J Mol Med (1999) 77:834–846
Digital Object Identifier (DOI) 10.1007/s001099900067

REVIEW

Richard A. Alm · Trevor J. Trust

Analysis of the genetic diversity of *Helicobacter pylori*: the tale of two genomes

Received: 13 September 1999 / Accepted: 8 November 1999 / Published online: 4 December 1999

Abstract Infection with *Helicobacter pylori* has been linked to numerous severe gastroduodenal diseases including peptic ulcer and gastric cancer. Several techniques have been used to measure the genetic heterogeneity of *H. pylori* at several different levels and to determine whether there is any correlation with severity of disease. The availability of two completed genome sequences from unrelated strains (J99 and 26695) has allowed an analysis of the level of diversity from a large-scale yet detailed perspective. Although the two chromosomes are organized differently in a limited number of discrete regions, the genome size and gene order of these two "high-virulence" (*cagA*$^+$ and *vacA*$^+$) *H. pylori* isolates was found to be highly similar. The regions of organizational difference are associated with insertion sequences, DNA restriction/modification genes, repeat sequences, or a combination of the above. A significant level of variation at the nucleotide level is seen across the genome, providing an explanation for why the nucleotide-based typing techniques have such high discriminatory power among independent *H. pylori* isolates. This nucleotide variation together with the organizational rearrangements appears to have provided an over-estimation of the gene order diversity of *H. pylori* as assessed by pulse-field gel electrophoresis. Functional assignments are assigned to approximately only 60% of the gene products in each strain, with one-half of the remaining gene products of unknown function having homologues in other bacteria, while the remainder appear to be *H. pylori*-specific. Between 6% and 7% of the coding capacity of each strain are genes that are absent from the other strain, with almost one-half of these strain-specific genes located in a single hypervariable region called the plasticity zone. The majority of the strain-specific genes in each strain are also *H. pylori*-specific, with no homologues being identified in the public databases. Significantly, over one-half of the functionally assigned strain-specific genes in both *H. pylori* J99 and 26695 encode DNA restriction/modification enzymes. Analysis of the level of conservation between orthologues from the two strains indicates that the *H. pylori* specific genes have a lower level of conservation than those orthologues to which a putative function can be assigned. The plasticity zone represents one of several regions across each genome that is comprised of lower (G+C)% content DNA, some of which has been detected in self-replicating plasmids, suggesting that both horizontal transfer from other species and plasmid integration are responsible for the strain-specific diversity at this locus. These analyses have yielded results with important implica-

RICHARD A. ALM received his Ph.D. in molecular microbiology from the University of Adelaide in Adelaide, Australia. He is presently the Genomics Project Leader at the AstraZeneca R&D facility in Boston, Mass., USA, and is involved in the development of novel therapeutic strategies against *Helicobacter pylori*. His research interests include microbial genomics and molecular mechanisms of bacterial pathogenesis.

TREVOR J. TRUST received his Ph.D. in microbiology from the University of Melbourne in Melbourne, Australia. He is currently the Head of Infection Discovery at the AstraZeneca R&D facility in Boston, Mass., USA, which is involved in discovery of novel anti-infectives. His research interests include bacterial pathogenesis and anti-infective drug discovery.

R.A. Alm (✉) · T.J. Trust
AstraZeneca R&D Boston, 128 Sidney Street,
Cambridge MA, USA 02139-4239, USA
e-mail: richard.alm@astrazeneca.com
Tel: +1-617-5763900, Fax: +1-617-5763030

tions for understanding the genetic diversity of *H. pylori* and its associated diseases, and imply a need to reassess the respective roles of bacterial and host factors in *H. pylori* associated diseases.

Key words Diversity · Helicobacter · Genome · Gastritis · Ulcer

Abbreviations *PCR*: Polymerase chain reaction · *PFGE*: Pulse-field gel electrophoresis

Introduction

The discovery in the early 1980s of curved, flagellated bacilli in patients suffering from severe gastroduodenal diseases [1] has had a significant impact on world health. The responsible organism, *Helicobacter pylori*, represents the second most common bacterial infection in the world today, with approximately one-half the world's population being infected, and a rate approaching 100% in some less developed countries [2]. *H. pylori* colonizes the gastric mucosa and is now recognized as the aetiological agent of superficial and chronic gastritis, duodenal ulcer, and many gastric ulcers [3, 4, 5]. Furthermore, the strong association of *H. pylori* infection with gastric cancer and mucosa-associated lymphoid tissue lymphoma has resulted in this organism being classified as a class I carcinogen by the World Health Organization [6]. Indeed, the eradication of *H. pylori* has been shown to reverse low-grade mucosa-associated lymphoid tissue lymphoma [7].

The majority of persons infected with *H. pylori* do not present with clinical symptoms, with fewer than 10% progressing to serious disease. This enigma, together with hypotheses as to the wide spectrum of diseases associated with *H. pylori* infection, has been the cause of considerable debate. The human host-parasite relationship with *H. pylori* is an ancient one, with recent evidence detecting *H. pylori* in stool samples from pre-Columbian mummies which are over 1700 years old [8]. *H. pylori* is extremely well adapted to the human host, as evidenced by its ability to persist chronically in its unique gastric niche, and the finding that the bacterial surface carries carbohydrate structures which are identical to those found on human cells [9, 10]. Further, *H. pylori* is not isolated from other animals except some non-human primates and one colony of felines. The low incidence of severe disease associated with infection has led to the suggestion that there may be 'beneficial' *H. pylori* organisms in addition to those which cause disease, and that global eradication of this organism may not be justified [11, 12]. The clinical benefits of *H. pylori* eradication have been clearly demonstrated with duodenal ulcer patients, although several clinical trials with non-ulcer dyspepsia patients have resulted in conflicting data [13, 14]. In addition, there are reports that the lowering of intragastric acidity by *H. pylori* may protect against gastro-esophageal reflux disease [12], especially in patients producing high levels of gastric acid. *H. pylori* has been reported to be genetically extremely variable, and it has been proposed that this heterogeneity is involved in the ability of *H. pylori* to cause different diseases [15, 16], to play a role in the life-long chronicity of infection [17], and to result in some strains being beneficial to the infected human host [11].

H. pylori was the first bacterial species to have the genome from two independent isolates completely sequenced and compared [18, 19], and this analysis is available on the Internet (http://www.astra-boston.com/hpylori, strain J99 and genome comparison web site and http://www.tigr.org/tdb/CMR/ghp/htmls/Splash-Page.html, strain 26695 web site). *H. pylori* J99 was isolated in 1994 in the United States from a duodenal ulcer patient whereas strain 26695 was isolated in the mid-1980s in the United Kingdom from a patient with gastritis. The comparison of these two genomes should begin to provide a framework for understanding the level and mechanisms of genetic variability in this gastroduodenal pathogen.

Genetic heterogeneity in *H. pylori*

Several techniques examining nucleotide, amino acid and genomic diversity have been used to document the genetic heterogeneity of *H. pylori* (Table 1). Several studies, one using Japanese strains [20] and the other using North American strains [21], have been carried out to demonstrate macrodiversity in *H. pylori* genomes using

Table 1 Evidence of diversity by various methods in *H. pylori*

Type of diversity	Evidence
Macrodiversity	Gene organizational and physical map differences identified by PFGE
Microdiversity	Individual nucleotide changes observed during sequencing of orthologous genes from various strains.
Allelic variation	Global microdiversity measured using DNA-based techniques such as randomly amplified polymorphic DNA PCR and restriction fragment length polymorphism PCR and by multilocus enzyme electrophoresis.
Strain-specific genes	Some strains lack the *cag* pathogenicity island. Subtractive hybridization also demonstrates strain-specific differences.
Insertion elements	Identification of IS*605* and IS*606* at different physical locations or present in different numbers between strains.
Mosaicism	Shown with the vacuolating cytotoxin (*vacA*) gene and the 3' end of the *cagA* gene.

pulse-field gel electrophoresis (PFGE). These studies indicate that essentially all strains that can be digested with the *Not*I restriction endonuclease produce a unique profile, although two strains, UA800 and UA802, isolated from unrelated patients at different times, produce identical restriction patterns [21]. Sequential biopsy specimens taken from the same patient together with laboratory passage of the isolates suggest that the physical genomic organization of *H. pylori* is stable over time both in vivo and in vitro [22]. Physical maps of five *H. pylori* isolates have been generated by PFGE and subsequent hybridization with specific probes [23, 24]. Comparison of these maps suggest that, independent of the source of the strain, the gene order amongst *H. pylori* isolates is highly variable [23].

Population genetic analyses on *H. pylori* using multilocus enzyme electrophoresis to estimate allelic variation of 16 distinct enzyme loci in a total of 97 strains has demonstrated that *H. pylori* has a higher allelic variation than other human or animal pathogens [25, 26]. These data have led to the suggestion that *H. pylori* does not constitute a single species [26]. Two novel insertion elements, IS*605* and IS*606*, have been identified in *H. pylori*, and each contains two genes that appear to contain transposases from unrelated IS elements [27]. These IS elements are present in approximately one-third of *H. pylori* strains tested, often in multiple copies [18, 27], and IS*605* has been associated with the disruption of the *cag* pathogenicity island (*cag*PAI) in *H. pylori* NCTC11637 [28, 29].

The extent of microdiversity and allelic variation within *H. pylori* has been extensively studied using a variety of techniques, and these have been used to develop typing methods for differentiation between clinical isolates. These methods were used to examine specific genes as well as the genome randomly. The method of randomly amplified polymorphic DNA polymerase chain reaction (PCR) has been extremely successful in discriminating large numbers of isolates [30, 31] and has been used to distinguish differences between independent isolates from the same patient [32, 33]. Other DNA-based fingerprinting techniques that have been used to randomly sample the genome are repetitive PCR, restriction fragment end labeling, and oligofingerprinting [34, 35]. Sequencing or PCR restriction fragment length polymorphism of specific genes has also demonstrated a high degree of nucleotide diversity between *H. pylori* strains. Flagellin genes (*flaA a*nd *flaB*), several urease genes (*ureA-D*; the *ureC* and *ureD* genes are now called *lspA a*nd *glmM*, respectively [18, 36]), and ribosomal RNA patterns have been used to type clinical isolates [37, 38, 39, 40, 41, 42], although the level of discrimination depends on the number of enzymes used to generate a particular restriction fragment length polymorphism type. Kansau et al. [43] sequenced an internal portion of the *ureC* gene (now called *glmM*) from 29 clinical isolates and found that each one is unique. Recent sequence analysis of 9 gene fragments from a geographically diverse collection of 20 strains demonstrated that the majority of sequences are unique [44]. Indeed, only at two loci (*efp*: elongation factor EF-P and *ppa*: inorganic pyrophosphatase) were identical sequences seen between two individual strains [44]. These data, together with the fact that no orthologous gene pairs were identical between strains J99 and 26695 [19] suggest that identical gene sequences between *H. pylori* isolates are extremely rare. While all of these methods are highly discriminatory between independent isolates, there is no clear evidence that any particular typing method can be used to reproducibly predict the clinical outcome of a *H. pylori* infection. However, one study using amplification between repetitive DNA segments demonstrated a significant, albeit not complete, clustering between strains isolated from either duodenal ulcer or gastritis patients [45].

The presence of strain-specific genes is another measure of genetic diversity, and the variable clinical outcome of infections caused by *H. pylori* has been associated with strain-to-strain differences in genetic content with corresponding phenotypic diversity. The presence of the *cag*PAI has been associated with an enhanced risk for the development of duodenal ulcers and adenocarcinoma of the distal stomach [29, 46], likely due to the involvement of the gene products in the induction of interleukin-8 secretion from gastric epithelial cells [28]. The *cag*PAI in NCTC11637 is split into two segments, *cag*I and *cag*II, [28, 29] while from J99 and 26695 it is found as one locus [18, 19]. However, recent studies have confirmed that the *cag*PAI is not always present either as a conserved island, or as *cag*I and *cag*II, as deletions can occur within the island [47, 48], and that the clinical outcome of *H. pylori* infection cannot be reliably predicted by the presence or absence of any gene within the *cag*PAI [47]. This supports the evidence that the presence of anti-CagA antibodies does not differentiate between patients with peptic ulcer disease and those with asymptomatic gastritis [49]. In addition, while there is evidence that the majority of strains from Asian countries possess the *cagA* gene irrespective of the associated clinical disease state [48, 50, 51, 52], this finding is not always consistent [53].

The vacuolating cytotoxin gene (*vacA*) has been shown to possess a mosaic structure. Initially, three variations in the signal sequence (s-region; s1a, s1b and s2) and two variations in a middle domain (m-region; m1 and m2) of *vacA* were identified, and the s1 signal sequence was highly correlated with the presence of the *cagA* gene and peptic ulceration [54]. Recently the mosaic diversity in *vacA* has been expanded at both positions to include s1c and m2b variants [55]. The new m2b domain has to date been found exclusively in eastern Asian strains, and only 7 of the 114 strains with the s1c sequence were isolated from outside this region [56]. While other geographic regions also demonstrate a bias towards specific combinations of *vacA* ‘s’ and ‘m’ regions, s1 and the presence of the *cagA* gene remain highly associated [56]. As more data on *cagA a*nd *vacA* genotypes and their association with disease accumulate, it is becoming apparent that generalized conclusions can-

not be drawn, and that geographic differences are common.

Diversity analysis in two complete *H. pylori* genome sequences

The genomic comparison of two completely sequenced *H. pylori* isolates has provided some important information regarding genetic heterogeneity. There is evidence for some genetic loci (*cagA*, *ureC/glmM* and *vacA*) that European and Asian strains represent two distinct lineages of *H. pylori* strains [56, 57, 58]. Using *vacA*, North American strains bear closer resemblance to western European strains [56]. Recent sequence analysis of several other genetic loci from geographically diverse strains also demonstrate distinction of these lineages [44]. It will be interesting to determine whether this delineation holds true for all genes. Does the comparison of J99 and 26695 [19], then, represent a true indication of the diversity within *H. pylori*? It is important to note that many of the studies using the techniques in Table 1 were from single geographical populations. Therefore the comparison of two strains (J99 and 26695) isolated a decade apart on two continents from patients presenting with different diseases would be expected to capture many aspects of genetic diversity within *H. pylori*.

Comparison of the two genomes shows that *H. pylori* J99 is 24,036 bp smaller and contains 57 fewer predicted open reading frames than strain 26695. Each strain contains a set of specific genes which are absent from the other (89 in J99, and 117 in 26695), and in both strains almost half of these are clustered into one locus called a plasticity zone [19]. The majority (approximately 60%) of the strain-specific genes in both J99 and 26695 are also *H. pylori* specific and share no sequence similarity to information in the public databases.

Analysis of *H. pylori* J99 and 26695 PFGE maps

PFGE is an essential tool for construction of physical maps of bacterial genomes and for epidemiological differentiation between strains, and enormous diversity of gene order has been observed between the five published physical maps of *H. pylori* [23]. There are 1406 genes in J99 that have orthologues in 26695, and to align these orthologues requires several artificial translocation and/or inversions. The extent of gene order conservation was determined by comparing each of the 1495 J99 genes to its orthologous 26695 partner (if present) with respect to its immediate neighbouring gene flanking each side. A total of 1267 genes (84.7%) have the same gene located on both sides in both genomes. There are 161 (10.8%) genes where the gene order is disrupted on one side by the insertion/deletion of a strain-specific gene(s) while maintaining the gene order on the other side. Forty genes (2.7%) are flanked by strain-specific genes on both sides, although only one (JHP1295) even possesses an orthologue in strain 26695 at all. Surprisingly, only 27 genes (1.8%) have the same neighbouring gene on one side and are flanked on the other side by a gene common to both strains which is out of order due to an organizational rearrangement such as a translocation and/or inversion. Importantly, the only gene used by Jiang et al. [23] to construct the physical maps that does not have identical neighbouring genes in both J99 and 26695 is *vacA*, as J99 has a single strain-specific gene (JHP820) inserted downstream. The area downstream of *vacA* has been shown to contain a *vapD* homologue in other strains, which has been shown to exhibit positional diversity [59]. In addition, one of the 23S-5S rRNA loci is associated with the plasticity zone, and is flanked by strain-specific genes in both strains. There were nine strings of conserved genes over 50 genes in length, representing 46% of the total number of common genes,

Table 2 Comparison of *Not*I fragment sizes and specific gene localization in the genomes of *H. pylori* strains J99 and 26695 (*boldface* genes not used in the physical maps generated experimentally by Jiang et al. [23] to compare diversity)

*Not*I fragment (relative size)	*H. pylori* strain J99		*H. pylori* strain 26695	
	Calculated size (bp)	Gene localization	Calculated size (bp)	Gene localization
1	361,657	16S rRNA, 23S-5S rRNA, *tsaA*, *lpp20*	435,377	*ureA*, ***fucU***, ***hopA***, *flaB*, *sodB*
2	340,979	*flaA*, *cagA*, *copA*, *gyrB*	427,615	16S rRNA, 23S-5S rRNA, *flaA*, **M. *Hpy*I**, *tsaA*, *lpp20*
3	211,595	*ureA*, *flaB*	397,668	23S-5S rRNA, ***fucT***, *cagA*, *pfr*, *gyrB*
4	137,024	23S-5S rRNA	245,518	16S rRNA, *copA*
5	122,233	***fucT***, *pfr*	74,479	*hpaA*
6	120,364	***hopA***	43,787	*vacA*, ***hopB***
7	82,042	16S rRNA	43,423	*katA*
8	76,944	*hpaA*		
9	61,271	**M. *Hpy*I**		
10	40,240	*katA*		
11	36,497	***fucU***, *sodB*		
12	25,610	***hopB***		
13	22,902	*vacA*		
14	4,473			

with the longest being 133 genes. The limited gene shuffling observed is consistent with a low level of evolutionary divergence within *H. pylori* [19].

The sizes of *Not*I restriction fragments were generated from the complete genomes of *H. pylori* J99 and 26695 (Table 2). *H. pylori* J99 has 14 *Not*I sites, compared to 7 in 26695, which would provide a totally different PFGE restriction pattern. All seven of the 26695 *Not*I sites are also found in the corresponding orthologues in J99. Nucleotide changes in orthologous genes, which fail to affect the sequence of the encoded protein, are responsible for 6 of the additional 7 *Not*I sites in J99, whereas the last difference is due to a single amino acid change. When the positions of 20 representative genetic loci were mapped to specific *Not*I fragments (Table 2), few genes mapped to similar fragments. Given the gene order similarity between *H. pylori* J99 and 26695, it appears that PFGE alone provides an over-estimated view of macrodiversity in *H. pylori*, and any such results should be viewed cautiously. Furthermore, of the ten organizational differences that were required to align the orthologous genes from the two genomes [19], only the two large inversions would have been detected with a *Not*I restriction digest. However, these two inversions, coupled with the additional *Not*I fragments introduced due to minor nucleotide variations in the restriction sites, provide a diverse picture of two genomes which have 85% of their genes preserved in the same genetic order.

Comparison of orthologous sequences between *H. pylori* J99 and 26695

Comparative annotation of the two completed *H. pylori* genomes predicts the functions of 877 (58.7%) and 898 (57.9%) genes for strain J99 and 26695, respectively, and grouped them into a functional hierarchy (Table 3). The remaining genes of unknown function are either *H. pylori*-specific (343 and 364 for J99 and 26695, respectively) or share similarity with other gene products in the database (275 and 290 for J99 and 26695, respectively).

Several *H. pylori* genes have been examined at the nucleotide level among independent strains. Nucleotide variation has been detected in the *cagA, vacA, flaA, flaB, cysS* and *ureC/glmM* genes, as well as the *tnpA/B* genes from IS*605* [43, 60, 61, 62, 63]. This variation is often in the range of 5–10%, and in many cases every isolate possesses a unique sequence.

The extent of nucleotide variation for all the orthologous genes from *H. pylori* J99 and 26695 has been determined, and averages based on functional group designations are presented in Table 3. The average nucleotide identity for all orthologues is 92.6%. However, the average nucleotide identity for the orthologues with a predicted function is higher, at 94.0%. The expected increase in the predicted protein similarity due to 'silent' nucleotide changes is evident in both classes, being 93.4% and 95.4% for all the orthologous proteins and those with a predicted function, respectively. Within the genes that have been assigned a predicted function, the genes (and corresponding proteins) predicted to be involved in DNA restriction and modification display the highest level of divergence with only 90.6% identity, similar to those orthologues that have no predicted function (Table 3). The lower identity level among some of the members in the DNA restriction/modification group suggests that some of the pairs may not be true orthologues but possess different DNA specificities. Of all the designated groups the *H. pylori*-specific genes have the lowest level of identity (Table 3). As a further measure of divergence, the lengths of all the orthologous proteins have also been compared. A total of 162 protein pairs (19%) with a predicted function are different in length, with an average difference of 13.9 residues (Table 3). However, a few protein classes are responsible for the majority of this divergence, including the three large VacA homologues, the Vir homologues, DNA restriction/modification enzymes, and the outer membrane proteins (Table 3). Analysis of the length divergence in the "metabolic" functional groups demonstrates that only 49 protein pairs (6%) differ in length, and these only by an average of 3.6 residues. In contrast, the length divergence of the proteins with no predicted function is an average of 29.4 residues and occurs in 28% of the genes, with the *H. pylori* specific genes contributing the majority of the diversity (Table 3).

The level of nucleotide sequence variation and corresponding protein variation seen between J99 and 26695 would provide the individuality that has been detected with various fingerprinting methods such as PCR with randomly amplified polymorphic DNA, PCR with amplification between repetitive DNA segments, and oligofingerprinting. Analysis of 12 of the enzymes used in multilocus enzyme electrophoresis studies [25, 26] (acid phosphatase, diaphorase, valine-leucine peptidase [26] and indophenol oxidase [25] were not unambiguously identified in either J99 or 26695) demonstrates that the average nucleotide and amino acid identities are 95.3% and 97.4%, respectively (amino acid similarity 98.9%). Moreover, all 12 enzymes are of identical length between the two strains.

Phenotypic diversity

Each of the two strains contains a set of specific genes absent from the other genome (89 in J99 and 117 in 26695), of which almost one-half lie in the hypervariable plasticity zone. Relatively few of these strain-specific genes can be attributed a function (25 and 26 for J99 and 26695, respectively) with the majority (approx. 60%) in both strains encoding DNA restriction/modification enzymes (Table 4). Several strain-specific genes encode products that are likely to be able to alter the complexion of the bacterial cell envelope and subsequently may alter the interaction with the host immune system (Table 4). A comparison of functionally attributed genes in these two

Table 3 Sequence similarity of orthologous genes between two *H. pylori* strains

Functional group and subgroup	Number of genes in			Average identity		Genes of different length	
	J99	26695	Both	nt%	aa%	No.	Av. diff. (aa)
Amino acid biosynthesis							
Aromatic family	14	14	14	94.0	95.5	2	3
Aspartate family	14	14	14	94.8	96.1	1	5
General	6	6	6	94.5	96.3	1	1
Glutamate family	3	3	3	93.9	95.5	0	0
Pyruvate family	2	2	2	94.8	96.0	0	0
Serine family	5	5	5	93.8	96.0	0	0
Totals	44	44	44	94.3	95.9	4	3
Biosynthesis of cofactors							
Biotin	5	5	5	93.9	95.0	1	3
Folic acid	8	7	7	93.4[b]	92.9[b]	2[b]	6.5
Heme and porphyrin	10	10	10	94.0	94.7	2	2
Menaquinone and ubiquinone	5	5	5	94.6	95.7	0	0
Molybdopterin	8	8	8	91.8	91.5	2	4.5
Pantothenate	4	4	4	94.3	95.0	0	0
Pyridine nucleotides	3	3	3	93.8	95.7	0	0
Pyridoxine	3	3	3	94.3	96.6	0	0
Riboflavin	6	6	6	94.5	95.3	0	0
Thiamine	3	3	3	89.4	88.4	1	2
Thioredoxin, glutaredoxin, glutathione	5	5	5	95.7	98.4	0	0
Totals	60	59	59	93.6	94.4	8	3.9
Cell envelope							
Membranes and porins	61	65[c]	60	92.6	93.4	28	24.1
Murein sacculus and peptidoglycan	16	16	16	93.9	95.2	3	10.7
Surface polysaccharides and antigens	41	40[d]	38	92.4	92.5	14	10.4
Surface structures	42	43[e]	42	93.6	95.7	7	17.6
Totals	160	164	156	92.9	93.9	52	18.8
Cellular processes							
Cell division	9	9	9	94.2	96.2	3	4.7
Cell killing	5	6[f]	5	91.3	90.5	4	37.3
Chaperones	8	8	8	95.1	96.2	1	2
Detoxification	4	4	4	95.9	98.1	0	0
DNA transfer and Cag island	35	38[g]	33	93.4	93.8	7	39.3
General	9	10	9	93.6	94.9	1	45
Phosphorus compounds	2	2	2	96.4	98.0	0	0
Polyamine biosynthesis	3	3	3	95.1	96.8	0	0
Protein and peptide secretion[h]	13	13	13	94.2	95.8	5	2.2
Insertion elements	2	14	0	[i]	[i]	[i]	[i]
Urea	6	6	6	96.4	98.5	1	2
Total	96	113	92	94.1	95.1	22	22.8
DNA replication							
Totals	23	23	21	94.1	95.1	8	5.5
DNA restriction, modification, recombination and repair							
Totals	67	68	51	90.6	90.6	25	21.1
Energy metabolism							
Amino acids and amines	7	6	6	95.5	96.7	1	2
ATP-proton motive force interconversion	9	9	9	95.6	97.6	0	0
Electron transport	52	52	52	95.0	97.1	4	3.5
Entner-Doudoroff	2	2	2	94.7	97.1	0	0
Fermentation	4	4[j]	3	92.6	93.9	3	3
Gluconeogenesis	10	10	10	95.3	97.1	0	0
Other	6	7	6	94.7	95.4	0	0
Pentose phosphate pathway	5	5	5	94.3	95.5	0	0
Sugars	1	1	1	95.8	98.5	0	0
TCA cycle	8	8	8	95.8	98.1	1	2
Totals	104	104	102	95.0	96.9	9	3
Fatty acid and phospholipid metabolism	28	29	28	95.0	96.6	3	3.7

Table 3 (continued)

Functional group and subgroup	Number of genes in			Average identity		Genes of different length	
	J99	26695	Both	nt%	aa%	No.	Av. diff. (aa)
Purines, pyrimidine, nucleoside and nucleotide biosynthesis							
2'-Deoxyribonucleotide metabolism	4	4	4	95.5	98.4	0	0
Purine ribonucleotide biosynthesis	3	3	3	95.0	96.7	0	0
Pyrimidine ribonucleotide biosynthesis	11	11	11	95.0	96.6	0	0
Salvage and interconversion	13	13	13	94.9	96.3	0	0
Sugar-nucleotide biosynthesis and conversions	3	3	3	94.6	96.2	0	0
Totals	34	34	34	95.0	96.7	0	0
Regulatory functions							
General	22	22[k]	21[k]	94.2	96.2	3	4.3
Chemotaxis and motility	10	10	10	94.7	96.8	2	6.5
Totals	32	32	31	94.4	96.4	5	5.2
Transcription							
Degradation of RNA	3	3	3	94.4	94.8	0	0
DNA-dependent RNA polymerase	3	3	3	96.0	97.9	1	2
RNA processing	2	2	2	93.4	93.3	0	0
Transcription factors	5	5	5	96.5	98.9	0	0
Totals	13	13	13	95.4	96.9	1	2
Translation							
Amino acyl tRNA synthetases	21	21	21	94.4	95.7	4	3.5
Degradation of proteins and peptides	23	23	23	94.3	96.0	5	3
Nucleoproteins	2	2	2	94.0	95.4	0	0
Protein modification	5	5	5	94.6	97.1	0	0
Ribosomal proteins: synthesis and modification	55	55	55	95.8	98.3	2	4
Translation factors	12	12	12	95.5	97.8	2	3
tRNA modification	10	10	10	92.9	92.9	3	2.3
Totals	128	128	128	95.0	96.9	16	3.1
Transport and binding proteins							
Amino acids, peptides and amines	18	18	18	94.8	97.0	2	2
Anions	4	4	4	94.9	96.2	0	0
Carbohydrates, alcohols and acids	6	6	6	94.4	96.7	1	11
Cations	26	26	26	94.5	96.6	4	5
General	33[l]	32	32	94.3	96.4	2	4.5
Nucleosides, purines and pyrimidines	1	1	1	94.1	95.5	0	0
Totals	88	87	87	94.5	96.6	9	4.9
Conserved with no function	275	290	267	92.5	93.3	62	21.0
H. pylori specific with no function	343	364	288	88.4	87.3	95	34.9
Totals for predicted function	877	898	846	94.0	95.4	162	13.9
Total for no known function	618	654	555	90.4	90.2	157	29.4
Total	1495	1552	1401[m]	92.6	93.4	319	21.5

[a] Relative to the number of JHP genes
[b] Does not include JHP862 which represents a partial duplication of the *folE* gene
[c] Includes eight HP genes which make up four JHP genes (73, 238, 851, 1138).
[d] Includes two HP genes which make up JHP86
[e] Includes two HP genes which make up JHP625
[f] Includes two HP genes which make up JHP556
[g] Includes four HP genes which make up JHP35 and JHP36
[h] Does not include the recently identified JHP1126A (HP1203A) gene which displays similarity to the functionally required domains of the *E. coli* SecE protein [64]. The *secE* gene and its predicted protein display 95.5% and 96.6% identity, respectively, between the J99 and 26695 strains, and both proteins are 59 amino acid residues in length
[i] See text; J99 has one IS606 copy, while 26695 has two IS606 copies and five IS605 copies. The orthologous transposase genes in IS606 were not included in this analysis
[j] Includes two HP genes which make up JHP841
[k] Includes two HP genes which make up JHP151, J99 has two duplicate orthologues of HP1365
[l] Includes an additional partial duplication of HP0818 in J99
[m] The 94 gene difference represents the 89 J99-specific genes, three duplications and *tnpA/tnpB* from IS*606*

strains has suggested that they would be physiologically quite similar [64]. Importantly, the genomic locations of the J99- or 26695-specific genes are very similar. In approximately one-half of the loci in which one strain contains a specific gene(s), the other strain also contains a specific gene(s), with orthologous flanking genes [19], suggesting that the *H. pylori* genome has limited regions where these genes can be located. Other *H. pylori* strains would be expected to contain their own set of 'specific' genes, possibly both in a plasticity zone and elsewhere

Table 4 Distribution of *H. pylori* J99 and 26695 strain specific genes

Annotation	*H. pylori* J99	*H. pylori* 26695
H. pylori specific with no known function	56	69
Conserved hypothetical with no known function	8	22
Genes with predicted function	(25)[a]	(26)[a]
DNA restriction/modification	15	16
DNA replication	2	2
Cell envelope	4	2
Cellular processes	2	4
Energy metabolism	2	1
Phospholipid metabolism	0	1
Total strain-specific genes	89	117

[a] The total genes with predicted function are shown in their predicted functional groups

on the chromosome, and this has been shown experimentally in an additional *H. pylori* strain using subtractive hybridization [65].

Mechanisms of diversity generation in *H. pylori*

While not all the mechanisms of generating diversity in *H. pylori* can be identified from the genome analyses, there are several important features identified from the comparison, and they can be separated into genetic and antigenic diversity. Genetic diversity can arise from within the *H. pylori* species or by acquisition of heterologous DNA from other species. There are several areas of each *H. pylori* genome, including the plasticity zones, the *cag*PAI, and several DNA restriction/modification genes, which are of significantly different (G+C)% than the remainder of the genome [19]. All of the ten organizational differences which are required to align the orthologous genes from J99 and 26695 are associated with either an insertion element, a strain-specific DNA restriction/modification gene, a repeat element, or a combination of these.

The IS*605* element was first associated with a physical rearrangement of the *cag*PAI in strain NCTC11638 [28, 29]. The prevalence of this element is found in *H. pylori* irrespective of the *cag* status of the strain and is located, often in multiple copies, at various locations on the chromosome [60]. While *H. pylori* 26695 contains five complete copies of IS*605*, strain J99 lacks a complete copy [19]. Three of these elements are in the plasticity zone of 26695, whereas the other two are located at positions of organizational rearrangements between the two strains [19]. Further, both strains contain IS*605* fragments containing the left and/or right dyadic repeats. Both J99 and 26695 have complete copies of the related element IS*606* [18, 19, 27], although the single full copy in J99 is present in a different location than either of the two copies in strain 26695, and none is associated with any organizational differences. Both IS*606* transposase genes (*tnpA a*nd *tnpB*) in J99 differ from their counterparts in 26695. The J99 *tnpB* gene, as with that of *H. pylori* strain 84–183 (GenBank U95957), lacks the truncating frameshift seen in both 26695 copies. In addition, the J99 *tnpA* gene product is 74 codons shorter at the N-terminal end than both 26695 copies and the copy in 84–183. There are five cases of coincidence in the location of partial or complete copies of insertion elements in the two sequenced strains, suggesting either that certain loci are more receptive to insertion sequences, or that these elements are inherited from a common progenitor strain. The J99, but not the 26695, sequence contains an 80-basepair fragment of an IS*606* element transposase gene near the 31-basepair repeat on one side of the *cag*PAI and a 100-basepair fragment from another portion of the same IS606 gene near the other 31-basepair repeat. One of these fragments is found at the same location in strain NCTC11638. The location of these IS606 transposase fragments, i.e., adjacent to the 31-basepair repeats but proximal to the *cag*PAI genes, suggests that the *cag*PAI originated by horizontal transfer into, or by means of, an IS*606*-related element.

Both J99 and 26695 contain numerous strain-specific DNA restriction and modification genes (Table 4), and these are associated with four of the ten organizational differences between the strains. Moreover, many of these strain-specific genes have a lower (G+C)% DNA content, which may be consistent with them being acquired horizontally from other species. The large number of these genes present in *H. pylori* (approx. 3% of each genome), their maintenance in the genome, and their association with genomic rearrangements is consistent with the proposal that these systems can act as selfish genetic elements [66]. Moreover, the strain specificity of these genes (and presumably the specificity of the recognition sequences) could represent both a consequence of, and a factor promoting, high levels of recombination between shorter DNA segments, similar to that proposed by Milkman [67] for natural recombination in *Escherichia coli*. There are also several repeated areas in the *H. pylori* genome apart from the IS elements, and intragenomic recombination between them could give rise to organizational differences. The gene content of independent isolates is also a measure of genetic diversity, and this can be influenced by plasmids, either self-replicating or integrated into the chromosome. In support of this hypothesis, the highly variable plasticity zone of the *H. pylori* 26695 genome contains DNA sequence essentially identical to that found on a plasmid in an independent isolate [19].

Analysis of both completed *H. pylori* genomes indicates that several enzymes in the DNA mismatch repair

system are missing, such as MutL and MutH, suggesting that *H. pylori* has less stringent control of replicative errors. However, the absence of this system in other completed microbial genomes suggests that it is dispensable. It remains a possibility that a less effective mismatch repair system leads to the higher degree of nucleotide divergence seen between *H. pylori* isolates than in other species. However, a large degree of the nucleotide variation does not affect the encoded proteins that are under functionally selective constraints.

H. pylori is naturally competent for transformation in vitro [68], which allows for uptake of DNA and subsequent recombination/integration into the chromosome. Such exchange of genetic material has the ability to rapidly increase the level of genetic diversity by generating new combinations of sequence polymorphisms. A recent report [69] has provided the first evidence that recombination can occur in vivo, when genetic exchanges between two variants of *H. pylori* in a natural infection were identified. Indeed, it has been shown, using several genetic markers, that recombination in *H. pylori* is so common that each locus has undergone extremely frequent recombination and is in linkage equilibrium [44, 61]. Such promiscuous recombination among *H. pylori* strains causes considerable diversity, especially at the nucleotide level, but it remains to be demonstrated whether this phenomenon allows the pathogen to adapt to and to cause different diseases in, individual hosts. However, the apparent maintenance of geographically separate clonal groupings suggests that selection for certain alleles of particular genes exist.

The identification of homopolymeric tracts and dinucleotide repeats (and in some cases, associated frameshifts) in several genes in the genomes of strains J99 and 26695 has led to the prediction that "slipped-strand repair mechanisms" modulate gene expression. Indeed, differences in the length of these polymorphic regions were first identified during the sequencing of the J99 genome [19] (Fig. 1A). Such a mechanism could be involved in antigenic variation and adaptive evolution. Almost one-half of the genes containing these polymorphic sites are involved in the composition of the outer cell envelope of *H. pylori*, and differential expression of these genes would alter the appearance of the organism to the host immune system (Fig. 1B). The best studied examples of this phenomenon are the fucosyltransferase genes, involved in the final stages of lipopolysaccharide biosynthesis. The two copies of the α-1,3 fucosyltransferase genes and the single copy of the α-1,2 fucosyltransferase gene in *H. pylori* are involved in the synthesis of Lewis X and Y antigens of lipopolysaccharide, and may contribute to antigenic mimicry and autoimmune disease [9, 10]. There is a long homopolymeric C tract in the 5' coding region of each of the four previously sequenced *H. pylori* α-1,3 fucosyltransferase genes [18, 70, 71] and in the central domain of the α-1,2 fucosyltransferase gene [72, 73]. The serotype of lipopolysaccharide-phase variants of strain NCTC11637 and of several clinical isolates is correlated with the varying length

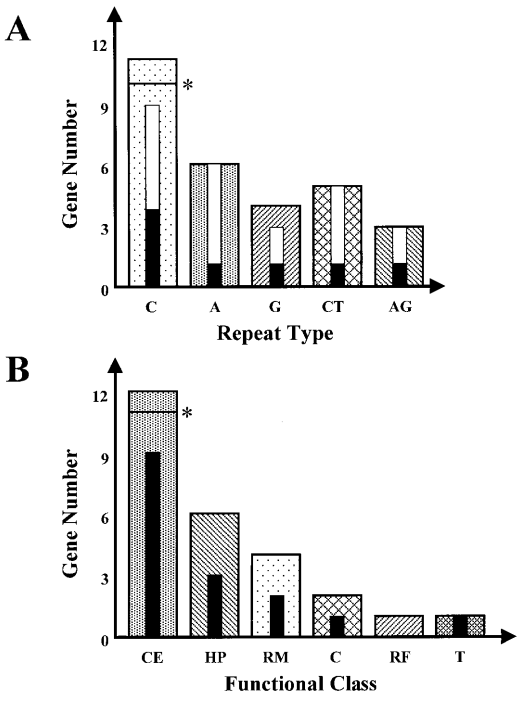

Fig. 1A,B Analysis of the *H. pylori* genes whose regulation may be regulated by slipped-strand repair. **A** *Large patterned bars* the number of genes in *H. pylori* J99 and 26695 with a particular type of mono- or dinucleotide repeat. Three genes contained two separate repeats [19], and in these cases both repeated areas were counted. *H. pylori* 26695 lacks an orthologue of JHP820 (C repeat); *asterisk* total gene number for this strain (10). *Height of white box (within each patterned bar)* the number of these genes whose repeat length differ between the orthologues present in J99 and 26695; *level of filling of white box* the number of genes whose repeats have been experimentally demonstrated to be polymorphic during the sequencing of the J99 genome. **B** *Large patterned bars* functional groups (as described in Table 3) of the genes that contain the mono- or dinucleotide repeats. *CE* Cell envelope; *HP H. pylori* specific with no function; *RM* DNA restriction/modification, recombination and repair; *C* conserved with no function; *RF* regulatory functions; *T* transport and binding proteins. *H. pylori* 26695 lacks an orthologue of JHP820; *asterisk* the total number of Cell envelope genes for this strain (11). *Height of solid box (within each patterned bar)* the number of orthologous genes whose predicted expression status (in-frame or out-of frame) are the same in the two *H. pylori* strains, regardless of the number of repeats

of the homopolymeric tract and the resulting expression status of these three fucosyltransferase genes [74]. In both *H. pylori* J99 and 26695 the α-1,2 fucosyltransferase coding region appears prematurely terminated due to the C string tract length, but is full length in strain UA802 [73].

Five members of the large outer membrane protein superfamily in both strains J99 and 26695 contain dinucleotide CT repeats in the coding region for the signal sequence (Fig. 1A). In each case the number of CT dinucleotide repeats in the two strains differ, but the predicted expression status is nevertheless unaffected [19]. Three of the genes are out of frame, and two, including the Lewis *babB* adhesin [75], are in frame.

Clinical relevance of diversity in *H. pylori* infection

Infection with *H. pylori* causes a wide spectrum of disease in humans, and attempts have been made to establish a correlation between these and genetic diversity. Although the presence of the *cag*PAI has been associated with an increased risk of peptic ulcers or stomach cancers in some European populations, this association appears to not be consistent in all populations (see above). Bacterial, host, and environmental factors can influence the susceptibility to and the clinical outcome of *H. pylori* infection, and since the discovery of VacA and the *cag*PAI significant emphasis has been placed on the identification of additional *H. pylori* factors that are involved in the colonization and disease process, or that can be correlated with various types of disease. From the bacterial perspective, several "diversity" factors could influence the outcome of infection.

Firstly, differences in gene expression could affect the clinical outcome. However, J99 and 26695 each possess only three identifiable σ factors (only one of which, σ^{54}, is known to be susceptible to environmental regulation), six independent two-component response regulators, and four transcriptional regulators with helix-turn-helix motifs, suggesting that regulation of gene expression in *H. pylori* is simpler than other bacterial pathogens. However, there are a small number of biologically relevant genes that are differentially expressed as a result of slippage in homopolymeric tracts or dinucleotide repeats. This method of regulation could affect the ability of the organism to colonize, persist, and produce disease. Furthermore, there are nine type II methyltransferases that appear conserved between the two strains and yet lack an identifiable restriction subunit partner, suggesting that *H. pylori* uses methylation as a method of controlling gene expression in a manner similar to that employed by *E. coli* to regulate pilin expression [76]. Furthermore, several genes in both strains possess one or more in-frame translational termination codons or frameshifts that would likely inactivate the gene. The presence of these 'pseudogenes' may represent either another method by which *H. pylori* can alter gene expression or simply unwanted genes undergoing genetic decay.

Secondly, the strain-specific genes are likely also to play a role in the pathophysiology of the diseases caused by *H. pylori* infection. The presence of the highly variable gene content in the plasticity zone may play an important role in this regard, although, as with J99 and 26695, there would be strain-specific genes outside the plasticity zone in other *H. pylori* strains. Although the majority of strain-specific genes with a predicted function resemble restriction/modification enzymes that may not directly participate in the disease process, there are a significant number of strain-specific genes of unknown function that may play a critical role in progression of disease or chronicity of infection. None of the strain-specific genes, at least in the two strains whose genomes have been sequenced, appear to have homologues known to be virulence factors in other pathogenic bacteria. The carriage and/or integration of plasmids also represent an important source of genetic material that may play a crucial role in genetic diversity and the outcome of the associated disease.

Finally, the genomic comparison, together with accumulating epidemiological evidence, suggests that host factor(s) may play a significant, and perhaps unappreciated, role in the susceptibility, severity, and outcome of a *H. pylori* infection. Different mice strains display markedly different susceptibility to *H. pylori* colonization and to clinical outcome, for example, in the case of gastric atrophy [77, 78]. There is also evidence of differences in susceptibility in certain human populations. For example, in Singapore's multiethnic Chinese, Malay and Indian population the *H. pylori* seroprevalence in healthy volunteers has been determined to be 79%, 13% and 34%, respectively [79]. In addition, the Malay subpopulation exhibit a lower incidence rate of gastric cancer among those infected [79]. These findings suggest either a genetic predisposition or a socio-cultural difference between the various ethnic groups.

Concluding remarks

H. pylori appears unique among bacteria in terms of the diversity it shows between strains, although the level of gene order diversity by PFGE appears to have been overestimated. As more intraspecies genomic comparisons are made in the future, the level of allelic or nucleotide diversity observed within *H. pylori* will be placed in better perspective. An initial report comparing two *Mycobacterium tuberculosis* strains suggests that a single nucleotide polymorphism occurs approximately every 5 kb (0.02%) [80], which is two to three orders of magnitude lower than that observed in *H. pylori*. Years of passage of an organism can propagate mutations even within a single strain. An example of this is seen by a comparison of approximately 55 kb of DNA sequence from 12 independent GenBank entries which used *Pseudomonas aeruginosa* PAO1, a laboratory standard for over one-quarter of a century, to the PAO1 genome sequencing project currently being performed (www.pseudomonas.com) which demonstrated a nucleotide divergence rate of 0.5%. Therefore, given the chronicity of infection with *H. pylori* and the potential rate of mutation combined with the promiscuous recombination, care needs to be taken when defining a 'different strain' as opposed to a quasi-species or subtype. The understanding of the mechanisms of diversity generation together with future identification of specific genes that are correlated with specific diseases will help elucidate the role of *H. pylori* in gastroduodenal disease. These mechanisms of generating diversity may also found to be responsible, at least in part, for the ability of *H. pylori* to survive and chronically persist in the harsh gastric environment. *H. pylori* pathogenesis is clearly a complex phenomenon, and the balance and interaction between host and bacteria, or the impact of environmental factors, should not be overlooked.

Acknowledgements We thank colleagues at both AstraZeneca R&D Boston and Genome Therapeutics Corporation for many stimulating discussions.

References

1. Marshall BJ, Warren JR (1984) Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet I:1311–1315
2. Matysiak-Budnik T, Megraud F (1997) Epidemiology of *Helicobacter pylori* infection with special reference to professional risk. J Physiol Pharmacol 48 [Suppl 4]:3–17
3. Cover TL, Blaser MJ (1992) *Helicobacter pylori* and gastroduodenal disease. Annu Rev Med 42:135–145
4. Hunt RH (1996) The role of *Helicobacter pylori* in pathogenesis: the spectrum of clinical outcomes. Scand J Gastroenterol Suppl 220:3–9
5. Labigne A, de Reuse H (1996) Determinants of *Helicobacter pylori* pathogenicity. Infect Agents Dis 5:191–202
6. Eidt S, Stolte M (1995) The significance of *Helicobacter pylori* in relation to gastric cancer and lymphoma. Eur J Gastroenterol Hepatol 7:318–321
7. Schmitt-Graff A (1996) Immunological and molecular classification of mucosa-associated lymphoid tissue lymphoma. Recent Results Cancer Res 142:121–36
8. Correa P, Willis D, Allison MJ, Gerszten E (1998) *Helicobacter pylori* in pre-Columbian mummies. Gastroenterology 114:A956
9. Monteiro MA, Chan KH, Rasko DA, Taylor DE, Zheng PY, Appelmelk BJ, Wirth HP, Yang M, Blaser MJ, Hynes SO, Moran AP, Perry MB (1998) Simultaneous expression of type 1 and type 2 Lewis blood group antigens by *Helicobacter pylori* lipopolysaccharides. Molecular mimicry between *H. pylori* lipopolysaccharides and human gastric epithelial cell surface glycoforms. J Biol Chem 273:11533–11543
10. Appelmelk BJ, Simoons-Smit I, Negrini R, Moran AP, Aspinall GO, Forte JG, Vries Td, Quan H, Verboom T, Maaskant JJ, Ghiara P, Kuipers EJ, Bloemena E, Tadema TM, Townsend RR, Tyagarajan K, Crothers Jr JM, Monteiro MA, Savio A, de Graaff J (1996) Potential role of molecular mimicry between *Helicobacter pylori* lipopolysaccharide and host Lewis blood group antigens in autoimmunity. Infect Immun 64:2031–2040
11. Blaser MJ (1997) Not all *Helicobacter pylori* strains are created equal: should all be eliminated? Lancet 349:1020–1022
12. Richter JE, Falk GW, Vaezi MF (1998) *Helicobacter pylori* and gastroesophageal reflux disease: the bug may not be all bad. Am J Gastroenterol 93:1800–1802
13. McColl K, Murray L, El-Omar E, Dickson A, El-Nujumi A, Wirz A, Kelman A, Penny C, Knill-Jones R, Hilditch T (1998) Symptomatic benefit from eradicating *Helicobacter pylori* infection in patients with nonulcer dyspepsia. N Engl J Med 339:1869–1874
14. Blum AL, Talley NJ, O'Morain C, van Zanten SV, Labenz J, Stolte M, Louw JA, Stubberod A, Theodors A, Sundin M, Bolling-Sternevald E, Junghard O (1998) Lack of effect of treating *Helicobacter pylori* infection in patients with nonulcer dyspepsia. Omeprazole plus Clarithromycin and Amoxicillin Effect One Year after Treatment (OCAY) Study Group. N Engl J Med 339:1875–1881
15. Atherton JC, Peek Jr RM, Tham KT, Cover TL, Blaser MJ (1997) Clinical and pathological importance of heterogeneity in *vacA*, the vacuolating cytotoxin gene of *Helicobacter pylori*. Gastroenterology 112:92–99
16. Dubois A, Berg DE, Incecik ET, Fiala N, Heman-Ackah LM, Perez-Perez GI, Blaser MJ (1996) Transient and persistent infection of non-human primates with *Helicobacter pylori*: implications for human disease. Infect Immun 64:2885–2891
17. Logan RPH, Berg DE (1996) Genetic diversity of *Helicobacter pylori.* Lancet 348:1462–1463
18. Tomb J-F, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA, Nelson K, Quackenbush J, Zhou L, Kirkness EF, Peterson S, Loftus B, Richardson D, Dodson R, Khalak HG, Glodek A, McKenney K, Fitzegerald LM, Lee N, Adams MD, Hickey EK, Berg DE, Gocayne JD, Utterback TR, Peterson JD, Kelley JM, Cotton MD, Weidman JM, Fujii C, Bowman C, Watthey L, Wallin E, Hayes WS, Borodovsky M, Karp PD, Smith HO, Fraser CM, Venter JC (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature 388:5 39–547
19. Alm RA, Ling LL, Moir DT, King BL, Brown ED, Jiang Q, Doig PC, Smith DR, Noonan B, Guild BC, de Jonge BL, Carmel G, Tummino PJ, Caruso A, Uria-Nickelsen M, Mills DM, Ives C, Merberg D, Mills SD, Taylor DE, Vovis GF, Trust TJ (1999) Genomic sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. Nature 397:186–190
20. Takami S, Hayashi T, Tonokatsu Y, Shimoyama T, Tamura T (1993) Chromosomal heterogeneity of *Helicobacter pylori* isolates by pulsed-field gel electrophoresis. Zentralbl Bakteriol 280:120–127
21. Taylor DE, Eaton M, Chang N, Salama SM (1992) Construction of a *Helicobacter pylori* genome map and demonstration of diversity at the genome level. J Bacteriol 174:6800–6806
22. Salama SM, Jiang Q, Chang N, Sherbaniuk RW, Taylor DE (1995) Characterization of chromosomal DNA profiles from *Helicobacter pylori* strains isolated from sequential gastric biopsy specimens. J Clin Microbiol 33:2496–2497
23. Jiang Q, Hiratsuka K, Taylor DE (1996) Variability of gene order in different *Helicobacter pylori* strains contributes to genome diversity. Mol Microbiol 20:833–842
24. Bukanov NO, Berg DE (1994) Ordered cosmid library and high-resolution physical-genetic map of *Helicobacter pylori* strain NCTC11638. Mol Microbiol 11:509–523
25. Go MF, Kapur V, Graham DY, Musser JM (1996) Population genetic analysis of *Helicobacter pylori* by multilocus enzyme electrophoresis: extensive allelic diversity and recombinational population structure. J Bacteriol 178:3934–3938
26. Hazell SL, Andrews RH, Mitchell HM, Daskalopoulous G (1997) Genetic relationship among isolates of *Helicobacter pylori*: evidence for the existence of a *Helicobacter pylori* species complex. FEMS Microbiol Lett 150:27–32
27. Kersulyte D, Akopyants NS, Clifton SW, Roe BA, Berg DE (1998) Novel sequence organization and insertion specificity of IS*605* and IS*606*: chimaeric transposable elements of *Helicobacter pylori*. Gene 223:175–186
28. Censini S, Lange C, Xiang Z, Crabtree JE, Ghiara P, Borodovsky M, Rappuoli R, Covacci A (1996) *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. Proc Natl Acad Sci USA 93:14648–14653
29. Akopyants NS, Clifton SW, Kersulyte D, Crabtree JE, Youree BE, Reece CA, Bukanov NO, Drazek ES, Roe BA, Berg DE (1998) Analyses of the *cag* pathogenicity island of *Helicobacter pylori*. Mol Microbiol 28:37–53
30. Akopyanz N, Bukanov NO, Westblom TU, Kresovich S, Berg DE (1992) DNA diversity among clinical isolates of *Helicobacter pylori* detected by PCR-based RAPD fingerprinting. Nucleic Acids Res 20:5137–5142
31. Han J, Yu E, Lee I, Lee Y (1997) Diversity among clinical isolates of *Helicobacter pylori* in Korea. Mol Cells 7:544–547
32. Kersulyte D, Chalkauskas H, Berg DE (1999) Emergence of recombinant strains of *Helicobacter pylori* during human infection. Mol Microbiol 31:31–43
33. Berg DE, Gilman RH, Lelwala-Guruge J, Srivastava K, Valdez Y, Watanabe J, Miyagi J, Akopyants NS, Ramirez-Ramos A, Yoshiwara TH, Recavarren S, Leon-Barua R (1997) *Helicobacter pylori* populations in Peruvian patients. Clin Infect Dis 25:996–1002
34. Doorn NEM van, Namavar F, Kusters JG, van Rees EP, Kuipers EJ, de Graaff J (1998) Genomic DNA fingerprinting

of clinical isolates of *Helicobacter pylori* by REP-PCR and restriction fragment end-labelling. FEMS Microbiol Lett 160: 145–150
35. Marshall DG, Coleman DC, Sullivan DJ, Xia H, O'Morain CA, Symth CJ (1996) Genomic DNA fingerprinting of clinical isolates of *Helicobacter pylori* using short oligonucleotide probes containing repetitive sequences. J Appl Bacteriol 81:509–517
36. Reuse H de, Labigne A, Mengin-Lecreulx D (1997) The *Helicobacter pylori ureC* gene codes for a phosphoglucosamine mutase. J Bacteriol 179:3488–3493
37. Owen RJ, Hunton C, Bickley J, Moreno M, Linton D (1992) Ribosomal RNA gene restriction patterns of *Helicobacter pylori*: analysis and appraisal of *Hae*III digests as a molecular typing system. Epidemiol Infect 109:35–47
38. Tee W, Lambert J, Smallwood R, Schembri M, Ross BC, Dwyer B (1992) Ribotyping of *Helicobacter pylori* from clinical specimens. J Clin Microbiol 30:1562–1567
39. Akopyanz N, Bukanov NO, Westblom TU, Berg DE (1992) PCR-based RFLP analysis of DNA sequence diversity in the gastric pathogen *Helicobacter pylori*. Nucleic Acids Res 20: 6221–6225
40. Forbes KJ, Fang Z, Pennington TH (1995) Allelic variation in the *Helicobacter pylori* flagellin genes *flaA a*nd *flaB*: its consequences for strain typing schemes and population structure. Epidemiol Infect 114:257–266
41. Shortridge VD, Stone GG, Flamm RK, Beyer J, Versalovic J, Graham DY, Tanaka SK (1997) Molecular typing of *Helicobacter pylori* isolates from a multicenter U.S. clinical trial by *ureC* restriction fragment length polymorphism. J Clin Microbiol 35:471–473
42. Ito A, Fujioka T, Kubota T, Nasu M (1996) Molecular typing of *Helicobacter pylori*: differences in pathogenicity among diverse strains. J Gastroenterol 31:1–5
43. Kansau I, Raymond J, Bingen E, Courcoux P, Kalach N, Bergeret M, Braimi N, Dupont C, Labigne A (1996) Genotyping of *Helicobacter pylori* isolates by sequencing of PCR products and comparison with the RAPD technique. Res Microbiol 147:661–669
44. Achtman M, Azuma T, Berg DE, Ito Y, Morelli G, Pan ZJ, Suerbaum S, Thompson SA, van der Ende A, van Doorn LJ (1999) Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. Mol Microbiol 32:459–470
45. Go MF, Chan KY, Versalovic J, Koeuth T, Graham DY, Lupski JR (1995) Cluster analysis of *Helicobacter pylori* genomic DNA fingerprints suggests gastroduodenal disease-specific associations. Scand J Gastroenterol 30:640–646
46. Figura N (1997) *Helicobacter pylori* factors involved in the development of gastroduodenal mucosal damage and ulceration. J Clin Gastroenterol 25:S149–S163
47. Jenks PJ, Megraud F, Labigne A (1998) Clinical outcome after infection with *Helicobacter pylori* does not appear to be reliably predicted by the presence of any of the genes of the *cag* pathogenicity island. Gut 43:752–758
48. Maeda S, Yoshida H, Ikenoue T, Ogura K, Kanai F, Kato N, Shiratori Y, Omata M (1999) Structure of *cag* pathogenicity island in Japanese *Helicobacter pylori* isolates. Gut 44:336–341
49. Graham DY, Genta RM, Graham DP, Crabtree JE (1996) Serum CagA antibodies in asymptomatic subjects and patients with peptic ulcer: lack of correlation of IgG antibody in patients with peptic ulcer or asymptomatic *Helicobacter pylori* gastritis. J Clin Pathol 49:829–832
50. Pan ZJ, Hulst RWvd, Feller M, Xiao SD, Tytgat GN, Dankert J, van der Ende A (1997) Equally high prevalences of infection with *cagA*-positive *Helicobacter pylori* in Chinese patients with peptic ulcer disease and those with chronic gastritis-associated dyspepsia. J Clin Microbiol 35:1344–1347
51. Miehlke S, Kibler K, Kim JG, Figura N, Small SM, Graham DY, Go MF (1996) Allelic variation in the *cagA* gene of *Helicobacter pylori* obtained from Korea compared to the United States. Am J Gastroenterol 91:1322–1325
52. Mitchell HM, Hazell SL, Li YY, Hu PJ (1996) Serological response to specific *Helicobacter pylori* antigens: antibody against CagA antigen is not predictive of gastric cancer in a developing country. Gastroenterology 91:1785–1788
53. Ito A, Fujioka T, Kodama K, Nishizono A, Nasu M (1997) Virulence-associated genes as markers of strain diversity in *Helicobacter pylori* infection. J Gastroenterol Hepatol 12:666–669
54. Atherton JC, Cao P, Peek Jr RM, Tummuru MKR, Blaser MJ, Cover TL (1995) Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. J Biol Chem 270:17771–17777
55. Doorn LJ van, Figueiredo C, Sanna R, Pena S, Midolo P, Ng EK, Atherton JC, Blaser MJ, Quint WG (1998) Expanding allelic diversity of *Helicobacter pylori vac*A. J Clin Microbiol 36:2597–2603
56. Doorn LJ van, Figueiredo C, Megraud F, Pena S, Midolo P, Queiroz DMDM, Carneiro F, Vanderborght B, Pegado MD, Sanna R, Boer WD, Schneeberger PM, Correa P, Ng EK, Atherton J, Blaser MJ, Quint WG (1999) Geographic distribution of *vacA a*llelic types of *Helicobacter pylori*. Gastroenterology 116:823–830
57. Ende A van der, Pan ZJ, Bart A, van der Hulst RW, Feller M, Xiao SD, Tytgat GN, Dankert J (1998) *cagA*-positive *Helicobacter pylori* populations in China and The Netherlands are distinct. Infect Immun 66:1822–1826
58. Ito Y, Azuma T, Ito S, Suto H, Miyaji H, Yamazaki Y, Kohli Y, Kuriyama M (1998) Full-length sequence analysis of the *vacA* gene from cytotoxic and noncytotoxic *Helicobacter pylori*. J Infect Dis 178:1391–1398
59. Cao P, Cover TL (1997) High-level genetic diversity in the *vapD* chromosomal region of *Helicobacter pylori*. J Bacteriol 179:2852–2856
60. Hook-Nikanne J, Berg DE, Peek Jr RM, Kersulyte D, Tummuru MK, Blaser MJ (1998) DNA sequence conservation and diversity in transposable element IS605 of *Helicobacter pylori*. Helicobacter 3:79–85
61. Suerbaum S, Smith JM, Bapumia K, Morelli G, Smith NH, Kunstmann E, Dyrek I, Achtman M (1998) Free recombination within *Helicobacter pylori.* Proc Natl Acad Sci USA 95:12619–12624
62. Garner JA, Cover TL (1995) Analysis of genetic diversity in cytotoxin-producing and non-cytotoxin-producing *Helicobacter pylori* strains. J Infect Dis 172:290–293
63. Evans Jr DJ, Queiroz DM, Mendes EN, Evans DG (1998) Diversity in the variable region of *Helicobacter pylori cagA* gene involves more than simple repetition of a 102-nucleotide sequence. Biochem Biophys Res Commun 245:780–784
64. Doig P, de Jonge BL, Alm RA, Brown ED, Uria-Nickelsen M, Noonan B, Mills SD, Tummino PJ Carmel G, Guild BC, Moir DT, Vovis GF, Trust TJ (1999) *Helicobacter pylori* physiology predicted from a genomic comparison of two strains. Microbiol Mol Biol Rev 63:675–707
65. Akopyants N, Fradkov A, Diatchenko L, Hill J, Siebert P, Lukyanov S, Sverdlov E, Berg D (1998) PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori.* Proc Natl Acad Sci USA 95:13108–13113
66. Kobayashi I (1998) Selfishness and death: raison d'etre of restriction, recombination and mitochondria. Trends Genet 14:368–374
67. Milkman R (1997) Recombination and population structure in *Escherichia coli*. Genetics 146:745–750
68. Nedenskov-Sorensen P, Bukholm G, Bovre K (1990) Natural competence for genetic transformation in *Campylobacter pylori*. J Infect Dis 161:365–366
69. Kersulyte D, Chalkauskas H, Berg DE (1999) Emergence of recombinant strains of *Helicobacter pylori* during human infection. Mol Microbiol 31:31–43
70. Ge Z, Chan NW, Palcic MM, Taylor DE (1997) Cloning and heterologous expression of an alpha1,3-fucosyltransferase gene from the gastric pathogen *Helicobacter pylori.* J Biol Chem 272:21357–21363

71. Martin SL, Edbrooke MR, Hodgman TC, van den Eijnden DH, Bird MI (1997) Lewis X biosynthesis in *Helicobacter pylori*. Molecular cloning of an alpha (1,3)-fucosyltransferase gene. J Biol Chem 272:21349–21356
72. Saunders NJ, Peden JF, Hood DW, Moxon ER (1998) Simple sequence repeats in the *Helicobacter pylori* genome. Mol Microbiol 27:1091–1098
73. Wang G, Rasko DA, Sherburne R, Taylor DE (1999) Molecular genetic basis for the variable expression of Lewis Y antigen in *Helicobacter pylori*: analysis of the a (1,2) fucosyltransferase gene. Mol Microbiol 31:1265–1274
74. Appelmelk BJ, Martin SL, Monteiro MA, Clayton CA, McColm AA, Zheng P, Verboom T, Maaskant JJ, van den Eijnden DH, Hokke CH, Perry MB, Vandenbroucke-Grauls CM, Kusters JG (1999) Phase variation in Helicobacter pylori lipopolysaccharide due to changes in the lengths of poly (C) tracts in alpha3-fucosyltransferase genes. Infect Immun 67: 5361–5366
75. Ilver D, Arnqvist A, Ogren J, Frick IM, Kersulyte D, Incecik ET, Berg DE, Covacci A, Engstrand L, Boren T (1998) *Helicobacter pylori* adhesin binding fucosylated histo-blood group antigens revealed by retagging. Science 279:373–377
76. Blyn LB, Braaten BA, Low DA (1990) Regulation of pap pilin phase variation by a mechanism involving differential dam methylation states. EMBO J 9:4045–4054
77. Lee A, O'Rourke J, DeUngria MC, Robertson B, Daskalopoulos G, Dixon MF (1997) A standardized mouse model of *Helicobacter pylori* infection: introducing the Sydney strain. Gastroenterology 112:1386–1397
78. Sakagami T, Dixon M, O'Rourke J, Howlett R, Alderuccio F, Vella J, Shimoyama T, Lee A (1996) Atrophic gastrities changes in both *Helicobacter felis* and *Helicobacter pylori* infected mice are host dependent and separate from antral gastritis. Gut 39:639–648
79. Fock KM, Khor CJL, Goh KT, Teoh YL, Seow A, Chia KS, Ho B, Ng HC (1998) Seroprevalence of *Helicobacter pylori* infection. Gastroenterology 114:A596
80. Fleischmann RD, White O, Hickey E, Clayton R, Dodson R, Gwinn M, Heidelberg J, Peterson J, McDonald L, Utterback T, Bishai W, Delcher A, Fraser C, Venter JC (1998) Sequencing of the *M. tuberculosis* genome; comparison of a recent clinical isolate with the laboratory strain. Microb Comp Genomics 3:C14