Helicobacter pylori – molecular genetics and diagnostic typing

Zhongming Ge and Diane E Taylor

Department of Medical Microbiology and Immunology, University of Alberta, Edmonton, Alberta, Canada

The genome of H. pylori is 1.68–1.73 Mb in size and contains a relatively low GC content (an average of 32.5 mol%). Physical and genetic maps of five H. pylori strains (NCTC 11637, NCTC 11638, NCTC 11639, UA 802 and UA 861) have been constructed and the complete genome sequence of strain 26695 has been determined. At least 50 genes, some of which play important roles in the physiology and pathogenicity of the bacterium, have been cloned. Marked genomic sequence variability has evolved from strain to strain demonstrated by random arrangement of 17 known genes on the chromosome and frequent mutations within individual genes. Based on such variability, sensitive and efficient molecular typing techniques such as ribotyping, AR-PCR, PCR-RFLP, PCR-DNA sequencing and PFGE-RFLP have been developed and widely applied in both epidemiological and clinical studies of this pathogen. Subtypes of vacA (encoding a vacuolating cytotoxin) and the intermediate forms of a pathogenicity island (the cag region) have been identified in different H. pylori strains and these individual vacA subtypes are associated with specific clinical manifestations of H. pylori infection. Further studies on relationships between the genetic diversity and pathogenicity of H. pylori strains would lead to the development of novel and efficient therapeutic strategies for eradication of this microorganism.

After Helicobacter pylori was first described by Warren and Marshall in 1983, most studies in the early years focused on classification, in vitro culture and histopathology of this human pathogen. In the 1990s, with the aid of powerful molecular biological techniques including polymerase chain reaction, DNA cloning, gene expression, mutagenesis and sequencing in combination of pulsed-field gel electrophoresis (PFGE) and electroporation/natural transformation, our knowledge of the genetics of H. pylori has expanded exponentially. Physical and genetic maps of the H. pylori genome have been constructed, and physiologically and pathologically important genes have been cloned and characterized from H.
**Genes and genome of *H. pylori***

**Overview**

The size of the *H. pylori* chromosome is estimated to be from 1.68–1.73 Mb\(^3\), determined by PFGE-restriction fragment length analysis of 29 *H. pylori* isolates. The genome sequence contains a relatively low GC content\(^4\), an average of 32.5 mol\% G+C with a range of 34.1–37.5 mol\%. At least 50 genes have been cloned from *H. pylori*. The important house-keeping genes include 16S\(^5\), 23S and 5S ribosomal RNA (rRNA) genes, gyrA involved in DNA replication\(^6\), recA responsible for homologous DNA recombination\(^7\) and ftsH essential for cell viability\(^8\). The *H. pylori* genome contains two copies of each of the 5S, 23S and 16S rRNA genes\(^9\). The 5S and 23S genes are often linked together, whereas in some strains the 16S rRNA gene is separated from the 23S/5S rRNA gene cluster\(^9\).

Several genetic determinants have been implicated in virulence of *H. pylori*, including a vacuolating cytotoxin gene (*vacA*), a cytotoxin-associated gene (*cagA*), flagelin genes (*flaA, flaB*), a urease gene cluster (*ureA and ureB* encoding structural subunits, *ureC and ureD* of unknown functions, *ureE, ureF ureG, ureH and ureI* required for urease activity) and *picB* having a potential ability to induce the production of interleukin-8 (IL-8) by gastric epithelial cells. These studies were recently reviewed by Labigne and de Reuse\(^10\) and will be addressed in detail elsewhere in this issue. More recently, a *cag* region consisting of *cag I* and *cag II* has been identified on the *H. pylori* chromosome\(^11\). These authors proposed that this region represents a pathogenicity island (PAI) within *H. pylori* similar to those found in *Salmonella, Yersinia*, and *Escherichia coli*. The *cag* I region contains at least 12 genes which are structurally similar to those constituting type III secretion systems involved in virulence and bacterial-host cell interaction\(^11\).
Insertional mutations within the genes including cagM, cagL, cagI, cagH, cagG, and cagE (or picB) markedly reduce the ability of H. pylori to induce secretion of IL-8 from gastric epithelial cells. Since IL-8 is a strong mediator of the inflammatory response, these results indicate the cag region is involved in the pathogenicity of H. pylori.

An insertion sequence, designated IS605, has been identified in the H. pylori genome. This new H. pylori IS605 contains two genes tnpA and tnpB which likely encode transposases. The tnpA product exhibits significant sequence similarity to the transposase of IS200 from E. coli, whereas the deduced amino acid sequence of tnpB is similar to the transposase found in the thermophilic bacterium PS3. A short inverted repeat with a core sequence CTTTAG flanks IS605.

Approximately 50% of H. pylori strains carry plasmids whose sizes range from 1.5–40 kb and whose restriction patterns are highly variable. These plasmids are cryptic, i.e. not found to be associated with any phenotype. The complete nucleotide sequences of two H. pylori plasmids pHPK255 (1.5 kb) and pHPM180 (3.5 kb) have been determined. The deduced Rep protein of HPK255 shares sequence similarity with the Rep proteins from a group of small plasmids utilizing the ‘rolling-circle-replication’ mechanism found in Gram-positive bacteria, whereas the deduced Rep protein (ORF1) of pHPM180 has significant sequence similarity to the Rep proteins which replicate by the ‘theta-type’ mechanism, suggesting that pHPM180 could also employ such a mechanism for its replication. Only two homologous regions between the nucleotide sequences of pHPM180 and pHPK255 have been found, which are restricted to the second long direct repeat of pHPM180.

**Sequence diversity**

Genetic studies have revealed a unique feature of the H. pylori genome, high genomic variability. Upon NruI and NotI digestion followed by PFGE, all 29 different strains examined gave discrete restriction fragment profiles. Comparison of the genetic maps of five H. pylori strains demonstrated that there is no characteristic arrangement of 17 known genes on the chromosome conserved by these strains (macrodiversity). DNA sequence diversity within this gastric pathogen was also demonstrated by using arbitrary-primer PCR. Sixty independent clinical isolates could be distinguished from one other by a single ‘random amplified polymorphic DNA’ (RAPD) primer.

Distribution of the cag region among different H. pylori strains is highly variable. Forty strains could be divided into eight groups. Two prototypes, namely type I and II, have been identified. This designation
Helicobacter infection appears to be an oversimplification of the situation as so many intermediate types exist. Nevertheless, in type I strains the cag I and cag II regions are linked together and there is no IS605 sequence in the chromosome. In contrast, type II strains contain neither the cag region nor IS605. Ten and six strains belong to type I and type II, respectively. The remaining 23 strains can be classified within each of the other 6 intermediate groups in which the cagI and cagII regions are separated by one or more copies of IS605, or coexist with IS605 (one or more copies) or are partially or completely deleted.

Sequence diversity also occurs at the level of individual genes of H. pylori (microdiversity). Early studies indicated that mutations have arisen within ureA and ureB\text Superscript 15,16, ureC and ureD\text Superscript 16 as well as flagellin genes (flaA and flaB) detected by PCR-based restriction fragment length polymorphism (RFLP) patterns\text Superscript 18. Recently, nucleotide divergence has been seen in genes such as ureC and vacA by a PCR-DNA sequencing approach\text Superscript 19,20. In the case of the H. pylori vacA gene, all H. pylori strains carry the vacA gene but only approximately 45% produce active cytotoxin (referred to as an ability to induce vacuoles in epithelial cells \textit{in vitro}). Comparison of nucleotide and amino acid sequences of a 0.73 kb region of vacA revealed 88–97% sequence identity for both nucleotide and amino acid sequences between vacA\text Superscript + and vacA\text Superscript – strains, whereas the degree of sequence identity between vacA\text Superscript + and vacA\text Superscript – strains showed a significant decrease (~70% at the nucleotide level and ~60% at the amino acid level)\text Superscript 20. Further sequence analyses revealed the presence of five vacA subtypes within 59 strains, of which three signal sequence (namely sla, sib and s2) and two middle region motifs (ml and ml) were identified. Fifty-nine strains can be grouped by combination of signal sequence and middle region motifs into 5 types\text Superscript 20: sla/ml (8), sla/ml2 (12), slb/ml (14), slb/ml2 (6) and s2/ml2 (19).

Mechanisms causing genomic diversity are poorly understood at present. For macrodiversity, one hypothesis suggests that this diversity reflects a long evolutionary association with the human host\text Superscript 9. Different strains remain within each individual human and his/her offspring for many generations and may independently undergo evolution. Alternatively, the diversity could result from gene re-arrangement within the chromosome caused by transposon-mediated gene mobility and recombination between repeated sequences\text Superscript 9. On the other hand, microdiversity in H. pylori may result from two mechanisms: either by a progressive accumulation of point mutations in different strains over time or by horizontal gene transfer followed by allelic exchange between strains\text Superscript 9. However, there is no solid evidence so far to support any of the above hypotheses to explain either macrodiversity or microdiversity.
Molecular typing of *H. pylori*

Molecular diagnosis and typing of *H. pylori* can provide information valuable for epidemiological and clinical studies, to appreciate population genetic structure, and to understand evolution of the microorganism. Sensitive and efficient techniques are essential for such purposes. Based on the high degree of genomic variability within *H. pylori* strains, various molecular techniques have been developed to differentiate clinical isolates.

**Ribotyping**

The principle of this method is that chromosomal DNA from different *H. pylori* strains are completely digested by selected restriction endonucleases (e.g. HindIII, HaeIII and Dral for *H. pylori*) and subsequently such restriction digestion patterns of the ribosomal RNA (rRNA) genes are visualized by hybridization with either radioactively or nonradioactively labeled specific DNA or RNA probes. This technique has been successfully applied in typing clinical isolates of *H. pylori*.

**Restriction fragment length polymorphism (RFLP)**

Intact chromosomal DNA is digested by restriction endonucleases and separated by pulsed-field gel electrophoresis (PFGE). Genetic maps can be constructed by hybridization of the restriction fragments with DNA probes prepared from known genes.

**PCR-RFLP**

DNA fragments are first amplified using PCR with primers generated from known genes, followed by restriction digestion. *H. pylori* strains are differentiated by this technique in the urease genes including *ureA*, *ureB*, *ureC* and *ureD*.

**Arbitrarily primed PCR (AP-PCR) or random amplified polymorphic DNA (RAPD)**

This technique involves amplification of random DNA fragments with primers consisting of arbitrary nucleotide sequences. Profiles of randomly amplified DNA fragments can discriminate among strains.

**PCR-DNA sequencing typing**

This technique is based on the PCR amplification of a gene known to be polymorphic and common to all *H. pylori* strains. Subsequently, the
nucleotide sequences of the fragments amplified from different isolates are determined and compared by direct DNA sequencing\textsuperscript{19,20}.

**Direct PCR typing**

This method has recently been developed to detect so-called ulcerogenic strains based on the representative motifs in \textit{vacA} or \textit{cagA}\textsuperscript{20}. Type-specific primers can be generated from the individual motifs including s1a, s1b, s2, m1 and m2 and then used to diagnose types of clinical isolates by PCR. In addition, since approximately 45\% of \textit{H. pylori} strains contain the \textit{cagA} gene, which is considered to be a marker of ulcer-producing strains, primers derived from \textit{cagA} can be easily used to distinguish \textit{cagA}\textsuperscript{+} and \textit{cagA}\textsuperscript{-} strains.

The standard protocols for each of these methods are reproducible and available in the recent book edited by Clayton and Mobley\textsuperscript{22}. Readers should be aware that although emerging evidence suggests that the types of the \textit{vacA} motifs are associated with clinical outcomes caused by corresponding \textit{H. pylori} strains\textsuperscript{23}, we believe that these data are still preliminary and further work is required to confirm a definite link. As a result, at present caution is advised before the techniques based on the \textit{vacA} motif divergence are widely applied to clinical practice.

**Key points for clinical practice**

- In the past 14 years, genetic studies have considerably increased our understanding of this important human pathogen, including pathogenicity, virulence factors and pathogen-host interaction using animal models and \textit{in vitro} mutagenesis techniques.

- The discovery of high genomic variability in \textit{H. pylori} offers simple, accurate and efficient molecular typing techniques which are useful in both epidemiological and clinical studies.

- Importantly, some exciting progress has been made on genetic diversity–pathogenicity relationships among \textit{H. pylori} strains. For example, \textit{H. pylori} strains containing s1a type \textit{vacA} are more highly associated with severe gastric inflammation and duodenal ulceration (89\% of 18 patients with s1a strains) than \textit{vacA} s2 strains which are associated with less inflammatory and ulcerous presentations (20\% of 10 patients with s2 strains)\textsuperscript{23}. 
The types of the PAI present in *H. pylori* may also contribute to the pathogenicity, since some of the genes in this region are involved in induction of a pro-inflammatory cytokine IL-8 in gastric epithelial cells *in vitro*.[11]

Further studies on whether or not genetic diversity in *H. pylori* is directly related to pathogenicity would explain, at least in part, why only approximately 10% of people world-wide carrying this pathogen develop ulcers (80-90% of a total population infected in developing countries and about 50% of 50-year-old people infected in industrialized countries).

Eventually, molecular studies may answer important public health issues such as who should be treated, when intervention should begin and what are the most efficient therapies to eradicate this gastric pathogen.

**Acknowledgements**

Work on *H. pylori* was supported by Glaxo-Wellcome Canada, the Medical Research Council of Canada and the Canadian Bacterial Diseases Network to DET who is an Alberta Heritage Foundation for Medical Research Scientist.

**References**

23 Atherton JC, Peek RM, Tham KT, Cover TL, Blaser MJ. Clinical and pathological importance of heterogeneity in vacA, the vacuolating cytotoxin gene of Helicobacter pylori. Gastroenterology 1997; 112: 92–9