Helicobacter pylori – a conundrum of genetic diversity

David G. Marshall, William G. Dundon, Sarah M. Beesley and Cyril J. Smyth

Author for correspondence: Cyril J. Smyth. Tel: +353 1 608 1195. Fax: +353 1 679 9294. e-mail: csmyth@tcd.ie

Department of Microbiology, Moyne Institute of Preventive Medicine, Trinity College, University of Dublin, Dublin 2, Republic of Ireland

Keywords: Helicobacter pylori, genetic diversity, horizontal transfer, natural transformation, repetitive DNA

Background

In the past 15 years the recognition of Helicobacter pylori infection as a cause of gastroduodenal morbidity has probably had more impact on health care worldwide than any other emerging infection barring HIV. H. pylori is one of the commonest human gastrointestinal pathogens, infecting almost 50% of the human population (Taylor & Blaser, 1991). Since its rediscovery as curved bacteria in the stomach by histological examination of gastric biopsies (Warren & Marshall, 1983; Marshall, 1989; Rathbone & Heatley, 1989b; Buckley & O’Morain, 1998) and its subsequent first isolation by Marshall & Warren (1984), H. pylori (a microaerophilic, Gram-negative spiral bacterium) has been shown to be associated with chronic superficial gastritis (inflammation of the gastric mucosa), chronic active gastritis (gastric mucosal inflammation with polymorphonuclear cell infiltration), peptic ulcer disease (gastric and duodenal ulcers) and gastric cancer (NIH Consensus Development Panel, 1994). Infection of the stomach with H. pylori appears to occur in early infancy. The bacterium colonizes the gastric mucosa and causes acute gastritis that progresses into chronic active gastritis, a condition that appears to persist for life (Taylor & Parsonnet, 1995). During the years or decades that follow the initial infection, the chronic active gastritis may remain asymptomatic, or may progress into more severe forms of the disease, such as a peptic ulcer, or the almost complete loss of gastric epithelial cell function observed in atrophic gastritis (NIH Consensus Development Panel, 1994).

A recent survey to analyse trends in microbiological research revealed that approximately 7.5% of all publications between 1991 and 1997 were concerned with the genus Helicobacter and that the annual publication rate on Helicobacter had tripled over that period (Galvez et al., 1998). Indeed, several books devoted entirely to H. pylori have been published in the last decade (Blaser, 1989; Rathbone & Heatley, 1989a; Malfertheiner & Ditschunet, 1990; Marshall et al., 1991; Goodwin & Worsley, 1993; O’Morain & O’Connor, 1994; Calam, 1996; Hunt & Tytgat, 1996; Lee & Mégraud, 1996; Moran & O’Morain, 1997; Farthing & Patchett, 1998), as well as comprehensive reviews in journals (e.g. Lee & Hazell, 1988; Dunn et al., 1997), numerous topical articles (e.g. Suerbaum, 1995; Covacci et al., 1997a; Kuipers & Blaser, 1998) and a journal launched in 1996, namely, Helicobacter.

One of the most intriguing aspects of H. pylori is its genetic diversity, which is at a level as yet unseen among other bacterial pathogens, the biological significance of which is enigmatic (Blaser, 1996; Logan & Berg, 1996; Covacci & Rappuoli, 1998). Digestive organs with acid-secreting epithelia, i.e. stomachs, arose over 300 million years ago in vertebrate evolution. Blaser (1996) has suggested that the diversity of H. pylori possibly reflects the diversity of humans, their ancestors, their environments and their diets, which has led to a state of balanced polymorphism in the H. pylori population in a genetically diverse human population during life-long infection of individuals. This review aims to describe aspects of the genetic diversity and genomic malleability of H. pylori.

The H. pylori genome sequence

A complete genome sequence of H. pylori was published by Tomb et al. (1997). The authors chose to sequence the genome of H. pylori strain 26695 which had been originally isolated from a patient with gastritis. This strain colonizes piglets, elicits an immune and inflammatory response and is also known to be toxigenic, plasmid-free and naturally competent (Tomb et al., 1997). It contains a single circular chromosome of approximately 1.7 Mb with 1590 predicted ORFs, one third of which are of unknown function (for review see Berg et al., 1997a). Analysis of this genome sequence...
may reveal important information on the genome structure of this H. pylori strain and perhaps provide clues as to how the genetic variability within this species may occur. However, despite the wealth of information that has already been gleaned and may be further gathered from the sequence of the H. pylori genome, unfortunately the genome sequence of strain 26695 does not hold all the answers. To fully understand the genetic diversity of H. pylori, a second if not a third genome should be sequenced and released into the public domain. The genome of a second strain, J99, has been sequenced in large part, but is licensed to the pharmaceutical company Astra (Hancock et al., 1998). This has allowed those with access to both genomic sequences to make some interesting observations on pairs of genes encoding outer-membrane proteins (Hancock et al., 1998). Indeed, Hancock et al. (1998) have suggested from comparative analysis of the complete genome sequence of strain 26695 and the incomplete genome sequence of strain J99 that many of the H. pylori genes are highly conserved.

Since the release of the genome sequence of strain 26695 it has become apparent that the assignment of roles to putative ORFs in H. pylori, and indeed in other genomes, must be performed with caution as, firstly, most of the putative functions have not as yet been subjected to biochemical or mutational confirmation, thereby overlooking the possibility of the evolution of a new function for the gene product, and, secondly, ORF assignment can vary depending on the system used for analysis of the genome sequence (Berg et al., 1997a; Saunders et al., 1998). Berg et al. (1997a) cite, for example, the tricarboxylic acid cycle enzyme 2-oxoglutarate oxidoreductase, which is present in H. pylori but is not listed among the ORFs in the genome sequence. Furthermore, the analysis of the genome sequence by Saunders et al. (1998) used a different strategy to that of Tomb et al. (1997) and in doing so revealed clearly identifiable differences in ORF assignment.

**Physical comparison of genomes**

**H. pylori**

Macrodiversity within a species is described as gross genetic differences between strains, a phenomenon that can be readily seen in H. pylori. Using various molecular techniques to compare the genomic DNA of unique isolates, i.e. isolates recovered from gastric biopsy samples from different individuals, data supporting the macrodiversity of H. pylori have been obtained.

The extent of genomic DNA sequence similarity among H. pylori isolates has been investigated by Yoshimura et al. (1993) and Takami et al. (1994) using thermal denaturation analysis. Both groups reported similar findings and calculated the DNA sequence similarity between the majority of isolates to be above 80%. However, a number of isolates had sequence similarities of less than 80%, a finding that Takami et al. (1994) postulated to be evidence for the existence of H. pylori subspecies.

The use of PFGE has allowed analysis of H. pylori genomes for macrorestriction patterns by digestion of DNA with restriction endonucleases that cut genomic DNA infrequently. The genome sizes in H. pylori have been assessed by this technique and were reported to vary from 1.0 Mb to 1.8 Mb (Takami et al., 1993). In addition, the presence of plasmids in H. pylori isolates is not uniform, with carriage rates reported to vary between 35 to 59% of isolates (e.g. Simor et al., 1990; Kleanthous et al., 1991; Xiang et al., 1995).

The degree of conservation of gene location between the genomes of isolates has also been determined using PFGE, which, in combination with Southern hybridization with gene probes, generates information about gene order within a strain, thereby allowing a map of the genetic loci to be deduced. In all, four H. pylori strains have been mapped by PFGE, namely, strain UA802 which was mapped by Taylor et al. (1992) and a further three strains, NCTC 11637, NCTC 11639 and UA861, which were mapped by Jiang et al. (1996). An ordered cosmid library of strain NCTC 11638 was made by Bukanov & Berg (1994), which, in combination with PFGE, yielded a high-resolution physical genetic map of this strain. The gene orders of all five isolates were compared by Jiang et al. (1996), revealing that there is little conservation of gene location among the strains, with some genes being found in the same 25% of the genome but with little maintenance of their order. This inter-strain variability in gene location is in marked contrast, for example, to the inter-genus conservation of gene order observed between *Escherichia coli* and *Salmonella typhimurium* (Krawiec & Riley, 1990). The genome of *Neisseria gonorrhoeae* also displays considerable variability but, unlike H. pylori, this diversity is not manifest in variability in genome arrangement (Dempsey et al., 1995).

**Macrodiversity among other Helicobacter genomes**

Comparison of the level of genomic diversity of H. pylori with that of other Helicobacter spp. by PFGE revealed that *Helicobacter hepaticus* has a level of genomic variability which is intermediate between that of H. pylori and *Helicobacter mustelae* (Saunders et al., 1997). *H. mustelae* appears not to exhibit the heterogeneity present in H. pylori, although minor differences in PFGE pattern are present (Taylor et al., 1994). Why the profound genomic diversity of H. pylori is not present across the entire Helicobacter genus is an interesting question.

**Genotypic analysis of H. pylori populations**

**DNA fingerprinting of H. pylori**

A number of genotyping methods have been applied to H. pylori for epidemiological analysis. Random amplification of polymorphic DNA analysis (RAPD)
(Akopyanz et al., 1992), oligofingerprinting (Marshall et al., 1996) and repetitive element-PCR (Rep-PCR) (Go et al., 1995) are the most discriminatory fingerprinting methods applied to *H. pylori*. All can easily distinguish unrelated isolates by the fingerprint profiles they generate. Other DNA-fingerprinting methods, namely, HindIII-RFLP (Langenberg et al., 1986), HaelIII-ribotyping (Owen et al., 1992), PCR-RFLP analysis of fla (Forbes et al., 1995) and ure genes (Foxall et al., 1992; Fujimoto et al., 1994; Shortridge et al., 1997), have also been frequently applied to the differentiation of *H. pylori* strains, but fail to discriminate all epidemiologically unrelated isolates tested. From the DNA-fingerprinting studies applied to *H. pylori*, it is clear that isolates display a level of heterogeneity in fingerprint profiles unseen in other bacteria, with isolates recovered from different individuals generally having different fingerprint profiles by most methods.

**Evidence for multiple strain colonization**

The use of the above techniques on isolates from multiple biopsies recovered from colonized individuals has revealed data on the proportion of individuals colonized by more than a single strain. Colonization of individuals with more than a single strain appears to be rare among Europeans and North Americans. For example, Taylor et al. (1995), using RAPD, found more than one colonizing strain in only three out of 15 patients tested, Marshall (1996), using oligofingerprinting, found two strains in only one out of 19 patients screened, and Shortridge et al. (1997), using PCR-RFLP of ureC genes, found multiple strains in only two out of 81 patients examined. In comparison, in developing countries, where colonization with *H. pylori* is more common, up to 34% of individuals tested have been found to be colonized with multiple strains (Berg et al., 1997b).

**Substrain diversity revealed by genotypic analysis – evidence for microevolution**

When the most discriminatory DNA-fingerprinting methods are used to examine isolates recovered from different biopsy sites within individuals, i.e. on epidemiologically related isolates, very similar but not identical fingerprint profiles are often observed, suggesting that rearrangements are occurring in the genomes of strains during infection. The first evidence to support the occurrence of genomic rearrangements in *H. pylori* came from HindIII-RFLP analysis of isolates recovered from multiple biopsies obtained from individuals, where the profiles generated for approximately 33% of isolates displayed subtypic variations (Oudbier et al., 1990). Furthermore, using oligofingerprinting, up to 60% of individuals tested were found to be colonized with subtypic isolates (Marshall, 1996). Moreover, when 10 biopsies were recovered from each patient, Jorgensen et al. (1996) found that 16 out of 17 individuals tested displayed subtypic *H. pylori* populations. Several studies have looked at the fingerprint profiles of multiple single colony isolates from the same biopsy specimen using RFLP or RAPD analysis (Oudbier et al., 1990; Marshall et al., 1995; Jorgensen et al., 1996). These studies have all found a homogeneous population to be present at a single site except for Jorgensen et al. (1996), who suggested that substrains may be present in a small proportion of biopsies. HaelIII-ribotyping analysis of isolates recovered from members of the same family has also revealed subtypic isolate profiles in different individuals, suggesting that a common clone may have initially infected the individuals (Nwokolo et al., 1992).

The minor band differences obtained by the most discriminatory techniques in fingerprint profiles of epidemiologically related *H. pylori* isolates probably represent rearrangements occurring in vivo in the genomic organization within single clones of *H. pylori* (Oudbier et al., 1990; Jorgensen et al., 1996; Marshall, 1996). Such changes in genetic loci between clones over a relatively short period of time are referred to as microevolution (Achtman, 1994; Hobbs et al., 1994). These events probably occur in most, if not all, *H. pylori* strains colonizing individuals and are, therefore, likely to be a major contributory factor to the heterogeneous nature of the genomic organization of this species. It appears likely that different strains undergo microevolution in isolation in different individuals which would undoubtedly contribute to the extensive genetic diversity present between isolates.

The discovery of transposable elements in *H. pylori* (IS605; see section on cag-pathogenicity island below) is of interest in this regard (Censini et al., 1996; Akopyants et al., 1998; Höök-Nikanne et al., 1998). Movement of insertion sequences (ISs) in a genome occurs with about the same frequency as other mutational events (Berg & Berg, 1996). Höök-Nikanne et al. (1998) have suggested that tests for the copy number and chromosomal distribution of these IS elements could be useful in tracing the microevolution of a strain during long-term carriage in individual hosts and in tracking the transmission of a strain among epidemiologically linked individuals.

**Population structure of *H. pylori***

**Natural variability in *H. pylori***

Population genetics is the study of the natural variability of bacterial populations and has led to the formulation of theories to account for this variability. By using multilocus enzyme electrophoresis (MLEE) data, the frequency of recombination in natural populations can be estimated by calculating the index of association (I_A) value between loci. *H. pylori* has an I_A value that does not differ significantly from zero and, therefore, has a panmictic or freely recombining population structure (Go et al., 1996), with genetic diversity in the sample tested exceeding the level of diversity recorded in all other bacterial species studied by MLEE. Non-clonal bacterial species, like *H. pylori*, are unlikely to show the genetic stability of clones. There is, therefore, little
rationale in attempting to compare the genetic relatedness of epidemiologically unrelated isolates of non-clonal species by phylogenetic trees produced from DNA-fingerprinting profiles (Maynard Smith, 1995). Accordingly, if meaningful conclusions are to be drawn from such data, studies of isolate populations should probably be restricted to epidemiologically related isolates.

Quasi-species

The possibility that within H. pylori there may exist subspecies or what has also been termed a species-complex or quasi-species has been put forward because of the relatively low level of clonality observed in this species (Hazell et al., 1997; Covacci & Rappuoli, 1998). While such hypotheses are valid, a study by Eckloff et al. (1994), which examined 16S rDNA sequences, classified H. pylori as a unique species. They found that the sequence of 16S rDNA, the defining molecule of a species, among five H. pylori isolates showed 0.2-0.5% divergence. In comparison the 16S rDNA sequences of other Helicobacter species revealed divergences between species which ranged from 27 to 80%.

Genetic diversity in virulence-associated loci

Virulence-associated loci

The association of H. pylori infection with the pathogenesis of a spectrum of diseases has been recognized since its isolation but remains poorly understood. The disparate outcome of infection in different individuals suggests the existence of strains of H. pylori with enhanced virulence potential and the possibility of the presence in some strains of virulence-associated loci (Atherton, 1997). The contribution of host and environmental factors to the outcome of H. pylori-associated disease cannot be discounted, but is beyond the scope of this review (Riccardi & Rotter, 1994; Pérez-Pérez & Parisonnet, 1994; Kuipers & Blaser, 1998). Accordingly, the variance present among H. pylori isolates at the level of (i) carriage of specific genes by sub-groups of isolates and their reported associations with clinical outcome and (ii) the genetic diversity at the level of gene structure and gene expression of specific genes is highlighted below.

vacA

The vacA gene which encodes the vacuolating cytotoxin is present in all H. pylori strains but active toxin is produced by only 50% as determined by the ability of strains to induce vacuolation of epithelial cells in vitro (Cover & Blaser, 1992). It was thought up to very recently that lack of expression of the vacuolating cytotoxin in 50% of strains was due to sequence divergence in the vacA genes of cytotoxin-negative strains which had approximately 65% nucleotide sequence identity with the corresponding loci in cytotoxin-producing strains (Cover et al., 1994). However, Forsyth et al. (1998) have shown that the vacA gene is transcribed in all strains, but that transcription occurs at higher levels in cytotoxin-positive strains than in cytotoxin-negative strains. Therefore, they suggest that differences in vacA transcription levels among H. pylori strains may be a factor that contributes to different vacuolating cytotoxin phenotypes (Forsyth et al., 1998).

In addition to this inter-strain heterogeneity in VacA expression, the vacA alleles also exhibit a type of genotypic variation known as mosaicism (Fig. 1a), with conserved regions and highly divergent regions occurring within these alleles (Atherton et al., 1995). Three signal sequence types exist, namely, s1a, s1b and s2. Until recently, all mid-regions were classified as types m1 and m2. However, additional variants have been identified in German, designated m1a (Strobel et al., 1998), in Taiwan, designated m1T and m1Tm2 (a chimaeric type) (Wang et al., 1998), and in China, designated m1b and m1b–m2 (a hybrid type) (Pan et al., 1998), suggesting that vacA typing schemes may need to be revised.

There is much evidence for a strong correlation between the vacA genotype of a strain and disease outcome. Strains of the signal sequence/mid-region combination type s1/m1 were identified by Atherton et al. (1995) as producing high levels of the vacuolating cytotoxin in vitro and were associated with increased gastric epithelial damage, enhanced gastric inflammation and duodenal ulceration (Atherton et al., 1997). Furthermore, strains of signal sequence type s1 are associated with the presence of the cagA gene (see below) and numerous studies have highlighted the prevalence of type s1 strains in individuals with peptic ulcer disease (Rudi et al., 1998; Strobel et al., 1998; van Doorn et al., 1998). Interestingly, strains of signal sequence type s1a and mid-region type m1 were isolated with remarkable prevalence (almost 100%) in Japan (Ito et al., 1997), although no correlation with clinical consequences was reported. The same high prevalence of type s1a strains was also found in Taiwan and China, but, in contrast to Japan, the dominant mid-region type was m2 (Pan et al., 1998; Wang et al., 1998). However, strains with the type m1b and m1b–m2 alleles produced significantly higher levels of the VacA cytotoxin than did those with type m2 alleles (Pan et al., 1998) and strains with the type m1T allele occurred statistically more often in peptic ulcer disease patients than did strains with type m2 alleles (Wang et al., 1998).

fla and ure

The flagella and urease of H. pylori are important factors involved in the initial colonization of the gastric mucosa (Josenhans & Suerbaum, 1997; Mobley et al., 1997) and are present in all H. pylori isolates. The gene sequences which encode the flagellin, flaA and flaB, showed extensive sequence variability, as evidenced by RFLPs when PCR products were digested with restriction endonucleases (Ohata-Tada et al., 1997). Forbes et al. (1995) presented evidence that the sequence diversity found in the flaA and flaB genes had resulted
from reassortment of sequences between strains in the bacterial population. Similarly, the urease structural genes ureA, ureB and ureC exhibit extensive restriction site diversity, which has been exploited for the molecular typing of strains (Foxall et al., 1992; Fujimoto et al., 1994; Shortridge et al., 1997).

**babA**

The BabA adhesin of *H. pylori*, which binds to the fucosylated Lewis b (Leb) histo-blood group antigen, is a putative virulence factor and is strongly associated with the presence of the *cag*-pathogenicity island (see below). The gene encoding BabA was recently cloned by Ilver et al. (1998). Three bab alleles have been identified (Fig. 1b), but it is not yet known whether bab alleles are present in all *H. pylori* strains. Two of these alleles, namely babA2 and babB, encode functional proteins. However, the third allele babA1 is identical to the babA2 allele with the exception of a 10 bp deletion of a repeat motif in the signal peptide sequence, which results in ablation of the translational initiation codon. This suggests that hot spots for phenotypic (phase) variation exist within the bab gene family. The bab genes belong to a family of over 30 *H. pylori* genes which have been proposed to encode outer-membrane proteins showing extensive amino acid sequence identity in the N-terminal and C-terminal domains, suggesting the potential for recombinational events leading to mosaicism (Tomb et al., 1997; Ilver et al., 1998).

**cagA**

*H. pylori* strains can be arbitrarily divided into six groups, based on whether they possess the cag genetic locus and express the CagA (cytotoxin-associated antigen) protein and active VacA toxin (Xiang et al., 1995). Isolates carrying the cag genetic locus are more commonly recovered from patients with duodenal ulcers and gastric tumours (Blaser et al., 1995). However, studies from Japan, Korea and China have shown that more than 90% of *H. pylori* strains were cagA+ irrespective of clinical presentation (Miehlke et al., 1996; Pan et al., 1997; Shimoyama et al., 1997; Maeda et al., 1998).

The *cagA* gene is part of the *cag*-pathogenicity island (see following section) and encodes a highly immunogenic outer-membrane protein of variable molecular mass, 120–140 kDa, and of unknown function (Covacci et al., 1993; Tummuru et al., 1993). Neither expression of the VacA toxin nor induction of interleukin 8 production was affected by inactivation of the *cagA* gene, suggesting that CagA may only be a marker for increased virulence (Tummuru et al., 1994; Crabtree et al., 1995). The *cagA* gene has a highly conserved 5′ region, while variation in the size of the CagA protein has been correlated with the presence of a variable number of repeat sequences encoded by the 3′ region of the *cagA* gene (Covacci et al., 1993; Tummuru et al., 1993).

Analysis of the 3′ sequences of 155 *cagA*-gene-positive isolates identified four types of *cagA* structure, designated A, B, C and D, based on their primary sequences and structural organizations (Fig. 1c; Yamaoka et al., 1998). Interestingly, part of the amino
acid sequence of the R3 region of the CagA protein in isolates from Japan, KIASAGKVGFGSGA, was different from that of strains from other geographical areas, FPLKRDVKDDLSKV (Covacci et al., 1993; Tummuru et al., 1993). Type C strains were associated with more severe degrees of gastric atrophy, higher levels of anti-CagA antibody and gastric cancer. The findings suggested that the cagA genotype may provide a useful marker for differences in virulence among cagA-positive isolates (Yamaoka et al., 1998).

**cag-pathogenicity island**

In an attempt to more fully understand the relationship between the CagA protein and virulence, Censini et al. (1996) analysed the DNA regions flanking the cagA gene. This led to the discovery of the cag-pathogenicity island (cag-PAI), recently confirmed by the extensive study of Akopyants et al. (1998).

The cag-PAI has many of the features of classic pathogenicity islands (Censini et al., 1996; Covacci et al., 1997b; Hacker et al., 1997), namely, (i) a G+C content of 35 mol% that differs from the mean of 39 mol% for the rest of the *H. pylori* genome, (ii) flanking direct repeats involved in the integration of the 40 kb locus into the genome, (iii) putative virulence genes and genes that encode a type IV secretion system and (iv) insertion sequences.

The nucleotide sequencing of the 40 kb cag-PAI revealed the presence of a novel insertion sequence, IS605, which consists of two genes, tnpA and tnpB, that, based on the amino acid sequence similarities of the translated gene products, are thought to encode transposases (Censini et al., 1996) (Fig. 2). These genes are flanked by two nucleotide sequences of 33 bp and 41 bp, respectively, that possess dyadic symmetry and a common core sequence. IS605 was shown to be present in 56% of the 44 clinical isolates tested and was found in varying copy number within these strains, inserted both into the cag-PAI and into sites on the rest of the genome. The maximal number of copies of IS605 was associated with cag-PAI+ strains, while cag-PAI- strains did not appear to possess the IS605 element (Censini et al., 1996). However, Akopyants et al. (1998) found that 9 out of 16 cag-PAI- strains lacked IS605 and Höök-Nikanne et al. (1998) reported that the prevalence of IS605 was similar in 37 cag-positive (27%) and in 13 cag-negative (23%) strains. The presence of IS605 did not correlate with production of the VacA toxin, with the m-type and s-type of vacA alleles, or with clinical manifestations of *H. pylori* infection (Höök-Nikanne et al., 1998). From one to nine copies of IS605 have been detected per genome (Censini et al., 1996; Akopyants et al., 1998; Höök-Nikanne et al., 1998). Interestingly, five full-length copies of IS605 occur in the genome sequence of strain 26695, but none of them is associated with the cag-PAI (Tomb et al., 1997).

The structure of the cag-PAI varies among strains (Fig. 3): (i) it can consist of a single uninterrupted unit; (ii) it can be split into two regions, cagl and cagII, by IS605; (iii) it can be split into cagl and cagII by a large piece of chromosomal DNA that is flanked by IS605; and (iv)–(vi) it can consist of partially deleted segments of cagl, cagII or both (Censini et al., 1996). This variation in the structure of the cag-PAI is most likely due to the insertion of IS605 followed by homologous recombination, which may explain the existence of cag-PAI+ and cag-PAI- strains and their intermediates (Censini et al., 1996; Covacci et al., 1997a).

The identification of the cag-PAI and its strong association with a functional *vacA* gene have led to some hypotheses on the evolution of cag-PAI+ *H. pylori*. It has been suggested that the inheritance of the cag-PAI resulted in the generation of a functional vacuolating toxin from a pre-existing molecule by a gain-of-function mechanism, thereby allowing for a more virulent clone, which could then out-compete less virulent strains in the stomach of the host (Covacci et al., 1997a). It has been reported that in some cases both cagA-positive and cagA-negative strains can be isolated from the same individual (van der Ende et al., 1996). These findings have previously been presented as evidence for multiple infections. However, it is now thought that a dynamic equilibrium may exist between the gain and loss of the cag-PAI (Covacci et al., 1997a; Covacci & Rappuoli, 1998).

**Strain-variable genetic loci**

In addition to the cag-PAI several novel genetic loci, which are only carried by a proportion of strains, have been identified. A recently identified region containing an ORF related to the virulence-associated protein D (VapD) from *Dichelobacter nodosus* was situated 3.5 kb downstream from the *vacA* gene in strain 60190 (Cao & Cover, 1997). In *H. pylori*, the *vapD* gene has been detected in approximately 61% of isolates. The presence of the *vapD* gene is not associated with a specific *vacA* allelle type (Cao & Cover, 1997).
The search for genes that are possibly induced by contact with epithelial cells led to the identification by Peek et al. (1996) of a locus known as ice (induced by contact with epithelium). There are two distinct ice subgroups, namely, iceA1 and iceA2. Only iceA1 expression increased following adherence to epithelium. The iceA1 gene product showed 65% identity with a restriction endonuclease, NalIII of Neisseria lactamica, suggesting a putative function (Peek et al., 1996; Xu et al., 1997). Strains of the iceA1 sub-group have been shown to be strongly associated with duodenal ulcer disease, while iceA2 strains were present at higher frequency in individuals with gastritis (Peek et al., 1996; van Doorn et al., 1998b). The iceA allelic type was independent of the cagA and vacA status (van Doorn et al., 1998b).

A genetic locus of unknown function designated trl (for tRNA-associated locus, formerly known as del), which was isolated using arbitrarily primed PCR of RNA, was only present in 48% of clinical isolates tested (Dundon et al., 1998). No correlation between the presence or absence of trl and the caga/Caga or vacA/VacA status or the clinical origin of strains has been revealed (Dundon et al., 1998).

**Phase-variable genes**

A number of homopolymeric tracts and dinucleotide repeat regions have been identified from analysis of the *H. pylori* genome sequence, many of which are within the promoter regions or ORFs of putative genes. Slipped strand mispairing of these repeat sequences within coding sequences may result in the loss or gain of a unit, such that the ORF may be in or out of frame with an upstream translation initiation codon, or may result in mutations preventing formation of a functional protein (Levinson & Gutman, 1987). The presence of the homopolymeric tracts and dinucleotide repeat regions may, therefore, facilitate the phase variation of the *H. pylori* ORFs they are associated with. Similar repeat sequences are presumed to be sites for slipped strand mispairing in *Haemophilus influenzae* (Moxon et al., 1994; Hood et al., 1996).

The *H. pylori* genome study by Tomb et al. (1997) identified a total of 30 genes with either dinucleotide repeats or homopolymeric tracts associated with them (Table 1). Eight of these genes, five of which encode outer-membrane proteins, have dinucleotide repeats at their 5′ end, suggesting that a mechanism for phase variation may exist in *H. pylori* (Saunders et al., 1998). There is also a suggestion from Tomb et al. (1997) that phenotypic variation may occur at the transcriptional level for 27 genes that have homopolymeric tracts in their promoter regions. Analysis of the genome sequence by Saunders et al. (1998), however, revealed nine such additional putative phase-variable genes (Table 1). These genes, along with the 27 described previously by Tomb et al. (1997), can be divided into five groups: (i) five are putatively involved in LPS biosynthesis; (ii) 14 encode putative cell-surface-associated proteins; (iii) six are part of the DNA restriction/modification system; (iv) six are of unknown function, showing no sequence similarity with genes or gene products in the public domain nucleotide databases; and (v) five encode miscellaneous proteins. However, Saunders et al. (1998) recommend caution when ascribing phase-variable functions based on the presence of homopolymeric tracts alone. They highlight that the absence of a number of the genes involved in the SOS response results in mismatch repair in *H. pylori* and may explain the prevalence of long homopolymeric tracts in the genome.

The O-antigen of the LPS of 85% of *H. pylori* isolates contains the tissue/blood group structure Lewis x (Le\(^x\)) or Lewis y (Le\(^y\)) or both, similar to those commonly occurring in human gastric mucosa (Aspinall et al., 1994; Aspinall & Monteiro, 1996). This molecular mimicry between *H. pylori* LPS and host Lewis antigens is thought to have a role in autoimmunity in *H. pylori*-linked disease (Negrini et al., 1993; Appelmelk et al., 1996; Aspinall & Moran, 1997). Evidence for phase variation was recently reported for the expression of LPS in *H. pylori* (Appelmelk et al., 1998). In a type strain of *H. pylori* (NCTC 11637) expression of LPS was not stable. The LPS in this given strain, which expresses Le\(^x\), displayed a high frequency of phase variation, resulting in the presence of several LPS variants in one bacterial cell population. Variants that expressed a derivative of Le\(^x\) lacking fucose, monomeric Le\(^y\), and both Le\(^x\) and Le\(^y\) were detected. Part of the variation observed was ascribed to altered levels of expression of fucosyltransferases that are encoded by genes known to possess oligonucleotide repeats, thereby suggesting possible...
Table 1. Phase-variable candidate genes [adapted from Tomb et al. (1997) and Saunders et al. (1998)]

<table>
<thead>
<tr>
<th>Repeat</th>
<th>Gene similarities</th>
<th>TIGR number</th>
<th>Reference*</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPS biosynthesis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 [AG], 11 A</td>
<td>RfaJ/glycosyltransferase</td>
<td>0208</td>
<td>1, 2</td>
</tr>
<tr>
<td>13 C</td>
<td>(1-3)-α-Fucosyltransferase</td>
<td>0379</td>
<td>1, 2</td>
</tr>
<tr>
<td>13 C</td>
<td>Lex2B</td>
<td>0619</td>
<td>1, 2</td>
</tr>
<tr>
<td>13 C</td>
<td>(1-3)-α-Fucosyltransferase</td>
<td>0651</td>
<td>1, 2</td>
</tr>
<tr>
<td>14 C</td>
<td>(1-2)-α-Fucosyltransferase</td>
<td>0093 and 0094</td>
<td>2</td>
</tr>
<tr>
<td>Cell-surface associated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 [CT], 14 A</td>
<td>Outer-membrane protein</td>
<td>0009</td>
<td>1</td>
</tr>
<tr>
<td>15 T</td>
<td>Outer-membrane protein</td>
<td>0025</td>
<td>1</td>
</tr>
<tr>
<td>15 C</td>
<td>Hypothetical protein; streptococcal M protein</td>
<td>0058</td>
<td>1, 2</td>
</tr>
<tr>
<td>12 C</td>
<td>Methyl-accepting chemotaxis protein</td>
<td>0103</td>
<td>2</td>
</tr>
<tr>
<td>14 T</td>
<td>Outer-membrane protein</td>
<td>0227</td>
<td>1</td>
</tr>
<tr>
<td>9 T</td>
<td>Haem-binding lipoprotein/transport protein</td>
<td>0298</td>
<td>2</td>
</tr>
<tr>
<td>6 [CT]</td>
<td>Outer-membrane protein</td>
<td>0638</td>
<td>1, 2</td>
</tr>
<tr>
<td>9 C</td>
<td>FipP</td>
<td>0684 and 0685</td>
<td>2</td>
</tr>
<tr>
<td>8 [CT], 16 T</td>
<td>Outer-membrane protein/adhesin</td>
<td>0722</td>
<td>1, 2</td>
</tr>
<tr>
<td>6 [CT], 14 T</td>
<td>Outer-membrane protein/adhesin</td>
<td>0725</td>
<td>1, 2</td>
</tr>
<tr>
<td>5 [GA]</td>
<td>Alginate O-acetylation protein</td>
<td>0855</td>
<td>2</td>
</tr>
<tr>
<td>16 T</td>
<td>Iron-regulated outer-membrane protein (frpB)</td>
<td>0876</td>
<td>1</td>
</tr>
<tr>
<td>11 [CT], 14 A</td>
<td>Outer-membrane protein/adhesin</td>
<td>0896</td>
<td>1, 2</td>
</tr>
<tr>
<td>13 T</td>
<td>Outer-membrane protein</td>
<td>0912</td>
<td>1</td>
</tr>
<tr>
<td>10 A</td>
<td>ABC transporter/secretion protein</td>
<td>1206</td>
<td>2</td>
</tr>
<tr>
<td>14 A</td>
<td>Outer-membrane protein</td>
<td>1342</td>
<td>1</td>
</tr>
<tr>
<td>9 [GA]</td>
<td>Conserved hypothetical protein; outer-membrane adherence protein associated protein</td>
<td>1417</td>
<td>1, 2</td>
</tr>
<tr>
<td>DNA restriction/modification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 [CT]</td>
<td>Cytosine-specific methyltransferase</td>
<td>0051</td>
<td>2</td>
</tr>
<tr>
<td>15 G</td>
<td>Type I restriction enzyme R protein</td>
<td>0464</td>
<td>1, 2</td>
</tr>
<tr>
<td>12 C and 15 C</td>
<td>Adenine-specific methyltransferase</td>
<td>1353 and 1354†</td>
<td>1, 2</td>
</tr>
<tr>
<td>10 C</td>
<td>R/M methyltransferase</td>
<td>1369 and 1370</td>
<td>2</td>
</tr>
<tr>
<td>14 C</td>
<td>Restriction enzyme β-subunit</td>
<td>1471</td>
<td>1, 2</td>
</tr>
<tr>
<td>12 G</td>
<td>Adenine-specific methyltransferase</td>
<td>1522</td>
<td>1, 2</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 [AT]</td>
<td>Hypothetical</td>
<td>0211</td>
<td>2</td>
</tr>
<tr>
<td>12 G</td>
<td>Hypothetical</td>
<td>0217</td>
<td>1, 2</td>
</tr>
<tr>
<td>14 A</td>
<td>Inner-membrane protein</td>
<td>0228</td>
<td>1</td>
</tr>
<tr>
<td>9 G</td>
<td>Hypothetical</td>
<td>0335</td>
<td>2</td>
</tr>
<tr>
<td>15 T</td>
<td>pyrG</td>
<td>0349</td>
<td>1</td>
</tr>
<tr>
<td>15 A</td>
<td>Inner-membrane protein</td>
<td>0350</td>
<td>1</td>
</tr>
<tr>
<td>14 A</td>
<td>cagA</td>
<td>0547</td>
<td>1</td>
</tr>
<tr>
<td>8 T</td>
<td>Hypothetical</td>
<td>0586 and 0585</td>
<td>2</td>
</tr>
<tr>
<td>15 T</td>
<td>Hypothetical</td>
<td>0629</td>
<td>1</td>
</tr>
<tr>
<td>14 T</td>
<td>Hypothetical</td>
<td>0733</td>
<td>1</td>
</tr>
<tr>
<td>9 [AG]</td>
<td>Hypothetical</td>
<td>0744</td>
<td>1, 2</td>
</tr>
<tr>
<td>15 C</td>
<td>Hypothetical</td>
<td>1353</td>
<td>1</td>
</tr>
<tr>
<td>16 A</td>
<td>fecA</td>
<td>1400</td>
<td>1</td>
</tr>
</tbody>
</table>

* 1, Tomb et al. (1997); 2, Saunders et al. (1998).  
† Saunders et al. (1998) only.

Involvement of slipped strand mispairing (Table 1; Ge et al., 1997; Martin et al., 1997; Tomb et al., 1997; Appelmelk et al., 1998). However, it should be noted that the fucT gene sequenced by Martin et al. (1997) lacks the oligo-C repeat described by Ge et al. (1997).

**Generation of genomic diversity in H. pylori**

The generation of genetic diversity within a non-clonal species of bacteria is likely to result from either of two events. Diversity can result from exchange of DNA
Genetic diversity of Helicobacter pylori

within a particular bacterial cell or lineage (intra-strain rearrangements) or it may result from exchange of DNA between different strains or in some rare cases even different species (horizontal transfer).

**Intra-strain generation of genomic diversity**

A variety of intra-genomic rearrangements (larger rearrangements than frameshifts) may occur in bacteria, namely, deletions, tandem duplications, inversions and translocations within a genome. Many of these rearrangements may occur as a result of the presence of repetitive DNA sequences. The repeated DNA sequences may be indigenous to the species, or may be derived from DNA introduced through horizontal transfer (Bacheller et al., 1996; Lloyd & Low, 1996). The genome sequence of H. pylori (Tomb et al., 1997) reveals that H. pylori possesses a large number of repetitive DNA sequences. As well as several full-length and partial copies of the IS605 insertion sequence, there are four copies (two full-length and two partial) of a novel insertion element, IS606. Due to the number of copies and their presence throughout the chromosome, these elements may lead to recombinational events resulting in chromosomal rearrangements. The short tandem repeats identified in H. pylori may also be associated with genomic rearrangements. Such repeats, like simple repeat DNA of eukaryotes, may be hot spots for intra-genomic rearrangements, such as gene conversion and illegitimate recombination.

Covacci (1996) has proposed the existence of ‘hinge regions’ (perhaps involving repeat sequences) in the chromosome which facilitate recombination while leaving regulatory and coding sequences intact, thereby permitting the rapid development of the macrodiversity identified by discriminatory fingerprinting techniques. The bacterial nucleoid is compacted in vivo with RNA, RNA polymerase components and protein (Trun & Marko, 1998). The propensity of H. pylori DNA to rearrange perhaps suggests that the loop domain structure of its DNA may be more highly dynamic than that of other bacteria, thereby facilitating intra- and inter-molecular recombination.

**Horizontal gene transfer**

The phenomenon of horizontal transfer is likely to have played a significant role in the genetic make-up of H. pylori. Analysis of G + C content and sequence data obtained from H. pylori prior to 1997 suggested the presence of sequences related to eukaryotes, archaea and other prokaryotes. For example, analysis of the H. pylori major sigma factor (RpoD) revealed that the amino acid sequence of RpoD showed strongest similarity to those of Gram-positive bacteria, e.g. Staphylococcus aureus, Listeria monocytogenes and Enterococcus faecalis (Solnick et al., 1997). This has since been confirmed by the publishing of the entire genome sequence of H. pylori strain 26695 and strongly suggests that inter-species horizontal transfer of genes has occurred during the evolutionary development of H. pylori. Many of the proteins described in this Gram-negative bacterium show greatest amino acid sequence similarity to proteins from the different major taxonomic groups. Indeed, Goldman & Kranz (1998) published evidence for the transfer of an entire biosynthetic pathway. The biogenesis of c-type cytochromes can be divided into three system types. Type I is found in most Gram-negative bacteria, while type II is characteristic of Gram-positive bacteria, cyanobacteria and chloroplasts. It is the type II system that is found in H. pylori. These findings suggest that transfer of genetic information to H. pylori has occurred from different phylogenetic groups.

A total of five regions within the H. pylori genome have significantly different %G+C contents compared to the mean content of 39 mol% for the rest of the genome (Table 2). Two of these regions contain at least one copy of the IS605 insertion sequence discussed earlier, flanked by a 5S rDNA sequence and a 521 bp repeat. The other regions include the cag-PAI, a region associated with ß and ß' RNA polymerase and elongation factor EF-G, and a region associated with two restriction/modification systems (Tomb et al., 1997).

**Mechanisms of horizontal gene transfer**

In order for exchange of DNA to occur between strains, DNA must be transferred from one bacterium to another and be taken up by the host cell. There are three main mechanisms by which naturally occurring DNA transfer may occur in bacteria, namely, natural transformation, conjugation and transduction. Despite the electron microscopic evidence for the occurrence of phage in H. pylori (Heintschel von Heineg et al., 1993), transduction has not been demonstrated to occur. Recently Kuipers et al. (1998), in a study to define conditions under which H. pylori strains could exchange DNA, found evidence for a deoxyribonuclease-resistant conjugation-like mechanism involving cell-to-cell contact which contributed to DNA transfer between H. pylori cells. A similar process of deoxyribonuclease-resistant transformation not involving conjugal plasmids has been described for N. gonorrhoeae (Catlin, 1981). Natural transformation has also been extensively demonstrated for H. pylori and, accordingly, it would appear to be an important mechanism for horizontal gene transfer in H. pylori (Nedenskov-Sorensen et al., 1990; Wang et al., 1993; Hofreuter et al., 1998).

**Natural transformation and competence**

*In vitro* studies have revealed that most strains of H. pylori are naturally competent for DNA uptake (Nedenskov-Sorensen et al., 1990; Wang et al., 1993). It is currently unknown how competence is regulated in H. pylori, although a gene cluster involved in the regulation of competence has recently been described (Hofreuter et al., 1998). This cluster consists of four tandemly arranged genes with partially overlapping ORFs (orf2, comB1, comB2 and comB3) constituting a single trans-
In several bacterial species the genes involved in natural competence are coordinately expressed under the influence of particular environmental conditions, suggesting that cells may actively respond to conditions in the habitat by adjusting their level of gene acquisition through natural transformation rather than by being passively subjected to environmentally forced gene entry (Lorenz & Wackernagel, 1994). The release of DNA from individual *H. pylori* cells for uptake by other *H. pylori* cells may be facilitated by autolysis (Phadnis et al., 1996). In *Streptococcus pneumoniae* the regulation of competence for genetic transformation depends on a quorum-sensing system (Alloing et al., 1996; Pestova et al., 1996) where a concerted response is initiated by signalling within the bacterial population. Such a mechanism, although not identified by analysis of the genome sequence, may possibly be present in *H. pylori* and may also be autolysis-linked (Phadnis et al., 1996). Hофреutеr et al., 1998) postulated that the 37-amino-acid-long secretory peptide encoded by *orf2*, the first ORF of the *comB* operon, may have a function in the regulation of competence in *H. pylori*.

In *Haem. influenzae* and *Neisseria meningitidis* the uptake of DNA by transformation is facilitated by the presence of uptake-specific sequences, i.e. a DNA consensus sequence of 29 bp and an invariant 10 bp sequence, respectively (Dempsey et al., 1995; Smith et al., 1995). Numerous copies of these uptake-specific sequences are present in the genomes of these bacteria, e.g. in the *Haem. influenzae* genome 1465 uptake-specific sequences have been identified (Fleischmann et al., 1995; Smith et al., 1995). However, as yet no such uptake-specific sequences have been identified in *H. pylori*. Following uptake of foreign DNA, repeats or repetitive elements often provide the substrate DNA for homologous recombination to occur between DNA molecules (Bachellier et al., 1996).

It is difficult to assess the importance of natural transformation in gene transfer within and between natural populations due to the absence of adequate *in vitro* models. Heretofore, what data there are may be of questionable relevance to the *in vivo* situation. Analysis of the *H. pylori* genome sequence identified 11 restriction/modification systems which have been presumed to protect the bacterium from integration of foreign DNA. Berg et al., 1997a questioned whether or not these systems are fully functional and whether or not they were acquired before or after the uptake of foreign genetic elements. However, more detailed analysis performed by Saunders et al., 1998 has provided evidence that all of these systems are unlikely to be functional. Furthermore, the integration of transformed DNA into the genomes of other naturally competent species, namely, *Bacillus subtilis*, *S. pneumoniae* and *Haemophilus* spp., appears to be relatively unaffected by the restriction systems of the recipient bacteria (Lorenz & Wackernagel, 1994). Sequence divergence is reported to be the most likely barrier to successful transformation with foreign DNA (Matic et al., 1996).

**Model for microevolution during infection**

Initial colonization of the gastric mucosa by founder isolates of *H. pylori* leads to a persistent infection by the bacterium. These founder colonies may all undergo microevolution of their genome structure. The failure to identify greater than a single strain or subtype within individual biopsy specimens indicates that within discrete sites (foci) within the gastric mucosa, identical clones colonize. The bacteria colonizing different sites within the stomach probably come under slightly different selection pressures, e.g. the pH level within different regions of the stomach is known to vary considerably, and different rearrangements may possibly be selected for by the different environmental conditions encountered at the different sites. *H. pylori* possesses flagella which facilitate its movement in the viscous gastric mucus and may perhaps play a role in the spread of infection within the gastric mucosa. Perhaps biopsies which are taken from adjacent sites are sampling the same focus of infection, whereas biopsies from regions of the gastric mucosa further removed from each other sample different foci of infection and contain different subtypes.

---

**Table 2. Genomic features that predict diversity (adapted from Tomb et al., 1997)**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Associated genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distinct G + C regions</td>
<td></td>
</tr>
<tr>
<td>Region 1 (33 mol % G + C)</td>
<td>IS605, SS rDNA and 521 bp repeat; virB4</td>
</tr>
<tr>
<td>Region 2 (35 mol % G + C)</td>
<td>cag-PAI</td>
</tr>
<tr>
<td>Region 3 (33 mol % G + C)</td>
<td>IS605, SS rDNA and 521 bp repeat</td>
</tr>
<tr>
<td>Region 4 (43 mol % G + C)</td>
<td>β and β′ RNA polymerase, EF-G ( fusA)</td>
</tr>
<tr>
<td>Region 5 (33 mol % G + C)</td>
<td>Two restriction/modification systems</td>
</tr>
<tr>
<td>Insertion sequences</td>
<td></td>
</tr>
<tr>
<td>IS605</td>
<td>tnpA and tnpB</td>
</tr>
<tr>
<td>IS606</td>
<td>tnpA and tnpB</td>
</tr>
</tbody>
</table>
Gene diversity of *Helicobacter pylori*

H. pylori lives predominantly in the mucus layer and must cope with the continual washing generated by peristalsis. The infecting bacteria are likely to be under significant selective pressure. In a hostile environment, like the gastric mucosa with low pH and an active immune response, it may be an important adaptation for a bacterium to mutate or rearrange its DNA frequently to generate genetic diversity within the population (Taddei et al., 1997). By continually generating cells with varied phenotypes, mixed populations are likely to exist, some members of which are better suited for particular environmental conditions (Dybvig, 1993). Many questions regarding the modes of genetic variation involved remain to be answered. Why *H. pylori*, which has phenotypic adaptations apparently specific for this habitat, e.g. flagella and urease, needs to generate such a level of genomic diversity and how it tolerates such levels of mutations and rearrangements within its genome are intriguing questions.

Pathogenesis is a complex and multifactorial process involving the coordinated action of a large assembly of virulence-associated loci. Genes associated with short repetitive sequences which encode putative surface-expressed proteins in *H. pylori* may represent examples of contingency loci, where phase variation of these proteins may result in adaptive advantage (Moxon et al., 1994). Horizontal gene transfer of DNA to *H. pylori* is facilitated by the fact that it is naturally competent for DNA uptake (Nedenskov-Sørensen et al., 1990). Mosaic gene structures, such as found in the vacA alleles of *H. pylori*, are indicative of horizontal gene exchange and associated recombination (Maynard Smith et al., 1991). Only approximately 10% of individuals in the developed world appear to be colonized with more than one strain, so frequent exposure to transformation of DNA from other strains appears to be unlikely. Therefore, the possibility for intra-species gene exchange appears to be limited. However, in the developing world, where *H. pylori* infection is more common, multiple infections are more frequent (Berg et al., 1997b).

Autolysis has been shown to occur in *H. pylori* (Phadnis et al., 1996), although recent data appear to conflict with this (Vanet & Labigne, 1998). Autolysis is associated with the release of DNA from cells which is then available for transformation. In other competent bacterial species, the autolysis of the bacterial population is correlated with the induction of competence genes (Lorenz & Wackernagel, 1994). Perhaps autolysis is associated with competence in *H. pylori*. Such a system would conform to a modular lifestyle, whereby the success of the individual is not as important as that of the population (Andrews, 1995). In this model, bacteria undergoing autolysis could be conceived as DNA donors, maintaining the genetic diversity and high rates of genome rearrangements to help the fitness of the population as a whole. Further analysis of the transformation and conjugation-like mechanisms of *H. pylori* is required to elucidate their role in DNA transfer and any association with the autolysis phenomenon.

**Concluding remarks**

*H. pylori* is a truly fascinating organism with many unique characteristics. The key issue raised by this review is why is this pathogen so genotypically diverse. The level of genomic diversity among *H. pylori* isolates indicates that this organism is an excellent model for studying the mechanisms of natural gene transfer and recombination. As a chronic human pathogen, which most likely causes life-long infections and is associated with several diseases including gastric cancer, further analysis of the macrodiversity of *H. pylori* populations is required as well as comparative genome sequences to unravel the details of its unique chromosomal structure.

**Acknowledgements**

The work of the authors has been supported by the Forbairt Basic Research Programme (SMB), the Irish American Partnership–Forbairt Scholars Programme (WGD) and the Health Research Board of Ireland (DGM). The authors are grateful to Professor Colm A. O Morain, Department of Gastroenterology, the Meath and Adelaide Hospitals, Dublin, for his valued contribution, research enthusiasm, and financial support and to Jane Twogood for technical support.

**References**


D. G. MARSHALL and OTHERS


