻¹], cefsulodin [5 mg liter⁻¹], trimethoprim [5 mg liter⁻¹], and cycloheximide [100 mg liter⁻¹] [Biogerm, Porto, Portugal]) and incubated at 37°C for 48 h in an anaerobic jar (Oxoid or BBL) with a gas generator system (CampyGen; Oxoid) (26, 27).

Genomic DNA extraction. The genomic DNA was extracted using the Pitcher method with minor modifications. The modifications included a final wash with 70% (vol/vol) ethanol and treatment with RNase A (Sigma) to a final concentration of 0.5 mg ml⁻¹ for 1 h at 37°C, after resuspension of the genomic DNA in 100 µl sterile double-distilled water (35).

REase choice. We chose 31 REases (New England Biolabs, Inc.) to evaluate the methylation status of the 51 strains (see Table S3 in the supplemental material). We verified the presence of the corresponding recognition sequences in the genomes of *H. pylori* strains NTCC 26695, J99, and HPAG1 (Table 1). This ensures that a lack of DNA digestion would strongly suggest that the corresponding recognition site is methylated (34).

Genomic DNA methylation. The genomic DNA was digested with REases, according to the manufacturer's instructions, in a volume of 50 µl. An excess of REase units was always used to ensure complete cleavage of the DNA. The DNA digestion status was observed after electrophoresis in 0.7% agarose gel (agarose LE; SeaKem, FMC) for 30 min and recorded (Polaroid 665 film or Kodak EDAS290). An undigested control of the genomic DNA was always present. The results were recorded on a binary matrix, where "0" indicates digestion (DNA is unmethylated) and "1" indicates no digestion (suggesting an active MTase).

Quality control. We set up two types of data quality control. First, we used four pairs of isoschizomers that differ with regard to which modified base prevents cleavage. One of each pair serves as an internal control for the presence of sites in the sequence and the ability of the DNA to be digested at all. These pairs were MspI and HpaII, BssKI and ScrFI, HhaI and HinPII, and NaeI and NgoMIV. Second, we did an REase replication assay, reisolating genomic DNA from seven randomly selected strains and repeating some of the DNA digestions (3.4% of the assays were repeated). The activity of each REase was verified by digesting a standard DNA from a sequenced phage or plasmid at the start and finish of each experimental series.

Data analysis. Kolmogorov-Smirnov and Welch-Satterthwaite statistical tests were used to determine whether samples X1 and X2 could have been drawn from the same population. The nonparametric Kolmogorov-Smirnov test determines whether the datasets, X1 and X2, differ significantly, and the Welch-Satterthwaite test is used to compare means without any a priori consideration of variances. The software NTSYS v.2.0 (Exeter Software) was used to construct a dendrogram from the binary matrix, using the unweighted pair group method with arithmetic average (UPGMA method) and the Jaccard coefficient. The Jaccard coefficient reflects the percentage of methylases common to two strains. To evaluate the data, the following parameters were determined: (i) the cophenetic correlation coefficient, which represents the goodness of fit of the dendrogram to the initial data; (ii) the discriminatory method capacity, which is the strain frequency per cluster and should be less than 5% ($n_j/N < 0.05$) (13); and (iii) Simpson's index of diversity, which reflects the capacity of the method to distinguish unrelated strains (25, 36).

RESULTS

Quality control. Validation of the data by control experiments yielded high confidence that the results reflect the state of modification of the genome of each strain. The four pairs of isoschizomers that are differentially sensitive to modification revealed the same methylation status in all tested strains. The replication assay (repeating DNA isolation and digestion) of 3.4% of total assays (54/1,581) showed 100% agreement with the first result obtained. For comparison, we calculated the reproducibility of methylation expression in two published studies. Ando et al. (2) included 10 common strains and 10 common REases previously used by Xu et al. (48). The reproducibility between these two studies was 97%, which clearly shows the stability of MTase expression in *H. pylori* during laboratory manipulation.

All strains in Table S1 (see the supplemental material) are obviously independent, as they were randomly selected. Could we consider any of the strains sequentially isolated from each patient independent (Table S2)? To answer this question we used two statistical tests, Kolmogorov-Smirnov (two-sided) and Welch-Satterthwaite (6). These tests results allowed us to consider 32 strains independent, all of which were from Table S1 (sample X1), and one of the multiple strains isolated from each of the 11 patients with repeated observations (sample X2p, that is, a partial set of sample X2 randomly chosen). The nonparametric Kolmogorov-Smirnov test showed that the two

TABLE 1. No. of recognition sites per sequenced genomes and comparison of resistance to REase cleavage in different studies

REase	Recognized sequence ^a	No. of recognition sites per sequenced genome ^b			% of strains resistant to cleavage (recognition site methylated)			
		26695	J99	HPAG1	Present study (n = 51)	A (n = 122) ^c	B (n = 19) ^d	C (n = 16) ^e
BseRI	GAGGAG	149	141	128	100			
EagI	CGGCCG	57	61	53	100			
HhaI	GCGC	5,924	6,035	5,638	100			100
Hinp1I	GCGC	5,924	6,035	5,638	100			
NaeI	GCCGGC	150	167	139	100			
NgoMIV	GCCGGC	150	167	139	100			
NlaIII	CATG	7,172	7,335	6,687	100	100	100	100
DpnI	GmATC	5,348	5,396	4,839	98			
DpnII	GATC	5,348	5,396	4,839	98			
BssHII	GCGCGC	282	303	266	96			
AseI	ATTAAT	484	424	426	90			
TacI	TCGA	303	336	275	88	86	63	
HpaII	CCGG	1,731	1,808	1,654	82	67		
HpyCH4V	TGCA	5,662	5,110	5,206	82	71		
MspI	CCGG	1,731	1,808	1,654	82			
HpyCH4IV	ACGT	282	500	278	65	74		
HpyCH4III	ACNGT	557	587	503	55	51		
DdeI	CTNAG	3,024	2,998	2,665	37	45	32	38
Hpy188I	TCNGA	1,396	1,244	1,209	37	21	84	13
FauI	CCCGC	1,657	1,637	1,535	24			
HaeIII	GGCC	1,607	1,598	1,540	24	38	21	13
BstUI	CGCG	2,478	2,664	2,358	14			
Hpy99I	CGWCG	247	267	244	14	62		
Sau96I	GGNCC	837	865	803	14			
AciI	CCGC	7,038	7,033	6,610	8			
BssKI	CCNGG	1,385	1,435	1,289	8			
ScrFI	CCNGG	1,385	1,435	1,289	8			
DraI	TTTAAA	4,905	4,692	4,338	6			
Fnu4HI	GCNGC	2,487	2,313	2,267	2			
FokI	GGATG	1,353	1,328	1,235	0	2	16	6
Sau3AI	GATC	5,348	5,396	4,839	0			

^a Data from REBASE (38).

^b Data from The Comprehensive Microbial Resource (34).

^c Data from Takata et al. (43).

^d Data from Ando et al. (2).

^e Data from Xu et al. (48).

data sets, X1 and X2p, do not differ significantly ($\alpha = 0.05$, $0.16 < D\alpha = 0.51$); i.e., both samples belong to the same population. To compare means without any a priori consideration of variances (X1 MTase mean = 14.62, standard deviation [SD] = 2.03; X2p MTase mean = 14.45, SD = 1.81), we used the Welch-Satterthwaite test. The result supported the same conclusion ($\alpha = 0.05$, $0.23 < t_{0.975}^{23} = 2.07$), i.e., there is an equality of mean values.

Simultaneous digestion by DpnI and DpnII. A peculiar result was obtained with the digestion of genomic DNA of strain P12 with DpnI and DpnII: DpnI cuts the sequence GATC only when the adenine residue is methylated (21), while DpnII does not cut GATC when that adenine residue is methylated (46). Theoretically, these enzymes should always give opposite results. In strain P12, both REases cut the DNA (Fig. 1a). This reaction was carried out several times, but always with the same outcome. A possible explanation for this result is that this strain has an adenine MTase that methylates a sequence that overlaps GATC, like M.SolI (GGmATCC), and the M.MboI isoschizomer is absent (GmATC) (14). We did not test this hypothesis because SolI is not commercially available. However, the simulation of a SolI digestion pattern for *H. pylori*

NCTC 26695 supports this hypothesis (Fig. 1b). Another explanation could be the presence of a regulatory protein that sequesters GATC sites and thus inhibits methylation, such as SeqA in *Escherichia coli* (23). However, simultaneous DpnI/DpnII digestion is an infrequent, but reproducible, observation, supporting the explanation of the presence of a rare MTase (Table 1). For this reason, the result of digestion of strain P12 by DpnI was considered negative (complete digestion results are in Table S4 in the supplemental material).

Dendrogram analysis. Of the 27 REases tested, isoschizomers and DpnI excluded, 7 gave the same result in all strains and thus are uninformative as a typing tool. These results were excluded from dendrogram analysis. A first dendrogram was constructed for the enlarged set of independent strains, with NTSYS version 2.0 software, using the UPGMA method and the Jaccard coefficient, for 32 strains [(X1 + X2p) \times 19 REases] (Fig. S1 in the supplemental material). The dendrogram has a cophenetic correlation factor of 0.79, which revealed a very good fit to the data matrix. Simpson's index of diversity was 100%, referring to the discriminatory power as the average probability that the typing method will identify as different types two unrelated strains randomly sampled in the

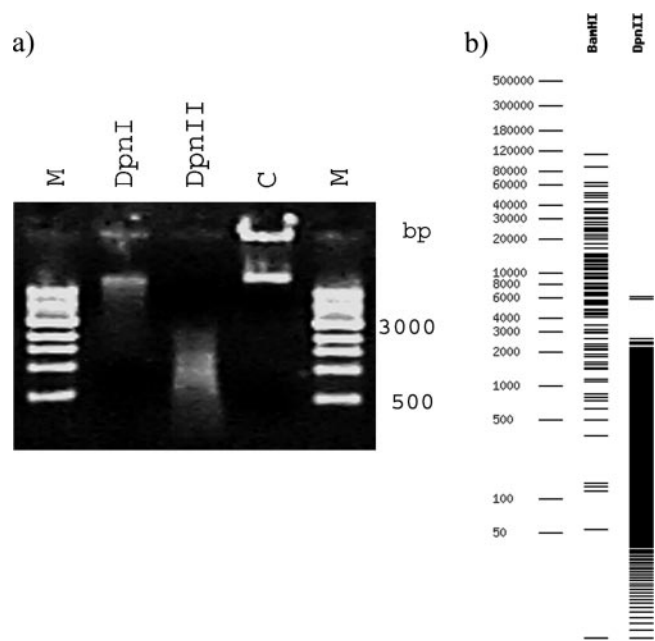


FIG. 1. (a) Strain P12 genomic DNA digested with DpnI and DpnII. Digestion with DpnI is clearly partial, because part of the DNA is the same size as the control (lane C). M, 1-kb DNA ladder (molecular size marker; New England Biolabs, Inc.). (b) Simulation of *H. pylori* NCTC 26695 genomic DNA digestion with BamHI and DpnII. BamHI digestion simulates the digestion pattern of DpnI if the DNA was methylated with M.SolI.

microbial population; the frequency of each group for a similarity level of 100% was less than 5% ($n_j/N = 0$) as suggested by Maslow et al. (25). The typeability, which is the proportion of strains that are assigned a type by the typing system, revealed that the 32 strains were considered different. When strains were chosen from set X2 by randomization, we obtained similar results (data not shown).

The inclusion of all strains (51 strains \times 19 REases) is more suitable for evaluating the performance of the typing method, especially in the present case, where quite similar strains were used in order to determine the discriminatory power of the method. The global dendrogram, constructed with NTSYS version 2.0 software with the UPGMA method and the Jaccard coefficient, had a cophenetic correlation factor of 0.79, which shows a very good fit of the basic dendrogram to the main matrix (Fig. 2). The Simpson index of diversity was excellent (0.996), and the frequency in each group for a similarity level of 100% was less than 5% ($n_j/N = 0.039$).

The global dendrogram analysis confirms that the method has a good discriminatory power. In the five pairs of strains isolated from two different locations in the same patient, antrum strains P22.1, P22.3, P23.1, P23.3, and P32.1 and corpus strains P22.2, P22.4, P23.2, P23.4, and P32.2, only three pairs, P22.1/P22.2, P22.3/P22.4, and P23.3/P23.4, have exactly the same methylation profile. This could be explained either by evolution in situ or by mixed infection.

Visual analysis of both dendrograms suggests that there are

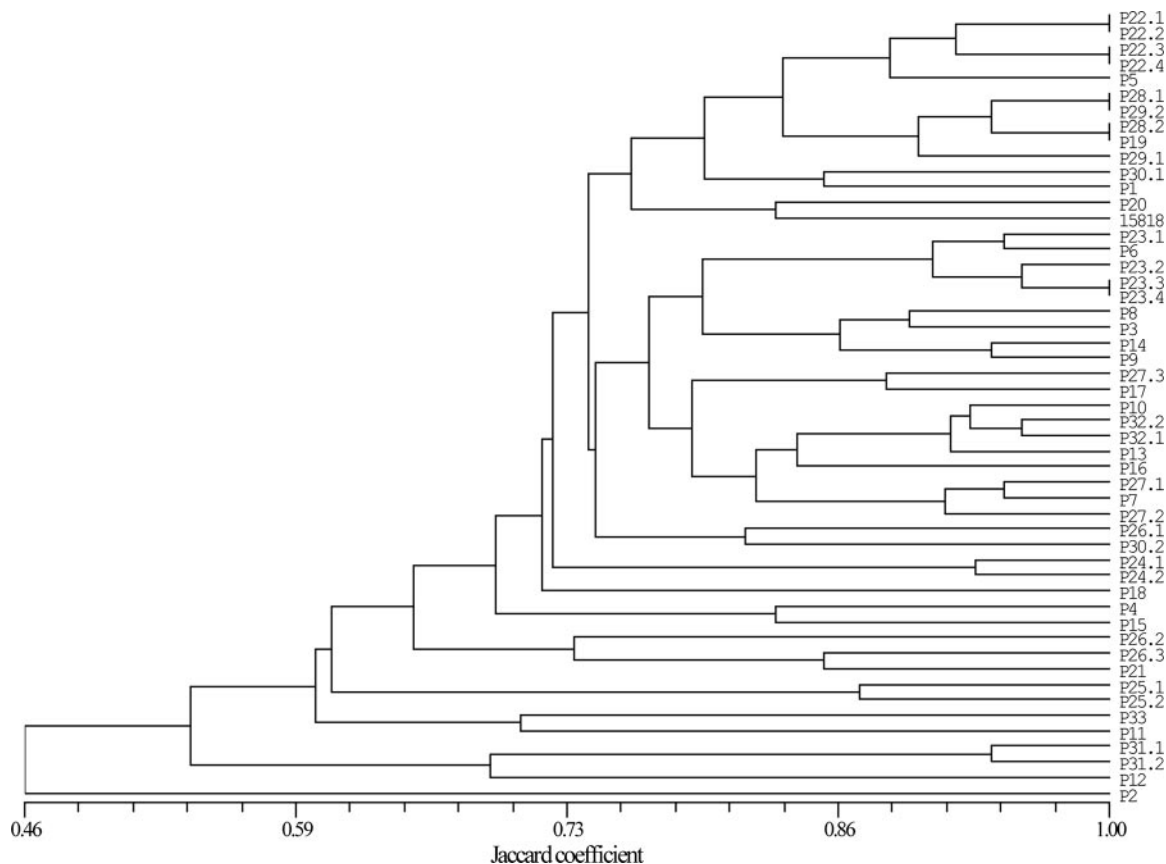


FIG. 2. Global dendrogram constructed with all strains (51 strains \times 19 REases), using the UPGMA method and the Jaccard coefficient.

TABLE 2. Observed changes in the methylation status of sequential isolates from the same patient

Patient	Theoretical maximal no. of changes ^a	No. of observed changes ^a	Strains	Observed changed methylation ^{b,c}	Same methylation status?	Same profile with different typing methods ^d ?	
						RAPD (% similarity)	PCR-RFLP (% similarity)
A	168 (28 × 6)	1	P22.1 and P22.2 P22.3 and P22.4	None None	Yes Yes	Yes Yes	Yes Yes
B	168 (28 × 6)	4	P22.1/P22.2 and P22.3/P22.4	+M.TacI	No	Yes	Yes
			P23.1 and P23.2 P23.3 and P23.4	M.HpyCH4III, M.DraI None	No Yes	Yes Yes	Yes Yes
			P23.1 and P23.3/P23.4 P23.2 and P23.3/P23.4	+M.HpyCH4III -M.DraI	No No	No (93) No (93)	Yes Yes
			P24.1 and P24.2	+M.FauI	No	No (97)	Yes
C	28	1	P25.1 and P25.2	+M.HpyCH4III, -M.Sau96I	No	No (90)	Yes
D	28	2	P26.1 and P26.2	-M.DdeI, +M.Hpy188I, -M.HpyCH4IV, -M.BssHII, -M.Sau96I	No	No (41)	No (55)
E	56 (28 × 2)	9	P26.2 and P26.3	-M.HpyCH4V, +M.BssHII, -M.FauI, +M.Sau96I	No	NT	NT
			P27.1 and P27.2 P27.2 and P28.3	-M.HpyCH4IV, +M.HpyCH4V +M.Hpy188I, -M.HpyCH4III, +M.HpyCH4IV, -M.HpyCH4V	No No	NT NT	NT NT
G	28	1	P28.1 and P28.2	+M.Hpy99I	No	NT	NT
H	28	2	P29.1 and P29.2	+M.Hpy99I, +M.HpyCH4V	No	NT	NT
I	28	5	P30.1 and P30.2	+M.MspI, +M.ScFI, +M.FauI, +M.TacI, +M.Sau96I	No	NT	NT
J	28	1	P31.1 and P31.2	-M.Sau96I	No	NT	NT
K	28	1	P32.1 and P32.2	M.AciI	No	Yes	Yes
Total	644	33					

^a The MTase alteration represented 5.1% (33/644) of the results, and the maintenance of the expressed MTase was 94,9% among the sequential isolated strains. ^b Isoschizomers not included.

^c +, MTase acquisition; -, MTase loss; no plus or minus sign, impossible to determine if the MTase alteration was an acquisition or loss.

^d "Yes" alone represents 100% similarity. When strains do not have the same profile, the similarity percentage is shown. NT, not tested. Data are from reference 32.

at least two clusters, with several subclusters (Fig. 2; also, see Fig. S1 in the supplemental material). It was also observed that there are two pairs of strains with similar methylation status that were not sequentially isolated from the same patient: P28.1/P29.2 and P28.2/P19.

The typing method based on the genomic methylation status is an excellent tool for *H. pylori* with respect to all performance criteria (42): i.e., (i) typeability (0.9; should be similar to 1); (ii) reproducibility (100%); (iii) stability (see below); (iv) discriminatory power (0.966; should be >0.95); (v) test population; (vi) epidemiological concordance (sequenced isolates show similar status of genomic methylation); (vii) typing system concordance (comparing with other methods showed similar results); and (viii) convenience criteria (simple, rapid, and inexpensive). These characteristics make the genomic methylation status a powerful method for typing *H. pylori* strains.

DISCUSSION

Diversity of MTase expression. We have observed a high number of expressed MTases, with an average of 14.6 distinct resistant recognition sequences (inferred to represent distinct MTases) per independent strain (SD, 1.9) and a maximum and minimum of 18 and 11 active MTases, respectively. A diversity of MTases expressed among independent tested strains was also observed. Both the large number and the diversity are in

agreement with results published elsewhere (4, 5, 22, 43, 45, 48). Among the tested MTases, eight gave the same result for all tested strains: five MTases were common to all (M.BseRI, M.EagI, M.HhaI, M.NlaIII, and M.NaeI), and two MTases were absent in all strains (M.FokI and M.Sau3AI). So far, only M.NlaIII (*iceA* or *hpyIM*) has been described as a conserved *H. pylori* MTase (49). The other nonmonotonous MTases can be classified into four groups accordingly to their frequency: (i) rare (M.AciI, M.DraI, M.Fnu4HI, and M.ScrFI), present in 2 to 8% of the strains; (ii) infrequent (M.BstUI, M.DdeI, M.FauI, M.HaeIII, M.Hpy99I, and M.Sau96I), present in 14 to 37%; (iii) frequent (M.Hpy188I, M.HpyCH4III, M.HpyCH4IV, M.HpyCH4V, and M.MspI), present in 55 to 82%; and (iv) very frequent (M.AseI, M.BssHII, M.DpnII, and M.TacI), present in >88% (Table 1). Inspection of Table 1 shows that in the other studies there are also four major groups of MTases. Most of them agree with the ones we found except for FokI, Hpy188I, and Hpy99I. The TacI results are in agreement with only one study (43) (Table 1). A possible explanation for this observation might be that some MTases are geographic markers. The other studies worked with groups of strains isolated from mixed geographic origins (Europe, Asia, North America, and Africa) (2, 43, 48).

It is also interesting that the numbers of recognition sites of the studied REases in the three sequenced *H. pylori* genomes

are remarkably similar (Table 1). This is in agreement with the observation that the overall genomic organization, gene order, and predicted proteomes of the two first sequenced strains are quite similar (12).

Maintenance of the expressed methylases. We performed a retrospective cohort study, treating sample X2 (Table S2 in the supplemental material) as a historical cohort, where all sequentially isolated strains available from the INSA collection and/or strains isolated from the stomach antrum and corpus were selected to determine the genomic methylation status. We observed a conservation of MTase expression of 94.9% (Table 2), revealing that after acquisition, the R-M system tends to be maintained and expressed (17, 18, 20, 28, 30).

We had also observed that strains isolated from stomach antrum and corpus of the same patient at the same moment had, in most cases, the same expressed methylases, except P23.1 and P23.2, with two observed differences, and P32.1 and P32.2, with one observed difference (Table 2; also, see Table S2 in the supplemental material). The differences in the methylation status between strains from stomach antrum and corpus may reveal an infection of the same patient with different *H. pylori* strains. Mixed infection with different *H. pylori* strains had already been described (9, 16, 47) and may provide the opportunity for horizontal gene transfer (10, 37).

Strains isolated from patients A to E and K were previously typed using PCR-RFLP (gene *ureA-ureB* digested with HaeIII) and RAPD (32). The strains from patients A and K were not differentiated, although genomic methylation revealed minor differences (Table 2). The other strains were discriminated with at least one method. Strains from patients B, C, and D were not identified as being different by PCR-RFLP but were discriminated with RAPD. Methylation analysis also discriminated these strains. Strains from patient E were considered different with RAPD, PCR-RFLP, and genomic methylation (Table 2).

The strains isolated from a couple (patients E and L) showed that the strain from the asymptomatic wife (P33) had a genomic methylation status different from that of all strains isolated from her husband (P26.1 to P26.3 [see Table S2 in the supplemental material]), as they were considered different when analyzed with PCR-RFLP and RAPD (32).

The results from the sequentially isolated strains demonstrate that the discriminatory power of the methylation status is superior to those of PCR-RFLP and RAPD, which are considered to have high discriminatory power and typeability (33). The stability of MTase expression in vivo (94.9%) in addition to the diversity of R-M systems present makes the use of genomic methylation status a quite attractive typing method. The 1-year follow-up of murine gastric colonization by three unrelated *H. pylori* strains did not show any temporal phase variation associated with R-M systems (40). Moreover, our follow-up of 10 patients during a period of 6 months to 4 years did not reveal substantial alteration in the MTase expression. Nevertheless, we cannot rule out the possibility that the patients were subject to reinfection and/or mixed colonization, which could have introduced a bias in this study.

Although MTase expression is quite stable, we found minor changes in methylation status in 5.1% of the epidemiologically related strains, reflecting mainly acquisitions of MTases, 63.3% (19/30). As expected for these strains, MTase acquisition was

almost two-thirds higher than MTase loss. The MTases may be acquired by horizontal gene transfer, as *H. pylori* organisms have access to DNA from bacteria that are introduced with food, because *H. pylori* is a naturally competent bacterium (29).

The loss of expression of a type II R-M system is more difficult due to the postsegregational killing effect (17, 18, 20, 28). However, it is possible that after REase loss, the orphan MTase suffers mutation, truncation, or deletion (43). In fact, among the sequenced strains there are several examples of orphan MTases: NCTC 26695 has two, J99 three, and HPAG1 five orphan MTases (38).

The typing method based on genomic methylation status is an excellent tool for *H. pylori*. An on-line database with the typing results from different origins would be easy to design and implement, allowing interlaboratory result comparison. This typing method can be applied to species with a high number of R-M systems, such as *Campylobacter upsaliensis*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Mycoplasma pulmonis*, *Thermoplasma volcanium*, *Methanococcus jannaschii*, *Methanosarcina mazei*, *Anabaena variabilis*, and *Xylella fastidiosa*, among others (38).

ACKNOWLEDGMENTS

We thank Lurdes Monteiro for the *H. pylori* strains, Dinis Pestana for valuable help on the statistical analysis, Elisabeth Raleigh for critical reading of the manuscript, Graça Malheiro for English review of the manuscript, Rick Morgan for many insightful discussions, and Carol Polisson for her assistance in our work.

This work was partially supported by New England Biolabs, Inc.

REFERENCES

- Alm, R. A., L. S. Ling, D. T. Moir, B. L. King, E. D. Brown, P. C. Doig, D. R. Smith, B. Noonan, B. C. Guild, B. L. deJonge, G. Carmel, P. J. Tummino, A. Caruso, M. Uria-Nickelsen, D. M. Mills, C. Ives, R. Gibson, D. Merberg, S. D. Mills, Q. Jiang, D. E. Taylor, G. F. Vovis, and T. J. Trust. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **397**:176–180.
- Ando, T., Q. Xu, M. Torres, K. Kusugami, D. A. Israel, and M. J. Blaser. 2000. Restriction-modification system differences in *Helicobacter pylori* are a barrier to interstrain plasmid transfer. *Mol. Microbiol.* **37**:1052–1065.
- Ando, T., Y. Goto, O. Maeda, O. Watanabe, K. Ishiguro, and H. Goto. 2006. Causal role of *Helicobacter pylori* infection in gastric cancer. *World J. Gastroenterol.* **14**:181–186.
- Aras, R. A., T. Takata, T. Ando, A. van der Ende, and M. J. Blaser. 2001. Regulation of the HpyII restriction-modification system of *Helicobacter pylori* by gene deletion and horizontal reconstitution. *Mol. Microbiol.* **40**:369–382.
- Aras, R. A., A. J. Small, T. Ando, and M. J. Blaser. 2002. *Helicobacter pylori* interstrain restriction-modification diversity prevents genome subversion by chromosomal DNA from competing strains. *Nucleic Acids Res.* **30**:5391–5397.
- Barnett, V. 2002. Sample survey—principles and methods, p. 184. Hodder Arnold, London, United Kingdom.
- Blaser, M. J., and J. C. Atherton. 2004. *Helicobacter pylori* persistence: biology and disease. *J. Clin. Investig.* **113**:321–333.
- Eslick, G. D. 2006. *Helicobacter pylori* infection causes gastric cancer? A review of the epidemiological, meta-analytic, and experimental evidence. *World J. Gastroenterol.* **12**:2991–2999.
- Ghose, C., G. I. Perez-Perez, L. J. van Doorn, M. G. Dominguez-Bello, and M. J. Blaser. 2005. High frequency of gastric colonization with multiple *Helicobacter pylori* strains in Venezuelan subjects. *J. Clin. Microbiol.* **43**:2635–2641.
- Gressmann, H., B. Linz, R. Ghai, K. P. Pleissner, R. Schlapbach, Y. Yamaoka, C. Kraft, S. Suerbaum, T. F. Meyer, and M. Achtman. 2005. Gain and loss of multiple genes during the evolution of *Helicobacter pylori*. *PLoS Genet.* **1**:e43.
- Han, S. R., H. C. E. Zschausch, H. G. Meyer, T. Schneider, M. Loos, S. Bhakdi, and M. J. Maeurer. 2000. *Helicobacter pylori*: clonal population structure and restricted transmission within families revealed by molecular typing. *J. Clin. Microbiol.* **38**:3646–3651.

12. Hancock, R. E., R. Alm, J. Bina, and T. Trust. 1998. *Helicobacter pylori*: a surprisingly conserved bacterium. *Nat. Biotechnol.* **16**:216–217.
13. Hunter, P. R., and M. A. Gaston. 1988. Numerical Index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J. Clin. Microbiol.* **26**:2465–2466.
14. Hwang, H. Y., and J. Yim. 1994. SolI, a novel isoschizomer of BamHI isolated from *Streptovorticillium olivovorticillatum*. *Nucleic Acids Res.* **22**: 2197.
15. International Agency for Research on Cancer. 1994. IARC monographs on the evaluation of carcinogenic risk to humans, no. 61. Schistosomes, liver flukes and *Helicobacter pylori*. IARC, Lyon, France.
16. Israel, D. A., N. Salama, U. Krishna, U. M. Rieger, J. C. Atherton, S. E. Falkow, and R. M. Peek, Jr. 2001. *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. *Proc. Natl. Acad. Sci. USA* **98**: 14625–14630.
17. Kobayashi, I. 1998. Selfishness and death: raison d'être of restriction, recombination and mitochondria. *Trends Genet.* **14**:368–374.
18. Kobayashi, I., A. Nobusato, N. Kobayashi-Takahashi, and I. Uchiyama. 1999. Shaping the genome—restriction-modification systems as mobile genetic elements. *Curr. Opin. Genet. Dev.* **9**:649–656.
19. Konturek, J. W. 2003. Discovery by Jaworski of *Helicobacter pylori* and its pathogenic role in peptic ulcer, gastritis and gastric cancer. *J. Physiol. Pharmacol.* **54**(Suppl. 3):23–41.
20. Kusano, K., T. Naito, N. Handa, and I. Kobayashi. 1995. Restriction-modification systems as genomic parasites in competition for specific sequences. *Proc. Natl. Acad. Sci. USA* **92**:11095–11099.
21. Lacks, S., and B. Greenberg. 1975. A deoxyribonuclease of *Diplococcus pneumoniae* specific for methylated DNA. *J. Biol. Chem.* **250**:4060–4066.
22. Lin, L.-F., J. Posfai, R. J. Roberts, and H. Kong. 2001. Comparative genomics of the restriction-modification systems in *Helicobacter pylori*. *Proc. Natl. Acad. Sci. USA* **98**:2740–2745.
23. Low, D. A., N. J. Nathan, and M. J. Mahan. 2001. Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. *Infect. Immun.* **69**:7197–7204.
24. Marshall, B. J., and J. R. Warren. 1984. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* **i**:1311–1315.
25. Maslow, J. N., M. E. Mulligan, and R. D. Arbeit. 1993. Molecular epidemiology: application of contemporary techniques to the typing of microorganisms. *Clin. Infect. Dis.* **17**:153–162.
26. Mégraud, F. 1996. Diagnostic bactériologique standard de l'infection à *Helicobacter pylori*, p. 249–266. *In* F. Mégraud and H. Lamouliatte (ed.), *Helicobacter pylori*, vol. 1. Elsevier, Amsterdam, The Netherlands.
27. Monteiro, L., and F. Mégraud. 1999. Par quels moyens rechercher *Helicobacter pylori* avant et après éradication? *Gastroenterol. Clin. Biol.* **23**:C3–C19.
28. Naito, T., K. Kusano, and I. Kobayashi. 1995. Selfish behavior of restriction-modification systems. *Science* **267**:897–899.
29. Nobusato, A., I. Uchiyama, and I. Kobayashi. 2000. Diversity of restriction-modification gene homologues in *Helicobacter pylori*. *Gene* **259**:89–98.
30. Nobusato, A., I. Uchiyama, S. Ohashi, and I. Kobayashi. 2000. Insertion with long target duplication: mechanism for gene mobility suggested from comparison of two related bacterial genomes. *Gene* **259**:99–108.
31. Oh, J. D., H. Kling-Backhed, M. Giannakis, J. Xu, R. S. Fulton, L. A. Fulton, H. S. Cordum, C. Wang, G. Elliott, J. Edwards, E. R. Mardis, L. G. Engstrand, and J. I. Gordon. 2006. The complete genome sequence of a chronic atrophic gastritis *Helicobacter pylori* strain: evolution during disease progression. *Proc. Natl. Acad. Sci. USA* **103**:9999–10004.
32. Oleastro, M. 2000. Master thesis. Universidade Técnica de Lisboa, Lisboa, Portugal.
33. Owen, R. J., D. Taylor, G. Wang, and L. J. van Doorn. 2001. Heterogeneity and subtyping, p. 363–378. *In* H. L. T. Mobley, G. L. Mendz, and S. L. Hazell (ed.), *Helicobacter pylori* physiology and genetics. ASM Press, Washington, DC.
34. Peterson, J. D., L. A. Umayam, T. M. Dickinson, E. K. Hickey, and O. White. 2001. The Comprehensive Microbial Resource. *Nucleic Acids Res.* **29**:123–125.
35. Pitcher, D. G., N. A. Saunders, and R. J. Owen. 1989. Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Lett. Appl. Microbiol.* **8**:151–156.
36. Priest, F., and B. Austin. 1993. Modern bacterial taxonomy, 2nd ed., p. 14–49. Chapman & Hall, London, United Kingdom.
37. Raymond, J., J. M. Thiberge, C. Chevalier, N. Kalach, M. Bergeret, A. Labigne, and C. Dauga. 2004. Genetic and transmission analysis of *Helicobacter pylori* strains within a family. *Emerg. Infect. Dis.* **10**:1816–1821.
38. Roberts, R. J., T. Vincze, J. Posfai, and D. Macelis. 2005. REBASE—restriction enzymes and DNA methyltransferases. *Nucleic Acids Res.* **33**: D230–D232.
39. Roberts, R. J., M. Belfort, T. Bestor, A. S. Bhagwat, T. A. Bickle, J. Bitinaite, R. M. Blumenthal, S. K. Degtyarev, D. T. Dryden, K. Dybvig, K. Firman, E. S. Gromova, R. I. Gumpert, S. E. Halford, S. Hattman, J. Heitman, D. P. Hornby, A. Janulaitis, A. Jeltsch, J. Josephsen, A. Kiss, T. R. Klaenhammer, I. Kobayashi, H. Kong, D. H. Kruger, S. Lacks, M. G. Marinus, M. Miyahara, R. D. Morgan, N. E. Murray, V. Nagaraja, A. Piekarowicz, A. Pingoud, E. Raleigh, D. N. Rao, N. Reich, V. E. Repin, E. U. Selker, P. C. Shaw, D. C. Stein, B. L. Stoddard, W. Szybalski, T. A. Trautner, J. L. Van Etten, J. M. Vitor, G. G. Wilson, and S. Y. Xu. 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* **31**:1805–1812.
40. Salaun, L., S. Ayraud, and N. Saunders. 2005. Phase variation mediated niche adaptation during prolonged experimental murine infection with *Helicobacter pylori*. *Microbiology* **151**:917–923.
41. Smith, M. G., G. L. Hold, E. Tahara, and E. M. El-Omar. 2006. Cellular and molecular aspects of gastric cancer. *World J. Gastroenterol.* **12**:2979–2990.
42. Struelens, M. J. 1996. Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. *Clin. Microb. Infect.* **2**:2–11.
43. Takata, T., R. Aras, D. Tavakoli, T. Ando, A. Z. Olivares, and M. J. Blaser. 2002. Phenotypic and genotypic variation in methylases involved in type II restriction-modification systems in *Helicobacter pylori*. *Nucleic Acids Res.* **30**:2444–2452.
44. Tomb, J. F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzegerald, N. Lee, M. D. Adams, E. K. Hickey, D. E. Berg, J. D. Gocayne, T. R. Utterback, J. D. Peterson, J. M. Kelley, M. D. Cotton, J. M. Weidman, C. Fujii, C. Bowman, L. Watthey, E. Wallin, W. S. Hayes, M. Borodovsky, P. D. Karp, H. O. Smith, C. M. Fraser, and J. C. Venter. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**: 539–547.
45. Vitkute, J., K. Stankevicius, G. Tamulaitiene, Z. Maneliene, A. Timinskas, D. E. Berg, and A. Janulaitis. 2001. Specificities of eleven different DNA methyltransferases of *Helicobacter pylori* strain 26695. *J. Bacteriol.* **183**:443–450.
46. Vovis, G. F., and S. Lacks. 1977. Complementary action of restriction enzymes Endo R.DpnI and Endo R.DpnII on bacteriophage f1 DNA. *J. Mol. Biol.* **115**:525–538.
47. Wong, B. C., W. H. Wang, D. E. Berg, F. M. Fung, K. W. Wong, W. M. Wong, K. C. Lai, C. H. Cho, W. M. Hui, and S. K. Lam. 2001. High prevalence of mixed infections by *Helicobacter pylori* in Hong Kong: metronidazole sensitivity and overall genotype. *Aliment. Pharmacol. Ther.* **15**:493–503.
48. Xu, Q., R. D. Morgan, R. J. Roberts, and M. J. Blaser. 2000. Identification of type II restriction and modification systems in *Helicobacter pylori* reveals their substantial diversity among strains. *Proc. Natl. Acad. Sci. USA* **97**: 9671–9676.
49. Xu, Q., R. M. Peek, Jr., G. G. Miller, and M. J. Blaser. 1997. The *Helicobacter pylori* genome is modified at CATG by product of *hpyIM*. *J. Bacteriol.* **179**:6807–6815.