2008

Analysis of Sequence Variation at Two Helicobacter pylori Genetic Loci Potentially involved in Virulence

George Warren Liechti
College of William & Mary - Arts & Sciences

Analysis of sequence variation at two *Helicobacter pylori* genetic loci potentially involved in virulence.

George Warren Liechti

Springfield, Virginia

Bachelors of Science, College of William and Mary, 2003

A Thesis presented to the Graduate Faculty
of the College of William and Mary in Candidacy for the Degree of
Master of Science

Department of Biology

The College of William and Mary
May, 2008

APPROVAL PAGE

This Thesis is submitted in partial fulfillment of
the requirements for the degree of

Master of Science

George Warren Liechti

George Warren Liechti

Approved by the Committee, April, 2008

Committee Chair
Associate Professor Mark Forsyth, Biology, The College of William and Mary

Professor Margaret Saha, Biology, The College of William and Mary

Associate Professor George Gilchrist, Biology, The College of William and Mary

ABSTRACT PAGE

Helicobacter pylori colonizes the gastric mucosa of nearly half the world's population and is a well documented etiologic agent of peptic ulcer disease (PUD) and a significant risk factor for the development of gastric cancer. The majority of H. pylori isolates that have been linked to peptic ulcer disease share a common genetic element known as the cag Pathogenicity Island (cag PAI). Recent studies comparing the genomes of more virulent and less virulent strains have revealed the presence of numerous pathogen-specific genes that may have been introduced to the more virulent, cag PAI positive H. pylori strains by horizontal gene transfer. In addition to the acquisition of virulence associated genes, it has been shown that gene loss also plays an important role in the development of numerous pathogens, although to date this avenue of research has been relatively unexplored in H. pylori. In this study, we show that HP1079 and HP1078, two genes of unknown function and currently lacking any orthologs in other sequenced bacterial species, exist as a co-varying locus with the cag PAI. In addition, we show that HP1079 has the potential to undergo phase variation and that HP1078 exists in two distinct allelic forms that are present in relatively equal frequencies in our study set of clinical isolates of H. pylori from North America. While a proposed role for HP1079 and HP1078 in virulence must await animal model studies, we hypothesize that this locus is a potentially novel virulence associated locus based upon its association with cag PAI positive H. pylori strains, the fact that it is subject to phase variable expression, and that one gene in this locus exhibits allelic variation. We also characterize another genetic locus, currently annotated as HP0937 and HP0938, and show that this locus appears to be actively undergoing genetic decay. The patterns of pseudogenesis are so dissimilar between isolates that we hypothesize that the selective pressure on the locus occurred subsequent to the divergence of these strains from their last common ancestor.

Table of Contents

List of Figures and Tables

INTRODUCTION

Background

Helicobacter pylorus is a spiral shaped, gram-negative, microaerophilic bacterium that colonizes the gastric mucosa of nearly half the world's population. It is an etiologic agent of peptic ulcer disease (PUD) as well as a significant risk factor for the development of gastric cancer. Socio-economic status is an established indicator of *H. pylori* infection rates, and in developing countries, infection rates in adults can reach as high as 90% [1, 2]. Because of its prevalence in the developing world and the emergence of antibiotic resistant strains [3], research involving the classification of novel *H. pylori* vaccine targets and a deeper understanding of this pathogen's interaction with its human host is of paramount importance. In this brief overview, some basic background on the pathobiology associated with infection by this organism will be examined, including its capability of eliciting disease as well as some of the environmental stresses thought to play a role in its initial and continued colonization of the human gastric mucosa.

There is still considerable debate as to the mode of transmission of this bacterium. *H. pylori* has been cultured from stool samples of infected individuals and screening of water samples from local sources using polymerase chain reaction (PCR) methods have shown the presence of *H. pylori* DNA, suggesting at least the possibility of a fecal-oral route of infection [4-6]. Currently, only one group has reported success in culturing *H. pylori* from an environmental reservoir [7]. Other studies indicate a potential oral-oral route of infection, with gastric reflux moving the bacterium into the oral cavity where it

has the potential to be spread orally [8, 9] However, the exact mechanism of *H. pylori* transmission remains unknown.

Infection/Colonization

Despite the accumulation of a large body of research over the last twenty years, many of the details involving the pathogenesis of *Helicobacter pylori* associated disease remain unknown. In rhesus monkeys, the infectious dose of *H. pylori* required to achieve colonization of the stomach was determined to be 10^4 [10]. The mammalian stomach is generally thought of as one of the immune system's early lines of defense against pathogens and the low pH associated with chemical digestion of food also provides an effective barrier to most bacteria. In order to survive in this relatively harsh environment, *H. pylori* colonizes the mucus layer that covers stomach cells and protects them from the harsh, acidic environment.

Once established in the mucus layer, the bacterium engages the host's immune response, which results in the recruitment of neutrophils, macrophages, and lymphocytes [11-13]. The level of immune response is quite variable among infected persons and it is believed that numerous factors including host and pathogen genotypes determine the severity of this response [11, 14]. Despite this induced inflammatory response, the infection is rarely cleared. The recruitment of increasing numbers of immune and inflammatory cells to the gastric mucosa often results in severe inflammation. The production of auto-antibodies specific for the gastric epithelium has also been reported as a consequence of the *H. pylori* induced immune response, indicating a level of auto-immunity is present during prolonged infections [15].

Gastric epithelial cells, in response to this extended inflammatory response, begin to lose adherence to each other due to a disruption of their tight junctions, allowing interstitial fluid to leak out into the mucosa and providing *H. pylori* with essential nutrients. It is this long term chronic inflammation of the stomach lining that is believed to be the root cause of ulcer disease. Mucin, a lubricant that protects many of the body's surfaces, is found at significantly reduced levels in the stomach over time in patients colonized with *H. pylori*, resulting in the inability of the stomach to neutralize free radicals from accumulating neutrophils [17-19]. Infected individuals have an estimated lifetime risk of 10-20% for the development of peptic ulcer disease, approximately 3-4 fold higher than in non-infected subjects. *H. pylori* infection can be diagnosed in 90-100% of duodenal ulcer patients and in 60-100% of gastric ulcer patients [16]. Approximately 5% of infected patients have malignancies, with a clinical progression beginning with gastritis, proceeding through atrophic gastritis, and intestinal metaplasia, and finally malignancy [11].

Bacterial Genetics

Since its discovery by Barry Marshall and Robin Warren in 1983 [1], *H. pylori* has become a prominent model system for bacterial pathogenesis, especially for studying persistent infection. This is due in large part to it being the only bacterium currently classified as a class 1 carcinogen by the World Health Organization and the fact that the species possesses remarkable genetic diversity [20] with hundreds of unique, classified strains. It is widely held that nearly every clinical isolate is unique and distinct from all others, with the exception of those isolated from family members. It is also believed that

these different strains rarely compete with each other in host colonization, and that infecting strains are commonly passed from mother to child [21-23]. The substantial variability between any two given strains has proven to be quite useful as a tool for tracking human migrations as far back as 12,000 years ago by comparing divergent *H. pylori* strains [24-26]. High mutation rates can lead to significant divergence of subclones from a primary isolate, with numerous phenotypic and genotypic differences [27, 28].

Despite the significant genetic variability demonstrated in numerous studies, estimates of the rate of genetic change vary greatly [25, 29]. Long-term studies have shown that while clonal variants may exist within a bacterial population, all isolates are generally related and derived from a common ancestor [29]. Using whole genome *H. pylori* DNA microarrays, *Sala et al.* demonstrated that approximately 22% of *H. pylori* genes identified in two sequenced strains of *H. pylori* (26695 and J99) are missing in other isolates [30]. This indicates the tremendous genetic variability present within this single bacterial species and supports the theory that microbial populations exist in the framework of a larger 'metagenome' where genetic elements are constantly being exchanged and expressed by various subpopulations. In this context, genes essential for survival in all environments or 'core-genes' are found in all strains, while genes essential only in specific environmental conditions are found in given subgroups of strains (Fig 1). Bacteria are thought to undergo rapid change as they take up, incorporate, and/or remove DNA from their genomes and this genetic environmental background consisting of all genetic material within the environment defines the metagenome.

Virulence Determinants

In comparing the genetic profiles of various *H. pylori* isolates, is has been possible to identify certain genes that are found more frequently in isolates from patients who exhibit severe disease. These genes encode virulence determinants and their presence is thought to confer a selective benefit to pathogenic bacteria. Early studies on *H. pylori* were based on identifying virulence determinants in pathogenic strains [31, 32]. This continues to be an avenue worthy of pursuit [33-35] and recent studies comparing the genomes of pathogens and non-pathogens have revealed the presence of pathogen-specific genes introduced to virulent bacterial strains by horizontal gene transfer [30, 33, 36]. Horizontal gene transfer is one of the principal mechanisms for genomic variability in *H. pylori* [37] and gene acquisition is thought to be the major factor responsible for observed strain variation [38]. *H. pylori* is naturally competent, i.e. it can freely import and incorporate DNA from the external environment into its genome. It preferentially incorporates DNA with methylation patterns similar to its own, so DNA acquired from the environment is generally from its own genetic "heritage". Other bacterial species do not take up environmental genetic material to this degree and it has been hypothesized that environmental stresses may stimulate competence in these species, yet the signals that trigger this event *in vivo* are currently unknown [39].

The origins of these laterally transferred genes are often difficult to determine as any distinguishing characteristics found within the transferred genetic material tend to reflect a signature of a particular lifestyle or local effect acting on the sequence rather then a previous genomic context [40]. Simply put, it is almost impossible to determine exactly where a gene obtained via horizontal gene transfer originally came from because

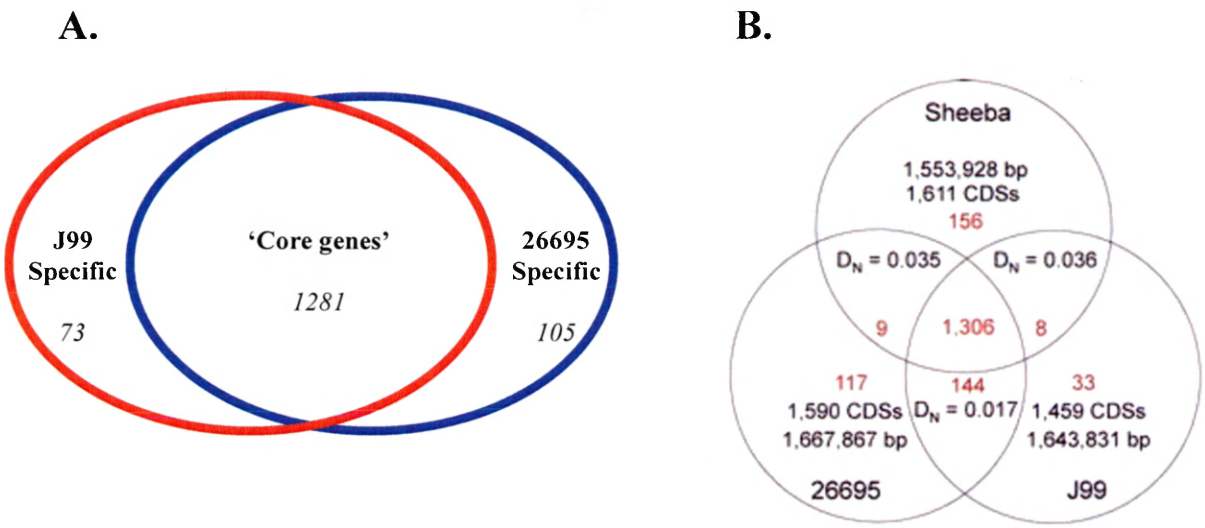


Figure 1. Identification of bacterial core genes through comparative genomics. A, The first whole genome microarray comparing the *Helicobacter pylori* strains 26695 and J99 found surprising variability between the two genomes [30]. Relatively large numbers of strain specific genes were identified, and the function of the majority of these genes remains unknown. **B,** Genetic comparisons between the two sequenced *H. pylori* strains and its close relative *Helicobacter acinonychis* strain Sheeba (isolated from the stomachs of large felines) found similar sets of strain specific and core genes indicating the potential for strain variation around a defined set of genes essential in most environments [98]. Identifying where overlap in gene content occurs among pathogenic bacterial

species allows attempts at correlating bacterial genetics with disease outcome and enables the identification of virulence determinants and potential targets for novel treatments. its genetic makeup was inevitably altered numerous times prior to being deposited into its new host genome. These alterations are thought to have been acquired largely by the imprecise insertion, replication, and removal of bacteriophages from bacterial genomes. Some postulate that these changes, whether the result of phage incorporation or not, directly affect the ability of a given gene to exist outside of a genomic context for extended periods of time. This has led researchers to suggest that selection may occur on genomic material existing free in the environment, such that only material of a specific composition can be taken up by a bacterium and that this composition may not reflect its ancestral, or even its most recent host genome [40].

In addition, some bacteria possess mechanisms that confer a level of control over gene expression of laterally transcribed genetic elements. The histone-like nucleoid structuring protein (H-NS) found in Salmonella has been shown to selectively silence sequences with GC content lower than the resident genome [41]. Regardless of origins, these genetic regions present in the environment represent an important resource for bacteria, enabling them to acquire functional gene products which would otherwise take long spans of evolutionary history to produce via mutational events alone.

One characteristic of pathogenic bacteria is the presence of numerous 'virulence genes' thought to have been acquired and incorporated via horizontal gene transfer and homologous recombination, respectively. Sometimes complex protein systems are encoded in large clusters of genes, and these gene clusters are acquired en masse by recipient bacteria. The majority of *H. pylori* isolates that have been linked to peptic ulcer

disease share a common genetic element known as a pathogenicity island (PAI) [31]. First coined by Hacker to describe unstable regions on the chromosome of uropathogenic *Escherichia coli* (UPEC) [42], PAI sequences generally contain numerous genes that bestow pathogenic capabilities on the previously benign or less virulent bacterium. *Yersiniae pestis*, for example, contains at least fifteen PAIs encoding everything from adhesin proteins and Type II/III secretion systems, to an antibiotic resistance mechanism [43]. These PAIs generally differ greatly from the rest of the bacterial genome in their G + C content, indicating they most likely are the result of horizontal gene transfer events. Due to the beneficial genes often carried on PAIs, the acquisition of genetic material in this manner has the potential to increase bacterial fitness, open up novel environmental niches for the bacterium, as well as increase pathogenicity [44-47].

Numerous genes are thought to aid *Helicobacter pylori* in its pathogenic lifestyle. Without these gene products, *H. pylori* is less capable of colonization and causing disease. The identification of these 'virulence determinants' is of enormous importance. Besides allowing us to more closely examine the exact interactions between the bacterium and their human hosts, these determinants allow us to design screens for the presence of these organisms in human patients, provide novel targets for potential treatments, as well as allow us to categorize the more dangerous bacterial isolates from the more commensal ones and treat patients accordingly. What follows is a list of some of the better characterized virulence factors of *Helicobacter pylori*. They are listed here to better familiarize the reader with the mechanisms of pathogenesis employed by *H. pylori* as well to demonstrate trends and common features that act as clues in the characterization of virulence factors in general.

Urease

One of the major obstacles to any bacterium attempting to colonize the GI tract is the strong acidic environment of the stomach. While other pathogens simply pass through the stomach on their way to the intestine, *H. pylori* proceeds to take up residence, making prolonged exposure to acid an environmental hurdle. It has been previously established that to limit the exposure to stomach acid, *H. pylori* actually colonizes the mucus layer of the stomach, essentially using the same protective coating the body uses to protect its own epithelial cells from the harsh pH of the stomach lumen. To further reduce the acidity of its environment, *H. pylori* uses the enzyme urease, which converts urea in the stomach into the basic compounds ammonia and bicarbonate, which act to buffer the pH of the periplasm and the area around the bacterium [48]. Urease makes up as much as 10% of the total protein content of a bacterium [49, 50], hinting at its relative importance. *H. pylori* strains containing mutated or deleted urease genes have proven incapable of gastric colonization in numerous model organisms [48, 51-53] and attempts to use urease as a vaccine candidate have given mixed results [50, 54, 55]. The ammonia produced by *H. pylori's* urease is also thought to have carcinogenic qualities [56, 57] and hypochloridia, a deficiency in the level of hydrochloric acid in the stomach, is a common symptom in *H. pylori* positive individuals [58-61].

Motility

It has been shown that *H. pylori* is incapable of colonization without functional flagella [62-64] indicating an absolute requirement for motility during infection. It is thought that despite its urease activity, *H. pylori* is only capable of surviving outside of

the mucus layer for a limited amount of time and that chemotaxis provides *H. pylori* a way to continually move away from higher acidity [65]. *H. pylori* is capable of discerning pH gradients in the stomach by using a chemotaxis receptor TlpB (encoded by the *H. pylori* gene HP0103) which enables it to continually move away from higher acidity, effectively keeping it closer to the stomach cells despite the high turnover rate of the gastric mucosa [64]. Thus motility and chemotaxis allow *H. pylori* to survive in the harsh and rapidly changing environment of the human stomach.

Because motility is required for the existence of *H. pylori* in its human host, this provides the host's immune system with an important target for mounting an immune response sufficient to potentially clear the infection. However, *H. pylori's* remarkable genetic diversity allows it to very rapidly change its outer surface markers that may come into contact with the host's immune system. These 'genetically plastic' regions allow for *H. pylori* to present a broad array of surface associated molecules to the immune system, potentially allowing a proportion of the bacterial population to evade detection [66, 67]. Flagellin, the major component of bacterial flagella, has such variable regions in its C-terminal D2-D3 domains that are exposed to the environment, while it's N-terminal D0-D1 domains that are buried within the flagella and essential for motility are highly conserved [68].

It is also noteworthy that flagella in *H. pylori* are completely sheathed in the bacterium's outer membrane [69, 70], effectively hiding the core subunits of the flagella from detection by the immune system through methods that will be discussed later. Flagellar sheaths are relatively rare phenomena, and have been characterized in the bacteria *Bacillus brevis*, *Bdellovibrio bacteriovorus*, *Pseudomonas stizolobii*, *Vibrio*

cholera and a few other *Vibrio* species [71-76]. It was previously reported that *Vibrio cholera*'s flagellar sheath was composed of a single protein and that lipopolysaccharide was almost entirely absent from the region [76]. Early studies of *H. pylori's* flagellar sheath concluded that it was comprised of mostly protein [77] and that a single protein, characterized as HpaA, appeared to be localized specifically to the sheath itself [78, 79]. However, upon further membrane characterization and electron microscopic imaging, it was accepted that the sheath is in fact comprised of a lipid bilayer, supporting a few earlier reports [80]. It was also noted that there may be distinct protein profiles between the sheath and the rest of the outer membrane [81, 82], though the protein HpaA does not appear to associate solely with the sheath region in all bacterial isolates [83]. Regardless of its composition, the sheath's predominant role is thought to be to protect the flagella from the acidic environment of the stomach. The existence of the flagellar sheath prevents interaction of the flagellin proteins with the host's defenses and only when damaged or when shed are flagella capable of interacting with the immune system, effectively announcing to the body the presence of a pathogen.

Despite being sheathed in the bacterial membrane and the large variation that exists in bacterial flagellin, the immune system still uses it as a target. The human innate immune system makes use of transmembrane proteins called Toll-Like Receptors (TLRs) to recognize pathogens and trigger an immune response [84-87]. Over time, these proteins have evolved to target regions that are essential to pathogens and thus are highly conserved. For example, TLR2 binds to peptidoglycan [88, 89], TLR4 binds to lipopolysaccharide [90], and TLR9 recognizes bacterial CpG DNA [91]. TLR5 recognizes bacterial flagellin [92], specifically the conserved N-terminal D0-D1 domain

that is essential for motility [68, 93]. Despite the highly conserved nature of this region, α and ε Proteobacteria appear to be different in at least three amino acid positions thought to be necessary for TLR5 binding [68]. In this way, these pathogens appear to have evolved to evade the host's innate immune response without loosing their ability to remain motile, while most other flagellated bacteria are recognized by and bind to TLR5. This illustrates a perfect example of the co-evolution of bacterium and host in the context of a single locus over evolutionary time.

Lipopolysaccharide (LPS)

Another major target for the immune system during a bacterial infection is lipopolysaccharide (LPS) [90, 94]. LPS is a major component of a gram-negative bacteria's outer membrane, contributing to the bacteria's structural integrity as well as providing a barrier against the immune system. A number of antimicrobial peptides recognize bacterial LPS and upon binding they induce an innate immune response [95, 96]. As mentioned above, the primary activator of the host response to LPS is TLR4 [90]. In many cases, the immune response resulting from the presence of LPS (often referred to as endotoxin) is so severe, that a dysregulated immune response to infection develops, known as sepsis, which can be life threatening. Studies are currently underway to develop molecular inhibitors of LPS, specifically targeting its ability to induce monocyte activation [97]. Currently, extensive filtration procedures are required to remove any contaminating LPS from reagents used during surgery or treatment of patients in order to remove the risk of accidentally inducing severe inflammation.

Because the host's immune response to LPS is so robust, many pathogenic bacteria modify their LPS in order to avoid this detection. *H. pylori* is a well studied example, as it can mimic the Lewis Blood Group (LBG) antigens presented on the O side chain of its LPS [98, 99]. Lewis Blood Group antigens are commonly associated with host monocytes, macrophages, granulocytes, and gastric epithelial cells. They are important for low-affinity adhesion to host gastric epithelial cells as well as in phagocytosis [100, 101]. It has been widely speculated that *H. pylori* uses this key feature of host cells to evade the immune system [98, 102]. The genes responsible for the creation of these bacterial LBG antigens are present throughout the genome, and many of them undergo phase variation [103]. The variation generated in the bacterial population, whether resulting from phase variation or other mechanisms [104] results in the production of a wide array of differing LPS profiles within a given population. This is believed to be a 'fine-tuning' of the bacterium's defenses, further enhancing the evasion of the immune response [103, 105-107].

In addition, *H. pylori* has modified the lipid A portion of its LPS compared to other bacteria. Lipid A is the 'hydrophobic anchor' of the LPS, and is the biological initiator of septic shock [108]. The modification of the lipid A portion of *H. pylori's* LPS results in reduced immunoreactivity by as much as 1000 times compared to the lipid A of other bacterial species [109-111].

cag Pathogenicity Island

In *H. pylori*, the major pathogenicity island (PAI) currently believed to be the best indicator of high virulence is known as the *cag* Pathogenicity Island [112, 113]. The 140

kDa product of the terminal gene in this PAI, CagA, is injected into host cells where it activates MAP kinase cascades that in turn lead to the activation of growth factors and alteration of epithelial gene expression, which resulting in cytoskeletal rearrangements [114]. Strains possessing the *cag* PAI are considered more virulent than isolates that lack the PAI. *cag* PAI-positive isolates elicit a much stronger mucosal inflammatory response [115, 116]. Other genes of the *cag* PAI encode structural proteins used in the translocation of CagA and possibly other proteins into the host's epithelial cells via a type IV secretion system. Numerous studies have identified loci apparently in linkage disequilibrium with *cagA*, potentially indicating that they may be important virulence determinants for *H. pylori* [33, 117-119]. It is worth noting that despite the *cag* Pathogenicity Island being generally accepted as a major virulence determinant of *H. pylori*, cases exist in which strains lacking this PAI have resulted in severe clinical outcomes [114].

Vacuolating Cytotoxin (VacA)

vacA, encoding a vacuolating cytotoxin (VacA) is present and secreted in all strains of *Helicobacter pylori* and encodes a single protein that is active in roughly 50% of identified strains [32]. It is secreted into the extracellular space by way of an autotransporter mechanism, during which the amino-terminal signal sequence and carboxy-terminal domain are cleaved from the 140-kDa protein [120]. The remaining mature 88-kDa protein, while not required for growth under laboratory conditions, acts as a cytotoxin. It's secretion results in the formation of vacuoles within the host cells *in vitro* and causes pores to form in the outer membrane of host cells *in vivo*, which may

subsequently release nutrients and cations from the gastric mucosa. Although present in all bacterial strains, *vacA* contains two variable domains that result in three distinct genotypes. The 5' terminal signal domain (S region) of the peptide encodes the secretion signal portion of the peptide and the mid-portion (M region) encodes the host cell-binding domain. Each can exist as two major subtypes, s1 or s2 and m1 or m2, respectively. This gives rise to three different allelic types, with the s1/m1 allele possessing the greatest vacuolating activity *in vitro*, the s1/m2 type having reduced activity, and the s2/m2 type having no activity [121]. In this manner, the s1/m1 allelic variant is associated with more severe disease in its human host [122].

babA

Helicobacter pylori outer membrane proteins (Hops) are thought to play critical roles in the bacterium's interaction with host cells. Among the most scrutinized of these is BabA (HopS). The BabA protein facilitates the binding of *H. pylori* to the Lewis B Blood Group antigen found on the glycoproteins of host gastric epithelial cells [123]. The *babA* gene exists as two distinct allelic types (*babA1* and *babA2*) with *babA2* being the only one believed capable of encoding a functional adhesion protein. Comparative studies have found high correlations between the presence of *cagA*, the *vacA* s1 allele, and the functional *babA* allele, making it a correlating virulence marker [124]. It has also been shown that sequence variation in *babA* exists between strains, potentially due to differing selective pressures for adhesion between hosts and even within a single host over a period of time [125]. These differences indicate the potential for a larger genetic family of BabA alleles whose altered makeup may be the result of close interaction with

varying host immune systems [126]. Surprisingly however, knockout studies have shown that the absence of *babA* does not prevent adhesion of *H. pylori* to gastric epithelial cells [127].

BabA has also been found to be metastable and is thought to respond to the degree of immune response by the host [128]. The genetic mechanism that allows for this degree of sensitivity is known as phase variation, and allows for multiple subpopulations of bacteria to exist simultaneously whose only differences are varying numbers of nucleotide repeats resulting in frame shifting in the open reading frames. This results in the phase variable gene being turned 'off' while in the rest of the population the gene is turned 'on'. In this manner, bacteria are capable of rapidly changing their phenotype within a population simply by adding or deleting a single base pair or dinucleotide within the coding sequences. These are often homopolymeric (eg. poly A) or dinucleotide repeats (eg. poly CT). Examples of such phase variation can found in numerous genes encoding adhesion proteins (*sabA, sabB, hopZ, oipA*).

The Outer Membrane Protein Adhesin, SabA

The gene *sabA* encodes a sialyl Lewis x (sLex)-binding adhesion, facilitating *H. pylori* binging to host membrane glycolipids [127]. It is hypothesized that this binding brings the bacterium into more intimate contact with the host cell then does BabA, however, the adherence is thought to be much less stable. This locus is known to undergo phase variation due to the presence of an ON/OFF frame-shift mutation, due to a homo-pyrimidine dinucleotide tract, giving rise to subpopulations. The local environment selects for the SabA expressing or non-SabA expressing population. Phase

variation of *sabA* may allow a varying degree of interaction between the host's immune response and the bacterium.

Additional studies looking more directly at the effects of the presence of SabA on the immune response have shown SabA to be an important virulence factor for the pathogenesis of *H. pylori* infection. SabA expression has been linked to the non-opsonic activation of human neutrophils in response to a bacterial infection resulting in the induction of oxidative metabolism and production of oxygen species that can damage the gastric epithelium [129]. SabA has been shown to be the key molecule in the activation of human neutrophils and has been shown to be required for phagocytic induction by neutrophils as well [130]. Taken together, these results linking SabA directly with immune system activation have led researchers to classify SabA as a major virulence determinant.

oipA

oipA encodes a pro-inflammatory, outer-membrane protein. The gene is found in all strains of *H. pylori*, however, its expression is regulated by phase variation, allowing bacterial populations consisting of bacteria that both express and fail to express this protein. Functional OipA appears to also be correlated with *cagA* status, however, little is known about its potential role in colonization and/or virulence other then that it elicits an inflammatory response by inducing IL-8 synthesis [131]. It has been shown that full activation of the IL-8 promoter is only achieved when both the *cag* PAI and OipA are present [131]. It has also been linked to the secretion of RANTES, a CC chemokine involved in the homing of classical lymphoid cells such as T cells and monocytes during

H. pylori infection [132]. It is thought to play a role in binding host cells, but no binding receptor has been established to date [133]. Studies have found that *OipA*-positive strains are found more often in patients with duodenal ulcers then in gastritis patients [134].

iceA

iceA is a novel gene found only in *H. pylori.* It is **i**nduced by **c**ontact with **e**pithelial cells and it exists as one of two distinct alleles, *iceA1* and *iceA2*. The presence of the *iceA1* allele is highly correlated with peptic ulceration as well as increased induction of IL-8 [135]. Both *iceA1* and *iceA2* were expressed *in vivo* by respective *H. pylori* strains in gastric biopsies. To date, little else is known about this gene or its product other then it appears to correlate with disease outcome in infected individuals.

dupA

dupA is a novel gene found to be associated with duodenal ulceration in patients. This gene is found in *H. pylori*'s "plasticity zone" (PZ), an area of the genome known for its variable gene content. The PZ accounts for roughly half of all of the strain-specific genes in any given isolate [136]. While very little is known about *dupA's* functional characteristics, it has been shown to correlate with a reduced risk for gastric atrophy and cancer, despite it also correlating with duodenal ulceration [137].

Gene Shuffling

Many known virulence determinants among pathogens facilitate host immune evasion. One method of immune evasion occurs at a genetic level and involves the

ability of a bacterium to rearrange its genome by means of DNA repeats and gene paralogs that can simultaneously alter the expression of disease-associated genes [39]. An example of this can be found in the bacteria *Neisseria gonorrhoeae* and *Neisseria meningitidis*. Repeat-mediated rearrangements cause genes encoding cell surface proteins to translocate to alternate sites on the chromosome. This positions 'silent genes' next to potential 'on switches' that causes their eventual transcriptional activation [39]. This gene shuffling has a direct effect on the proteins encoded, with an end result of each new bacterial generation presenting a new appearance to a host's immune system, making it harder for previous antibodies to identify the ever-changing bacterium [39].

Biofilms

A characteristic of many other persistent infections caused by other bacterial pathogens is the formation of an extracellular polysaccharide matrix, which acts as a semi-rigid structure for containing bacterial communities. Many bacterial species are capable of existing in two different phases, a planktonic phase in which a single bacterium exists autonomously or in a larger, communal construct made up of complex connections between numerous bacteria, potentially of numerous species. These three dimensional structures called biofilms are often found in the environment, usually involving dense communities of various bacterial species on water surfaces or at interfaces with rocks, glass, and other sediments. In the case of pathogenic bacteria, biofilms derived from monocultures are found to form *in vivo* in numerous bacterial species [138-141] and are generally associated with persistent bacterial infection. They have been shown to aid bacteria in both evading the immune system [142] and decreasing

bacterial susceptibility to antibiotics [138, 143] thereby allowing the genes encoding biofilm biogenesis to be labeled bacterial virulence determinants.

H. pylori has been shown to form biofilms *in vitro*, [144, 145] and studies have concluded that under low-shear stress conditions, *H. pylori* suspended in only water are capable of adherence to various materials at a wide range of temperatures while maintaining culturability for at least 6 hours [146] indicating a potential role for biofilm formation in *H. pylori*'s brief existence in the environment. The potential for formation of biofilms in the gastric mucosa has been cited [147], however, there remains little evidence to date that biofilms have a role during infection. Whether or not these biofilms definitively occur *in vivo* as well as whether they are involved in colonization, virulence, or immune evasion is currently unknown.

Each of the virulence factors described above may grant *H. pylori* a more efficient means of effectively colonizing the human stomach as well as causing significant disease. They bestow upon the bacterium a number of favorable capabilities, such as facilitating the successful evasion of the host immune system, allowing the bacterium to adhere to target epithelial cells, and eliciting inflammation sufficient to provide the bacterium with nutrients from host cells. Phase variation plays an important role in a number of the virulence determinants described, while allelic variation among bacterial strains provides an example of the co-evolution and highly dynamic interaction between bacterial pathogens and their hosts over evolutionary time.

Evolutionary Genomics

Horizontal gene transfer is hypothesized to play a major role in the acquisition of bacterial virulence determinants [148], however, DNA uptake is seldom beneficial to the

bacterium [149, 150]. Understanding how bacteria sort through an almost infinite amount of DNA in the environment is very important if we are to attempt to understand the root causes of bacterial pathogenesis. The environmental pressures involved in microbial evolution are important and warrant a thorough investigation.

Despite the fact that bacteria take up foreign DNA from the environment, bacterial genomes do not appear to grow larger over time [151] and isolates tend to remain relatively close in genome size [152, 153]. Additionally, unlike most eukaryotic organisms, prokaryote genomes generally consist largely of protein encoding or structural DNA [151]. It is currently believed that approximately 90% of a typical bacterial genome consists of protein-encoding sequence [151]. In order for bacteria to maintain relatively small, streamlined genomes consisting predominantly of protein-encoding DNA while simultaneously incorporating foreign DNA into their chromosomes, the rate of DNA influx must be balanced by the rate of DNA deletion [151].

Gene Removal : Black Holes

Despite the importance of horizontal gene transfer in *H. pylori*, other studies in other bacterial models have found it to be of limited importance in bacteria as a whole. These researchers cite gene loss and gene duplication as occurring much more frequently in bacterial genomes then horizontal gene transfer [154]. In bacteria that have acquired PAIs and subsequently express a more pathogenic phenotype, genes previously beneficial to a bacterium while existing in a benign association with its host, may no longer be under functional constraints or may possibly be inhibitory to the novel pathogenic lifestyle. The concept of selective loss of genes that inhibit pathogenesis has been shown

to occur in other bacterial models [155, 156]. A paradigm for this model is illustrated by *Shigella flexneri*, a virulent bacterium closely related to the commensal and pathogenic isolates of *Escherichia coli* [157, 158]. Studies have shown that lysine decarboxylase (LDC) activity is present in 90% of *E. coli* strains, but absent in all *Shigella* species [157]. When the gene for LDC (*cadA*) in *E. coli* was introduced into *Shigella flexneri*, virulence was attenuated [157]. The same study revealed a large deletion in the *cadA* region of *Shigella* species, revealing that bacteria evolve to become pathogens not only by acquiring virulence genes, but also by eliminating genes via deletions [157]. Similarly, a deletion in an arabinose assimilation operon has been demonstrated in *Burkholderia pseudomallei* to increase virulence [159]. It has recently been suggested that in some bacterial genera, gene loss rather then gene acquisition plays the dominant role in the evolution of pathogenesis [160].

Gene Removal : Bacteriophages

Bacteria have always existed in the presence of harmful genetic elements, and have evolved mechanisms by which they can defend themselves. Transposons and bacteriophages continue to be a constant danger to bacteria. Genetic parasites have the capacity to lysogenize the host bacterium's genome and induce their lytic cycles at a much later time. Until they enter their lytic phase, the phages multiply as a function of host bacterial replication. Upon activation of the phage's lytic cycle, the bacterium is killed. All descendants of the first lysogenized bacterium carry the risk of being killed by lytic phage. As a result of genetic parasitism, in bacteria, natural selection appears to favor high deletion rates in non-functional, foreign DNA [151]. This is countered,

however, by the substantial diversity seen in studied environmental isolates indicating a complex process of gene turnover and genomic uptake by bacterial hosts [161]. The presence of bacteriophage in the environment potentially increases bacterial diversity, essentially checking competitive dominants by 'killing the winner' and introducing a selection for acquiring phage-defense mechanisms [162]. Those bacterial species that would normally dominate under non-phage conditions or that are briefly dominant due to a beneficial gene encoded by a phage, are often at a selective, long-term disadvantage, and it is the bacteria that can effectively defend against phage that survive over evolutionary time.

Despite the potential advantages that might be gained from acquiring genetic material from bacteriophages, incorporated foreign DNA principally presents a danger to a bacterium unless the viral mechanism for propagation is removed. Examples of viral inactivation exist [36] but it is probable that gene loss via deletion in bacteria is predominantly driven by the presence of such genetic parasites. In the past, many believed simply that the accumulation of point mutations over time was enough to deal with dangerous parasitic DNA. *Nostoc punctiforme*, for example, has one of the largest microbial genomes identified and sequencing analysis indicates the presence of close to 150 Open Reading Frames (ORFs) that closely resemble tranposases in various stages of genetic decay [163]. Point mutations were thought to have eventually caused phages or transposable elements to 'get stuck' in the bacterial genome by preventing them from excising from the chromosome and in the case of phages, completing their lytic cycle. However, current studies in other bacterial models have concluded that the accumulation

of point mutations is an ineffective method of eliminating genetic parasites before they have the opportunity to kill their hosts.

It has been noted that few point mutations, in a single step, could eliminate all the mechanisms by which a prophage can kill a bacterium [151]. Pseudogenes in numerous bacterial genomes were recently compared, demonstrating that each genome contains a unique set of pseudogenes, indicating that pseudogene formation and subsequent deletion are relatively rapid processes [164]. Studies using comparative genomic approaches (microarrays) have found variability in genomic content in various bacteria to be largely due to prophages [165]. This is in stark contrast to the bacterium *Helicobacter pylori*, where tremendous genetic diversity exists despite the apparent absence of bacteriophage [164]. Only one *H. pylori* bacteriophage has been characterized (HP1), and the induction of a purely lytic propagation cycle required the supernate of another (supposed) lysogenic *H. pylori* strain [166]. Aside from this single report, the bacterium *H pylori* appears to lack any phage particles or even remnants of bacteriophages.

Numerous phages, upon entering their lytic cycle, excise from a genome bringing with them a portion of bacterial DNA. If this DNA, now part of the viral genome, contains a gene that may be of benefit to subsequent host bacteria, a level of selection may exist in which the bacteria infected with the phage may have an advantage over other bacteria in a given environmental context. Even though many phages have been shown to carry PAIs beneficial to bacteria, these benefits are short-lived if the prophage ultimately destroys the bacterium [151]. Non-functional genes are therefore believed to be removed quickly from a bacterial genome and many large non-coding sequences may represent DNA in the final stages of this degenerative process [167]. The rate at which

psuedogenes are removed from a bacterial genome differs greatly among microbes, however, it is currently believed that the presence of pseudogenes in a bacterial genome may indicate a relatively recent evolutionary event in the history of the bacterium [151]. In terms of genetic variability, this constant removal of foreign material from the genome results in traumatic change occurring continuously in bacterial genomes.

It is now widely believed that chromosomal deletions are the major bacterial mechanism in for eliminating dangerous, non-functional, or 'superfluous' DNA. In bacteria, essential genes have been shown to be more evolutionarily conserved than nonessential genes over both microevolutionary and macroevolutionary time scales [168]. Originally, deletions were thought to only prevent bacterial chromosomes from getting too large and cumbersome. It was believed that superfluous 'junk' DNA would place a burden on bacteria, slowing down their replication time while providing no benefit to the microbes at all. Since that time, however, evidence has accumulated countering this hypothesis. Genomic sizes among bacteria vary greatly, and this variation does not necessarily correlate with replication time. Some rapidly growing bacteria such as *Escherichia coli* have large chromosomes and many slow growing bacteria such as *Borrelia burgdorferi* have small chromosomes [151]. The argument that slower replication of large chromosomes is driving evolution of gene loss is now being carefully questioned. It is believed by some researchers that the cost of replicating even 'significant' amounts of extra DNA is paltry considering the huge 'energy budget' of the cell. The current belief among many microbiologists is that "high deletion rates are maintained not because the compact chromosomes that result are beneficial, but because the DNA that is removed is potentially detrimental" [151].

As mentioned previously, deletion events are believed to be the major mechanism for eliminating dangerous/non-functional DNA in bacteria. Whole-genome microarray studies of such pathogens as *Mycobatcerium tuberculosis* have concluded that single-nucleotide polymorphisms are very rare and that deletional events are important sources of genomic variation, more so even then horizontal gene transfer [20]. Significant studies in regard to pseudogene decay have been conducted in *Rickettsia prowazekii* and seem to contradict these previous findings. *R. prowazekii*a is a unique bacterium in that approximately 24% of its genome consists of non-coding DNA: a relatively high percentage for bacteria [167]. This apparent contradiction is thought to be a result of the microbe's environment. Evolutionary selection for deletional events is believed to vary given the specific selective pressures acting on the microbe. It has been largely accepted that inactivated genetic material within most bacterial genomes deteriorates spontaneously due to a mutation bias for deletion events. It is postulated that genetic elements provide such a threat that bacteria maintain high genomic deletion rates despite the increased risk of deleting a gene essential to their survival.

Environmental conditions also play a role in the evolution of a bacterial species. It is well established that environmental conditions have the potential to affect the bacterium's 'encounter-rate' with transposable elements. While phages themselves are found almost universally in the environment [169, 170], evidence exists that certain microbes may be less influenced by genetic parasites upon adoption of a pathogenic lifestyle. It has been reported that upon acquisition of a new host, an emerging bacterial pathogen undergoes large changes in its genome [171]. These changes may often include the accumulation of large numbers of mobile genetic elements, recombination with these

elements, pseudogene formation and subsequent deletions in the genome [171-173]. It is hypothesized that bacteria require many more genes to survive in the environment then they do within a host organism [174]. This is due to the presence of a larger variety of threats from numerous sources in the environment as compared with those presented by a single host [174], but also due to ephemeral nutrition sources in addition to more variable conditions then a bacterium would be exposed to within a host. A trend is becoming apparent among pathogens that they lose large amounts of DNA during the initiation of stable host colonization. Over longer periods of time the presence of extraneous genetic material as well as genetic parasites obtained from the external environment dissipates. It was long held that bacteria in general do not harbor many pseudogenes, though more recent studies of bacterial genomes have determined that perhaps many more pseudogenes exist in bacterial genomes then was once previously believed [171, 173]. It is thought that pseudogenes are removed by the same mechanisms that target bacteriophages, and that both genetic elements undergo similar selection [172].

Intracellular microbes may live a rather sheltered lifestyle in an environment relatively free of bacteriophages compared with other environments. Without the threat of these genetic parasites, it has been postulated that lower genome deletion rates may be selected for and that more 'junk' DNA will be allowed to accumulate [151, 167]. In comparison, environmental bacteria exist in a perpetual sea of genetic parasites, such as insertion elements, transposons, and bacteriophage [175], and there is likely strong selection for removal of foreign DNA under these conditions [176]. *H. pylori*, while not an obligate intracellular pathogen, does live in a unique environment, arguably free of genomic parasites as illustrated from the simple fact that its genome is completely free of

phage encoding DNA. Because the threat of genomic parasitism is reduced, *H. pylori* may tolerate foreign DNA to a greater degree then other free-living bacteria. Whether this relatively sterile environment plays a role in the establishment of the large degree of variability seen between any two given isolates of *H. pylori* remains to be determined.

While genomic uptake of DNA may play a major role in acquisition of pathogenic characteristics in *H. pylori*, DNA deletion is likely to remove dangerous phage elements as well as superfluous DNA that either is detrimental to a more pathogenic lifestyle or simply no longer subject to positive selection because it no longer encodes a functional product. In either case, the removal of this potentially dangerous DNA is thought to be quite rapid and the presence of non-functional genes or remnant viral DNA is thought to be evidence of a relatively recent evolutionary event.

Project Introduction

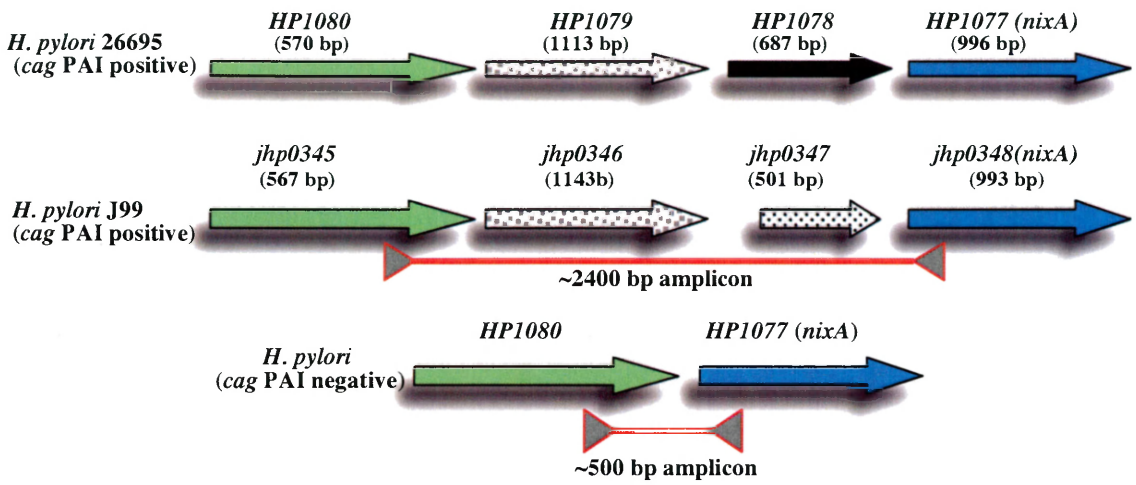
Novel Virulence Determinants

Previous work in our lab involved genomic comparisons of *H. pylori* strains containing the *cag* Pathogenicity Island (*cag* PAI) with those that lacked the *cag* PAI. We utilized DNA macroarrays containing 1,681 known *H. pylori* ORFs present in the genomes of the two *cag* PAI-positive sequenced strains of *H. pylori* (26695 and J99) [33]. In the five *cag* PAI-negative strains used in this study, DNA failed to hybridize with 27 genes in addition to genes that comprise the *cag* PAI. Two of the genes that did not hybridize are HP1079 (jhp0348) and HP1078 (jhp0347) (as annotated in the sequenced *H. pylori* strains 26695 and J99). Both of these genes are currently classified

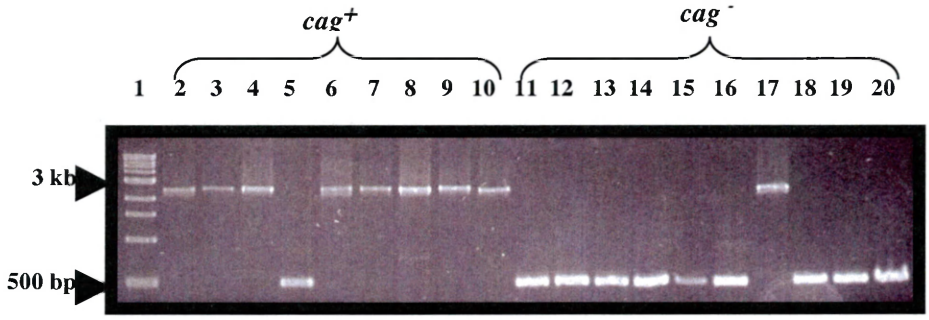
as genes of unknown function, and both are thought to express hypothetical proteins. Neither of these genes have an ortholog in any other bacterial species to date, though HP1079 does contain a highly conserved motif seen in numerous bacterial ATP hydrolyzing enzymes. The presence of both HP1079 and HP1078 is strongly associated with the presence of the *cag* PAI, suggesting a potential role in virulence (Fig. 2b). Gene specific screening for the presence of HP1078 has proven difficult using PCR techniques due to the significant variability present throughout the gene between bacterial orthologs in the sequenced strains of *H. pylori* [33]. However, 'empty site' PCR analysis of this region using primers to amplify sequences intervening HP1080 and HP1077 (*nixA*) yielded amplicons of sufficient size to contain both HP1079 and HP1078 from most *cag* PAI-positive isolates while much smaller amplicons resulted from most isolates lacking the *cag* PAI [33]. It was also noted that only large (~2kb) or small (~500bp) amplicons resulted, indicating that both genes are seemingly coupled, spatially in the genome and possibly in function as well (Fig. 2a). The presence of both genes might indicate a single insertion/deletion event in the evolutionary history of the bacterium. While the sequenced orthologs of HP1079 appear quite similar, HP1078 appears much more variable when comparing strains 26695 and J99 suggesting either a lack of functional constraints on the gene product or that multiple allelic forms exist as an adaptation to different host environments.

Figure 2. Analysis of the *HP1078-HP1079* locus in *H. pylori* clinical isolates. A, Arrangement of the genes in the variable locus containing *HP1079* and *HP1078* are shown for two sequenced strains of *H. pylori*, 26695 and J99 (both *cag* PAI-positive). Black triangles represent locations of empty-site primers (5' GCGGCGATTATGGAGCGGATTTG3' and 5' CGCATGCTTTGCCCCTAGCATG 3') designed to amplify sequences found between the highly conserved genes *HP1077* (*nixA*) and *HP1080*. Orthologous genes are represented by similar fill patterns and designated by the annotations of either Tomb *et al.* [136] for strain 26695 or Alm *et al.*[178] for strain J99. The annotated lengths of each open reading frame are shown in parentheses below the gene designation. Although *HP1078* of strain 26695 and *jhp0347* of strain J99 are located in orthologous positions, the deduced amino acid sequences are only 41% identical and 47% similar (data not shown). **B,** Inward facing primers designed to anneal within the highly conserved genes *HP1077* and *HP1080* (see Fig. 2 A) were used to amplify intervening sequences from 9 representative *H. pylori cag* PAI-positive isolates (**lanes 2-10**) and 10 representative *cag* PAI-negative isolates (**lanes 11-20**). **1.** 1 kb ladder, **2.** 26695, **3.** J99, **4.** 92-25, **5.** B125a, **6.** B185 **7.** B186, **8.** 87-33, **9.** 84-183, **10.** B105, **11.** J190, **12.** B211, **13.** Tx30a, **14**. 86-313, **15.** B80, **16.** 92-28, **17.** B166a, **18.** 92-24, **19.** J262, **20.** B141.

A.


H. pylori 26695
(cag PAI positive)
HP1080
(570 bp)
HP1079
(1113 bp)
HP1078
(687 bp)
HP1077 (nixA)
(996 bp)
H. pylori J99
(cag PAI positive)
jhp0345
(567 bp)
jhp0346
(1143b)
jhp0347
(501 bp)
jhp0348(nixA)
(993 bp)
~2400 bp amplicon
H. pylori
(cag PAI negative)
HP1080
HP1077 (nixA)
~500 bp amplicon

B.


cag+
cag -
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
3 kb
500 bp

In the current study, we sought to determine the degree of variation at the HP1079/HP1078 locus between *cag* PAI-positive and *cag* PAI-negative *H. pylori* strains. While our previous study had demonstrated HP1079 is highly associated with the presence of the *cag* PAI, we now present evidence that this gene is subject to phase variation based upon a poly adenosine tract. This gene was not previously recognized as a contingency gene under the control of phase variation. Additionally, the high level of sequence divergence between the sequenced orthologs of HP1078, including the absence of a consensus ribosome binding sequence of the J99 ortholog, led us to speculate a process of gene decay was occurring at this locus. However, when we examined the locus for potential genetic decay by sequencing the region using bacterial genomic DNA obtained from isolates collected six years after the initial isolates were collected (strain J99), we observed almost no genetic variability between isolates. In addition, our results show that HP1078 exists in two distinct allelic forms that are present in relatively equal frequency in our study sample of 73 North American clinical isolates of *H. pylori*. While a proposed role for HP1079/HP1078 in virulence awaits animal model studies, this locus is a potentially novel virulence associated locus based upon its association with *cag* PAI-positive *H. pylori* strains as well as the fact that it is subject to phase variable expression and the presence of allelic variants.

Bacterial Pseudogene Analysis

HP0938 (jhp0873) and HP0937 (jhp0872) are also classified as genes of unknown function, and have since been re-classified as pseudogenes [177]. They exist in nearly all *H. pylori* strains [30] regardless of *cag* PAI status. Using BLAST searches we found a single open reading frame (ORF) in several other microbes that shared homology with both of these genes, indicating that the full-length gene product may no longer exist in *H. pylori*. The function of the protein may potentially remain if the two pseudogenes work together as a dimer, however, this seems unlikely given the significant differences between the *H. pylori* strains examined, 26695 and J99. The HP0938/0937 locus in *H. pylori* strain 26695 strain appears to have a 322 bp deletion relative to the orthologous J99 locus. And the gene pair in strain J99 has a premature stop codon, resulting in the truncation of the ancestral gene. Given these observations, we hypothesize that two independent inactivation events have occurred within this conserved hypothetical protein-encoding gene. Potentially, this could be the result of strong evolutionary selection against this gene product in *H. pylori*. It could be argued that this gene product may have been detrimental to a pathogenic lifestyle and therefore was selected against in all *H. pylori* isolates observed. Alternatively, if the gene product was neutral, mutations in the gene would be entirely neutral and not reflect positive selection to eliminate protein function.

In this study, we attempted indirectly to ascertain whether this proposed full-length ancestral gene product was detrimental to a pathogenic lifestyle in *H. pylori*. The use of *H. pylori* as a model organism for this study is convenient given that, to date, no phage remnants and very few pseudogenes have been characterized in *H. pylori*, making

it a novel model system for examining bacterial pseudogenesis. In comparing *cag* PAI-positive and *cag* PAI-negative isolates, the genomic data always demonstrated the presence of more then one protein-encoding gene in the region between the two highly conserved adjacent genes, *yckJ* (HP0939) and *proP* (HP0936). This supports the hypothesis that the presence of multiple pseudogenes at this locus was not driven by selection for virulence in *H. pylori*. It is worth noting, however, that in each of the subsequent strains examined, the ancestral gene appears to be not only inactivated, but also inactivated in novel and unrelated ways. It is tempting to speculate that some change in the environment in which *H. pylori* lives has resulted in selection against a full-length gene product, and this selection has been applied universally in unrelated strains. This led us to speculate that rather then pathogenesis, this ancestral gene product may hinder initial colonization or survival in the gastric environment. We attempted to re-introduce a full-length "functional" ancestral gene into *H. pylori* to determine if the presence of the product significantly affected colonization/survivability. Unfortunately, we have been unsuccessful in this approach to date as the methodology for this task has proven somewhat unreliable. However, we did further analyze the region using a collection of clinical isolates of *H. pylori* to determine how widespread the existence of this ancestral gene's genetic decay was and whether any isolates shared an evolutionary history for this region.

METHODOLOGY

Bacterial Culture and Strains Used

Helicobacter pylori strains were cultivated on Trypicase Soy Agar II plates with 5% sheep blood (BBL) at 37°C and 5% CO_2. Genomic DNA was isolated by harvesting 24 hour blood agar plate cultures of *Helicobacter pylori* in 0.9% NaCl. Cells were then lysed using sodium dodecyl sulfate (SDS) at a final concentration of 0.5% and the genomic DNA was extracted with cetyltrimethylammonium bromide/sodium chloride (CTAB/NaCl) [178]. *Escherichia coli were grown* in Luria-Bertani (LB) broth or agar with appropriate antibiotic (100 μg/mL ampicillin). 74 North American isolates of *H. pylori* were used in this study (41 *cag* PAI positive and 33 *cag* PAI negative), of which 18 were from African-American patients. In addition, 10 African isolates (8 *cag* PAI positive and 2 *cag* PAI negative) were also used.

Molecular Techniques

Polymerase chain reactions were carried out at a total volume of 50 ul, containing 1X Mg^{2+}-Free Buffer (Promega), 2 mM $MgCl_2$ (Promega), 0.8 mM dNTP mix (0.2 mM of each nucleotide), 400 ng of each oligonucleotide, 100 ng of template DNA, and 2 units of Taq DNA polymerase (Promega). Thermal cycling (Perkin Elmer GeneAmp 2400) conditions included a hot-start of 5 minutes at 94°C followed by 30 amplification cycles and a final extension of 7 minutes at 72°C. Amplification cycles consisted of denaturation for 30 seconds at 94°C, annealing for 30 seconds at 3°C below the lowest oligonucleotide melting temperature, and extension at 72°C for one minute per kilobase

of expected amplicon. All oligonucleotides used in this study can be found in Table 1. Oligonucleotide primers used to amplify HP1080-HP1077 were HP1080ES, and HP1077ES. Oligonucleotide primers used to amplify the two proposed allelic types of HP1078 were HP1078 A-Fwd, HP1078 A-Rev., HP1078 B-Fwd, and HP1078 B-Rev. PCR amplicons were cloned using the pGEM®-T Easy vector system and T4 DNA ligase (Promega). Plasmid DNA was purified using Qiagen Miniprep spin columns (Qiagen). Restriction digests were preformed via standard protocols [179].

Cloning

All plasmid constructs were created using the pGEM-T Easy vector system (Promega). Restriction enzymes were used to remove ~300bp of sequence from HP1079 and insert a Chloramphenicol Resistance Cassette (Cat). Oligonucleotide primers used to verify the mutant plasmid were HP1079F(mut), and HP1079R(mut).

Sequencing.

DNA sequencing reactions were performed using the Big Dye V3.1 system (Applied Biosystems). The reactions consisted of 2 µl of Big Dye Terminator Mix, 1X Big Dye Sequencing Buffer, 100 ng DNA, and 200-400 ng oligonucleotide in a total volume of 20 µl. Inserts in pGEM®-T Easy-based plasmids were sequenced with both SP6 and T7 vector primers (Promega), or with sequencing primers specific for the amplified region. The conditions for sequencing cycles consisted of a 5 minute hot start at 94°C followed by 30 cycles of 45 seconds at 94°C, 30 seconds at 50°C, and one minute at 60°C. Dye Terminator Removal spin columns (Edge Biosystems) were used to purify sequencing

reactions, which were then dried via Speed Vac for between 30 and 45 minutes. Sequencing reactions were resolved using an ABI 3100 Avant Genetic Analyzer sequencer (Applied Biosystems). Analyses of sequencing results were analyzed with Sequencing Analysis 5.1.1 (Applied Biosystems) and MacVector 7.0 (Oxford Molecular) software.

Region Amplified	Primer Designation	Primer sequence
1079-1078 region	HP1080ES	5'-CGCCCAAAGCCATTCAAGAGGTGG-3'
	HP1077ES	5'-CCATAGGCGTTTTTGCCTTGTTGGGTGAG-3'
1078 - Allelle A	1078 A-Fwd	5'-GCTTGAAATTGAAAAATACGATAAAACAC-3'
	1078 A-Rev	5'-MTTTGTRTCCCAYKTGTATTGGAAA-3'
1078 - Allele B	1078 B-Fwd	5'-ACCAAGAAAGCGATGCAGG-3'
	1078 B-Rev	5'-GCCCTAGAAAATTTTTAAGGGG-3'
1079 Cat insertion site	HP1079F(mut)	5' -CACCAACAGCTATTTCACGC-3'
	HP1079R(mut)	5' -CTCATAGTCTTTATCCGC-3'
0938-0937 region	HP0939ES	5' -AAAAGCGAACGCCAGCTATG-3'
	HP0936ES	5' -GGGCTAAGCGTGTTTGGGAA-3'
Pseudogene repair	HP0938R	5' -CACAGCGACGCCTTGACAAATCC-3'
	HP0938R Mut	5' -TGGTCTTGCCCCATGTTTGAAAAGCC-3'
	HP0938F Mut	5' -ACATGGGGCAAGACCAGGTGCTAAAAAC-3'
	HP0938F	5' -CCTAACGCTTATGAAGAGAAGCG-3'
	intrgnc j99	5'-pGTTTAATGATAAAAAAGATCTAATAACCCACTTTTT-3'
	stpcodnfix	5' - pCTTTTCAAACATGGGGCAAGACCAGGTGCTAA-3'
	sma1mut j99	5' - pAAAAAGATAGAATAACCCGGGTTTTGAACGCTAA-3'
Knockin repair	jhp874F	5' -TCGCACAGCAAAAAGCGAACGCC-3'
	jhp874R	5' -TAAGCCCCATGCTATCCACGCC-3'

Table 1. Oligonucleotide primers used in the genetic analysis and knockout/knockin studies.

Allelic and Phase Variation in a *cag* Pathogenicity Island (PAI) Covarying Locus in *Helicobacter pylori*

RESULTS

HP1080-1077(*nixA*) Empty Site Analysis. PCR analysis was performed using 37 *H. pylori* isolates (18 *cag* PAI-positive and 19 *cag* PAI-negative) using primers based upon open reading frames flanking the HP1079/HP1078 locus. These genes (HP1080 and *nixA*) were previously demonstrated to be present in the vast majority of *H. pylori* isolates, regardless of *cag* PAI status [33]. This procedure distinguished isolates possessing an 'empty-site' (a 500 bp amplicon indicating the absence of HP1079/1078 between the HP1080 and *nixA* genes) from isolates possessing the HP1079/1078 pair (~ 2,400 bp amplicon). Large amplicons (~2.4kbp) were generated from 14 of 18 (77%) *cag* PAI-positive strains but only 2 of 19 (11%) *cag* PAI-negative strains. The smaller, empty-site sequences (~500bp) were amplified from only 1 of 18 (6%) *cag* PAI-positive strains but from 16 of 19 (84%) *cag* PAI-negative strains. A sampling of these results are presented in Figure 2b.

The region intervening HP1080 and *nixA* was sequenced from two *cag* PAI-positive strains that yielded large amplicons (isolates B105 and J166), one *cag* PAI-positive strain that yielded a small empty-site sequence (isolate B125), one *cag* PAI-negative strain that yielded a large amplicon (isolate B177), and one *cag* PAI-negative strain that yielded a small empty-site sequence (isolate 92-28). Sequencing was carried out using the cloned PCR products as templates and both strands were sequenced. The

large empty site sequences from the *cag* PAI-positive isolates contained two complete open reading frames. The amplicon generated from the *cag* PAI-negative isolate possessed HP1079 as a mutant pseudogene. BLAST results indicated strong homology between each of these ORFs and their corresponding genes (HP1079 and HP1078) found in both sequenced *H. pylori* strains (26695, J99). The small empty site sequences were found to contain no protein encoding sequences in the short intervening region separating the genes HP1080 and *nixA*.

We compared sequenced regions intervening HP1080 and *nixA* in 26695, J99, B105, and J166 at the nucleotide and predicted amino acid level. While HP1079 is quite well-conserved among these four *cag* PAI-positive strains, HP1078 demonstrated significant diversity between strains. An order of relationship was determined in which HP1078 alleles in strains 26695 and B105 appear highly similar at the predicted aa level and the HP1078 orthologs in J99 and J166 appear to be quite similar. This suggested the possibility of at least two distinct allelic variants of this *cag* co-varying locus. Taken together with our empty-site PCR data, these findings indicates that large empty site sequences, found predominantly in *cag* PAI-positive strains, contain both the HP1079 and HP1078 loci, and that significant sequence variation exists among HP1078 alleles.

HP1079 and HP1078 are stable over six years within a single host. Because of the significant alteration in nucleotide sequence around the 5' end of the HP1078 ortholog in strain J99 (annotated as jhp0347 in the J99 database [180]) including the apparent absence of a ribosome binding site, we entertained the possibility that this was a non-

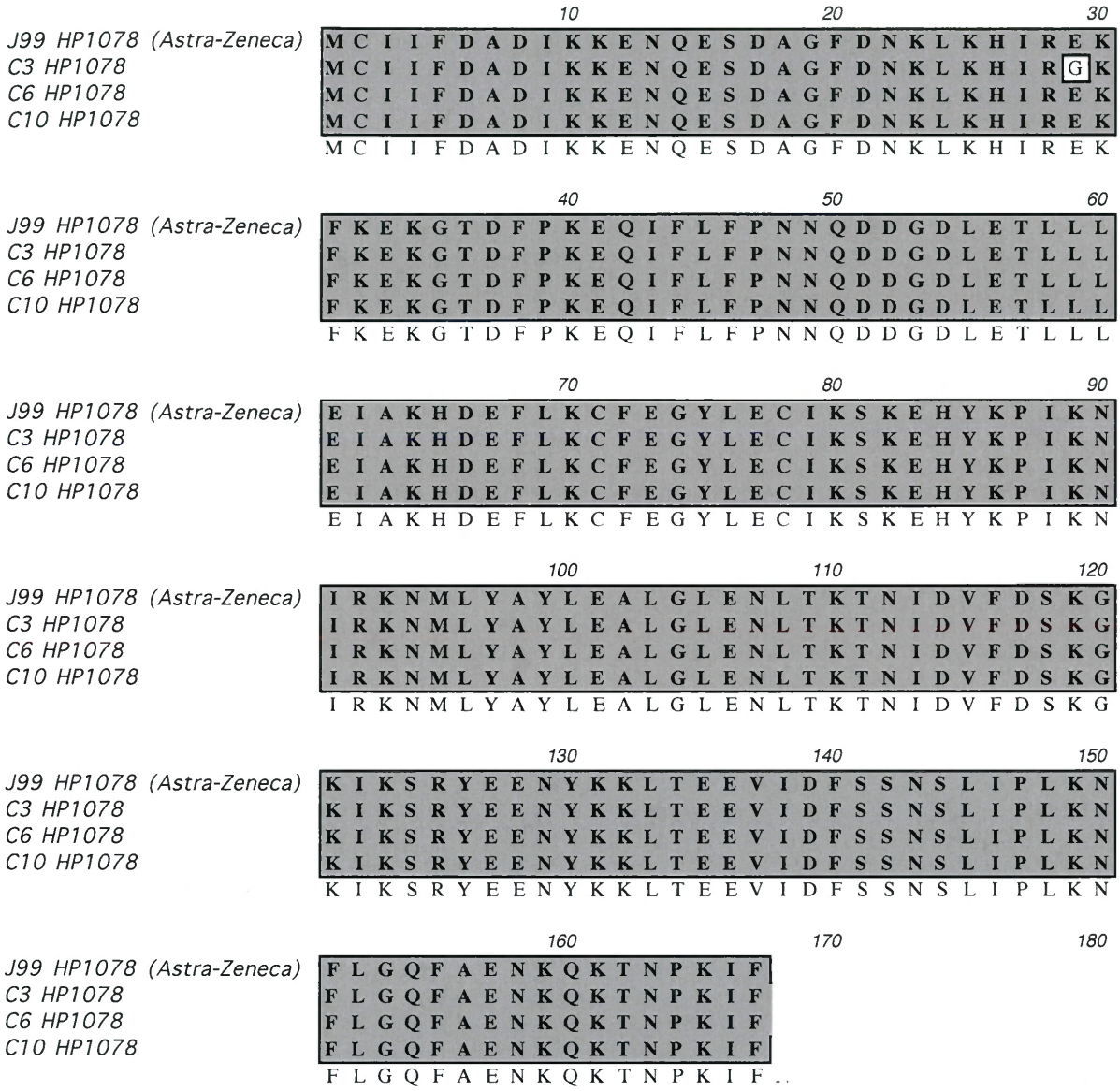
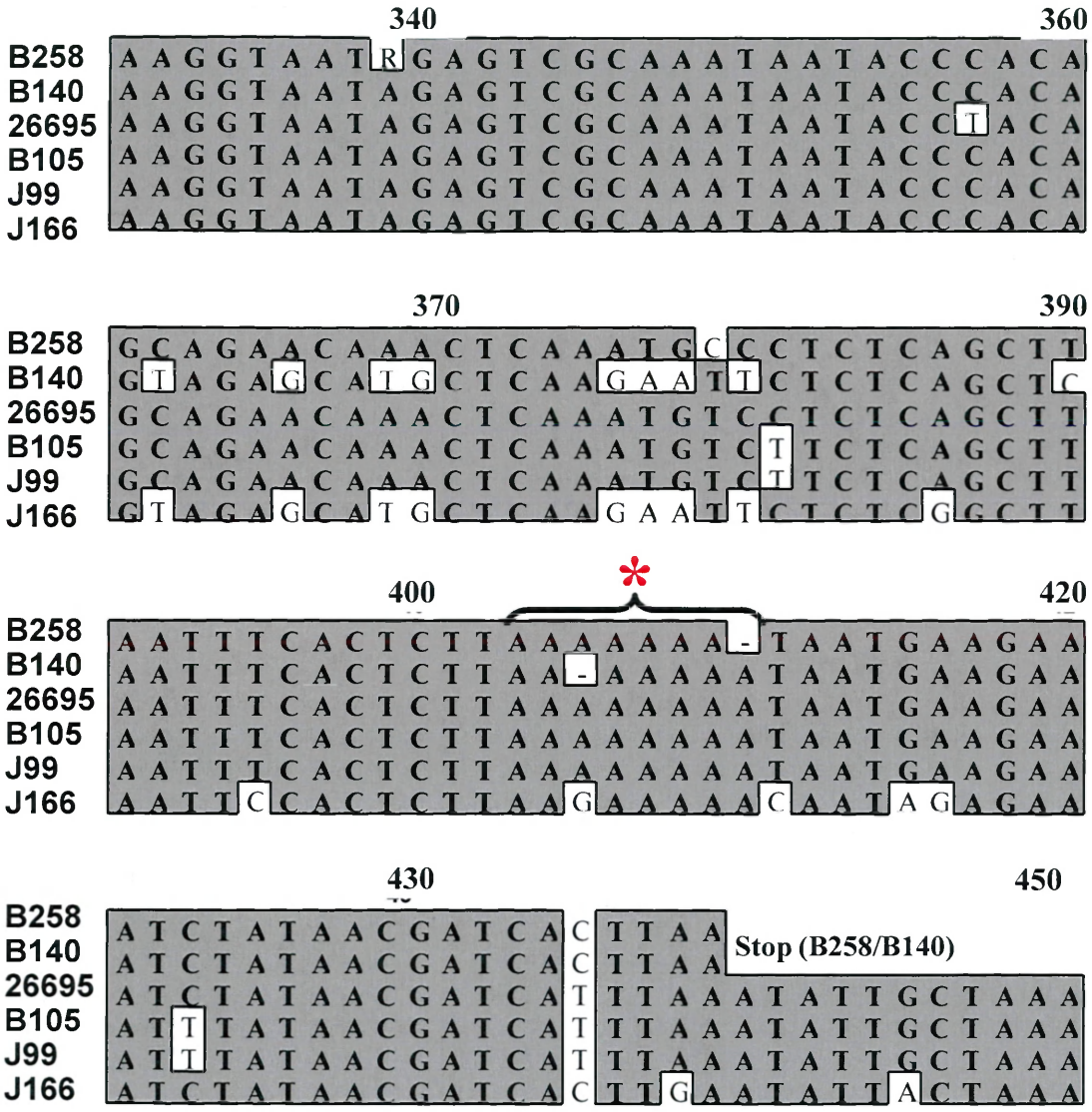


Figure 3. HP1078 is genetically stable over six years within a single host. The predicted amino acid sequences encoded by HP1078 from the initially characterized clinical isolate J99 and three re-isolates (C3, C6, and C10) recovered from the corpus biopsy of the original patient six years later. Only a single, non-synonymous mutation was observed in one of the re-isolates.

functional ortholog in strain J99. To begin to address this possibility, we performed PCR and sequencing analysis on the HP1080-*nixA* intervening sequence from three re-isolates of J99 (c3, c6, c10), recovered from corpus biopsy material from the original host approximately six years after the original isolation of strain J99 [28]. Of the three re-isolates, only a single non-synonymous mutation resulting in a predicted amino acid change occurred in HP1078 in one of the three re-isolates compared to the same ORF in the original J99 sequence (Fig. 3). Examination of the sequence of HP1079 revealed few mutations, all of which were synonymous mutations thus having no affect on the resulting predicted amino acid sequence. Although only six years after isolation of the original *H. pylori* strain J99, these data may suggest that both genes at this locus may remain under selective pressure due to constraints upon the gene products.

HP1079 is a phase variable gene. Comparative analysis was conducted for both the HP1079 and HP1078 orthologs from the *cag* PAI-positive isolates B105, B140, B258, and J166, which were determined during the current study, as well as the orthologs from 26695 and J99. In addition, the unusual strain, B177, which is one of the few *cag* PAI-negative isolates in our study to possess the HP1079/1078 orthologs was examined. All sequenced orthologs of HP1079 appear to share the same start codon and all are highly homologous with each other at the nucleotide level. Overall, HP1079 was found to lack much variation between strains with two notable exceptions. The first significant genetic variation is found approximately 403bp downstream of the start codon of HP1079 (Fig. 4). A poly-A track exists here and in four of the strains sequenced, the length of this

Figure 4. The *cagA* covarying gene HP1079 is phase variable. Nucleotide sequences of an internal portion of HP1079 orthologs of six different *cag* PAI-positive *H. pylori* isolates are shown (B258, B140, 26695, B105, J99 and J166). This region encodes from the highly conserved $lysine_{111}$ to $lysine_{150}$. The full-length HP1079 alleles in strains 26695 and J99 encode gene products of 370 and 381 amino acids, respectively. The poly adenosine region showing phase variation is indicated with an asterisk and the deletion of a single nucleotide within this region in strains B258 and B140 are indicated by dashes.

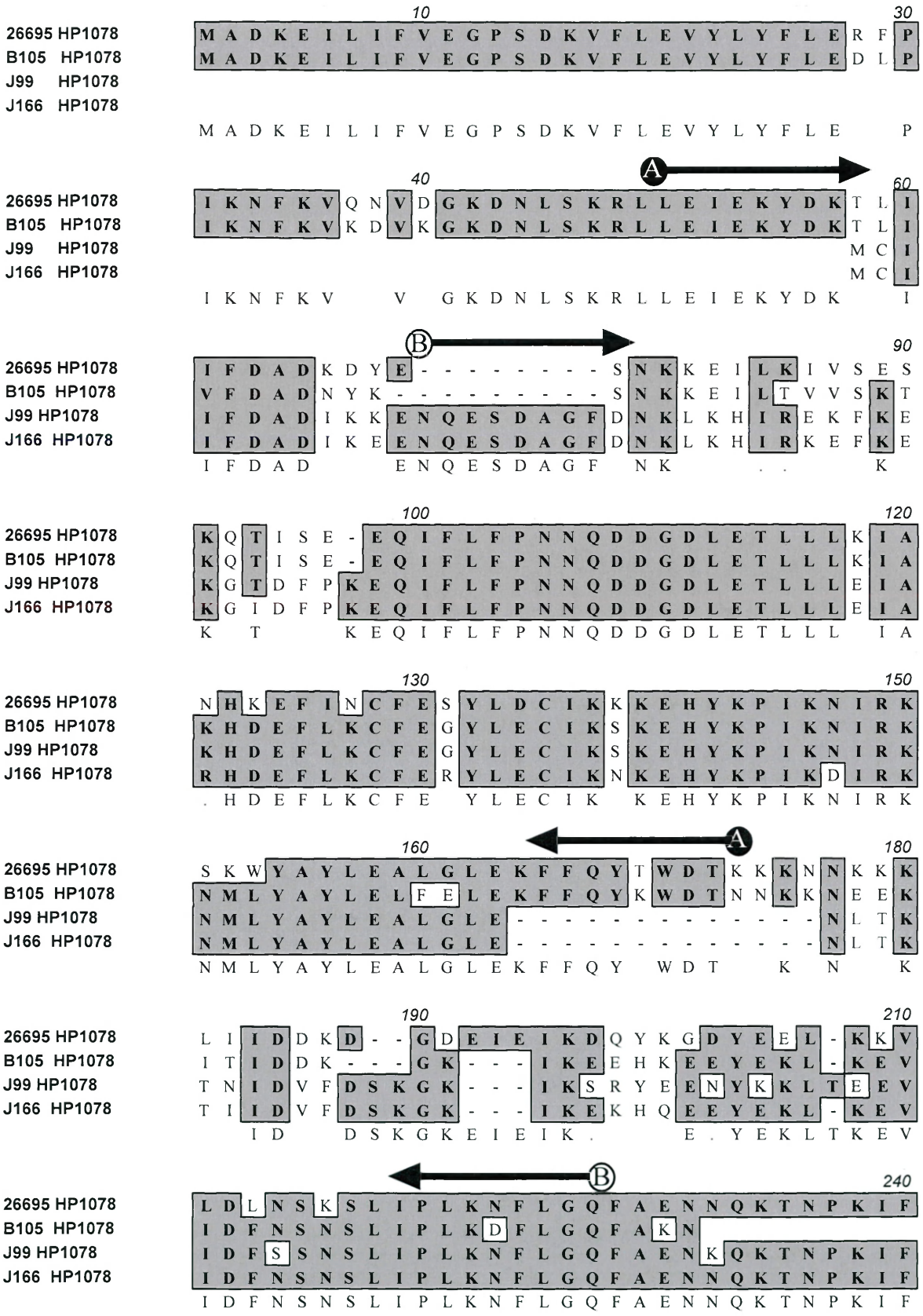

340
360
B258
B140
26695
B105
J99
J166
370
390
400
420
430
450
Stop (B258/B140)

homo-polymeric repeat numbers eight while in strains B140 and B258, this repetitive tract exists as seven adenine nucleotides (Fig. 4). The resulting frame shift in strains B258 and B140 causes the formation of an in-frame stop codon resulting in a predicted truncated gene product for HP1079.

Similarly, another, smaller, poly-A tract exists approximately 855bp downstream of the HP1079 start codon. In this instance, the poly-A region is five bases long in four isolates while in the fifth isolate (isolate B177) six adenines are present (data not shown). An in-frame stop codon exists 15 base pairs downstream of this sixth adenine in B177 while the remaining strains continue from 270bp to 300bp downstream before encountering a stop codon. The presence of this apparent slip-strand mispairing event leading to a frame-shift mutation in two *cag* PAI-positive isolates of the seven strains for which we have sequence data at this locus indicates that this particular locus is phase variable. The apparent divergence seen in the poly-A track in isolate B177 may simply indicate that the gene is no longer functionally relevant in a *cag* PAI-negative background, thereby allowing mutation without the presence of negative selection.

HP1078 allelic variation can be divided into two types. A collection of 36 North American *H. pylori* clinical isolates, with records containing patient backgrounds and *cag* PAI status [181] were used in our blind study. Based upon our demonstrated sequence diversity in HP1078 among isolates sequenced in the earlier part of this study, we performed PCR analysis on these 36 *H. pylori* isolates (23 *cag* PAI-positive and 13 *cag* PAI-negative) using primers whose design was based upon consensus HP1078 sequences found in strains 26695 and B105 (designated allele A) and in strains J99 and J166

Figure 5. The *cag* PAI covarying gene HP1078 exists as two major allelic variants. The predicted amino acid sequences encoded by HP1078 from four *cag* PAI-positive *H. pylori* strains, 26695, B105, J99, and J166 are shown. *Helicobacter* pylori strains 26695 and B105 possess the allelic variant designated type A while strains J99 and J166 possess type B alleles. Allele specific oligonucleotides (Table 1) were designed to differentiate isolates containing either of these alleles. Arrows ending with solid circles and the letter A indicate the approximate portion of the predicted gene product encoded by the binding oligonucleotide sequence for HP1078 allele A and likewise arrows ending with open circles and the letter B indicate the approximate binding for HP1078 allele B specific oligonucleotides. Note that the allele B "specific" reverse oligonucleotide is, in fact, not specific for allele B. The specificity of this amplicon in typing PCR relies upon the allelic specificity of the HP1078 allele B forward primer.

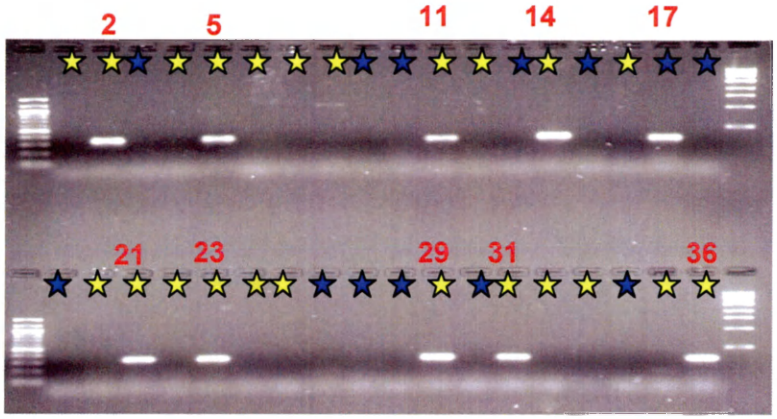

26695 HP1078
B105 HP1078
J99 HP1078
J166 HP1078
10
30
40
60
90
100
120
130
150
160
180
190
210
240
A
B

(designated allele B) (Fig. 5). We hypothesized that our previous inability to consistently amplify HP1078 sequences from a variety of *H. pylori* isolates [33] was due to allelic variation at this locus. In the current study, HP1078 sequences (either allele A or B) were amplified from 22 of 36 (61%) of these North American isolates (Fig. 6). Empty-site PCR was conducted on these same 36 strains for the HP1080-*nixA* region as described previously. Large empty site sequences were found in 25 of 36 (69%) isolates and the presence of an allele of HP1078 largely corresponded with the presence of the *cag* PAI (Fig. 2b).

In our HP1078 PCR analysis, we expanded our data set to include an additional strain collection used previously in the lab in characterizing the presence or absence of HP1079 and HP1078 by comparing the resulting sizes of the HP1080-*nixA* amplicon, as discussed previously. When combined with the data from our 'blind strain' collection, we found that only 3 of 51 (6%) *cag* PAI-positive isolates that contained the ~2.4kb HP1080-*nixA* PCR amplicon failed to yield an HP1078 specific amplicon with either set of allele specific oligonucleotides. These were isolates B140, B256, and B329. This locus from isolates B140 and B256 was sequenced and found to possess alleles of HP1078, however, minor sequence variations could be found to explain the failure of the polymerase chain reaction in these cases. These alleles are minor variants of HP1078 allele type A and do not rise to a level representing a third HP1078 allelic type. Our set of allele specific HP1078 oligonucleotides was successful at amplifying HP1078 sequences from 48 of 51 (94%) isolates yielding a large HP1080-*nixA* empty site amplicon (data not shown). We found no isolate from which we could generate a HP1078 amplicon with both sets of allele specific oligonucleotides (Fig. 6). We

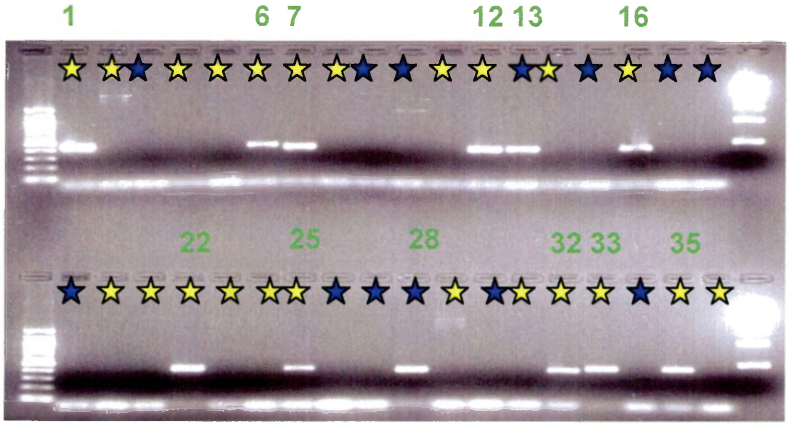
Figure 6. Allelic variation in *HP1078* among clinical isolates. HP1078 allelic typing PCR was performed on a collection of 36 *H. pylori* isolates [181]. 19 of the 23 *cag* PAI-positive isolates (83%) possessed one of the two possible alleles of HP1078, but only three of 13 *cag* PAI-negative isolates (23%) possessed one of the two alleles. Nine of 23 *cag* PAI-positive isolates (39%) possessed HP1078 allele A while 10 of 23 of these isolates (43%) possess allele type B. None of the *H. pylori* isolates generated amplicons with both primer pairs.

cagA positive
n=23



HP1078 allele type
A (26695-type).

cagA negative
n=13



HP1078 allele type
B (J99-type).

conclude that in the *H. pylori* isolate cohort examined, the variability seen could be characterized by the presence of two distinct allelic forms of HP1078.

Comparative sequence analysis of the HP1078 locus from the seven *H. pylori* isolates was conducted. Strong homology appears to exist between three of these strains (26695, B105, and B140) and between another three of these isolates (J99, J166, and B177). The remaining strain, B258, contains a stop codon approximately 360 bp downstream of its start codon, while the HP1078 ORF from each of the remaining strains continues for approximately another 360 bp. The sequence immediately following the premature stop codon in B258 remains highly similar to the 3' ends of HP1078 alleles A and B in 26695 and J99. It is interesting to note that the *cag* PAI-positive isolate B258 appears to have both an inactivated HP1079 and an inactivated HP1078 as well. It is tempting to speculate that this could indicate that the loss of one gene product might render the other gene dispensable due to a possible cooperative action of the two gene products. Immediately after the apparent gene-inactivating mutation in HP1078 in isolate B258, the sequence runs through the HP1078 stop codon found in 26695 and into *nixA*. This may be interpreted as further evidence of the loss of function in HP1078 in this isolate. Taken together, our data indicates that HP1078 is subject to allelic variation among isolates and that variation appears to be defined by two major allelic types that we designate HP1078A and HP1078B.

HP1078 allelic variation across racial demographics. Of the 22 isolates from North American patients possessing HP1078, 44% (14/22) possess allele type A while 56% (18/22) possess allele type B (Table 2). Among the 16 isolates from African-American

	North American		African	
	cagA+ (n=39)	cagA- (n=33)	cagA+ (n=8)	cagA- (n=2)
HP1078 (either allele)	32 (82%)	6 (18%)	8 (100%)	1 (50%)
Allele A *	14 (44%)	2 (33%)	1 (12%)	1 (100%)
Allele B **	18 (56%)	4 (67%)	7 (88%)	0 (0%)

* - # (%) of hp1078 positive isolates possessing Allele type A (26695-like)
** - # (%) of hp1078 positive isolates possessing Allele type B (J99-like)

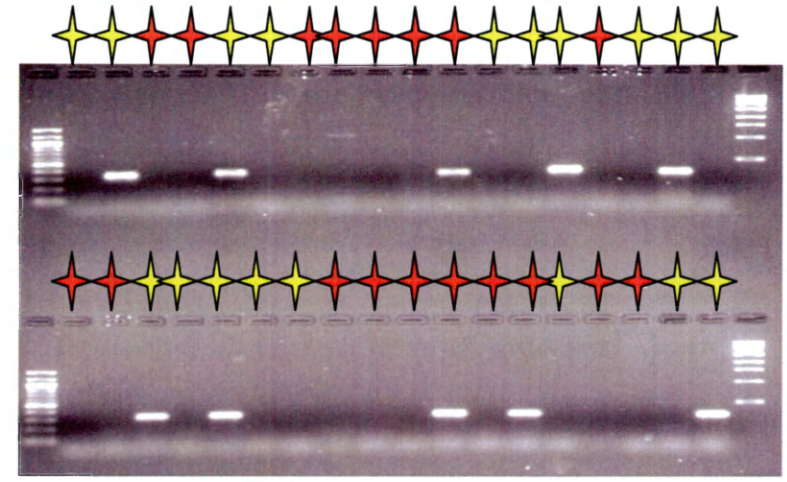
Table 2. Distribution of HP1078 Alleles. Strains are broken down by their geographic location as well as their potential virulence as determined by the presence of the *cag* Pathogenicity Island.

patients demonstrated to possess HP1078, 44% (7/16) were of allelic type A and 56% (9/16) of allelic type B compared to 50% allelic type A and 50% allelic type B seen in Caucasians. We also screened 10 isolates from patients from continental Africa (8 *cag* PAI-positives, 2 *cag* PAI-negatives), and found that 9 contained either an A or B allelic type of HP1078. Of the *cag* PAI-positive isolates, 7 (88%) typed positive for allele B while only 1 (22%) typed positive for allele A. Of the *cag* PAI-negative isolates, 1(50%) typed positive for allele A, and the remaining cag PAI-negative isolate did not appear to possess either HP1078 allelic type in our assays, suggesting the presence of even more allelic types. Thus data appears to show no substantial evidence for the predominance of one allelic type over the other when comparing isolates from North American Caucasians and African Americans (Fig. 7). It does, however, hint at a possible predominance of the B allele type in South African populations. Larger numbers of isolates from sub-Saharan Africans will be necessary to give this suggestion greater credence.

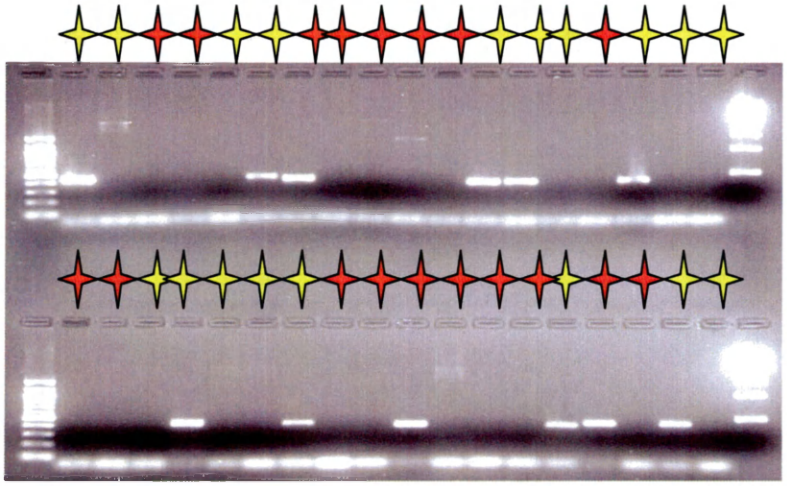
Figure 7. HP1078 allelic variation across racial demographics. 89% (16/18) isolates from African Americans possess an allele of *HP1078* (44% allele A and 56% allele B). 33% (6/18) isolates from Caucasians possess an allele of *HP1078* (50% allele A and 50% allele B) *however*, 83% of isolates from African-Americans were *cag* PAI-positive while 44% of those from Caucasians were *cag* PAI-positive.

African-American source patient

Caucasian source patient



***HP1078* allele type A (26695-type).**



***HP1078* allele type B (J99-type).**

Identification and Analysis of a Genetic Locus Undergoing Genetic Decay Simultaneously in Divergent Strains of *Helicobacter pylori*

RESULTS

HP0938-0937 locus shows evidence of extensive genetic decay. Homology screens conducted via BLAST searches closely matched the *H. pylori* strain 26695 genes HP0938 and HP0937 with portions of a single open reading frame (ORF) found in numerous other bacterial species. PCR analysis of the hypothesized pseudogenes HP0937/HP0938 was carried out on two *cag* PAI-negative strains (194a and J68) and two *cag* PAI-positive strains (B238 and J166). Primers HP0939 ES and HP0936 ES were used to amplify the sequence intervening the conserved genes *proP* (HP0936) and *yckJ* (HP0939). Amplicons (~2kbp) were generated and cloned into pGEM T Easy and both strains sequenced. Sequences were then compared with each other as well as with 26695 and J99 sequences using MacVector sequence analysis software.

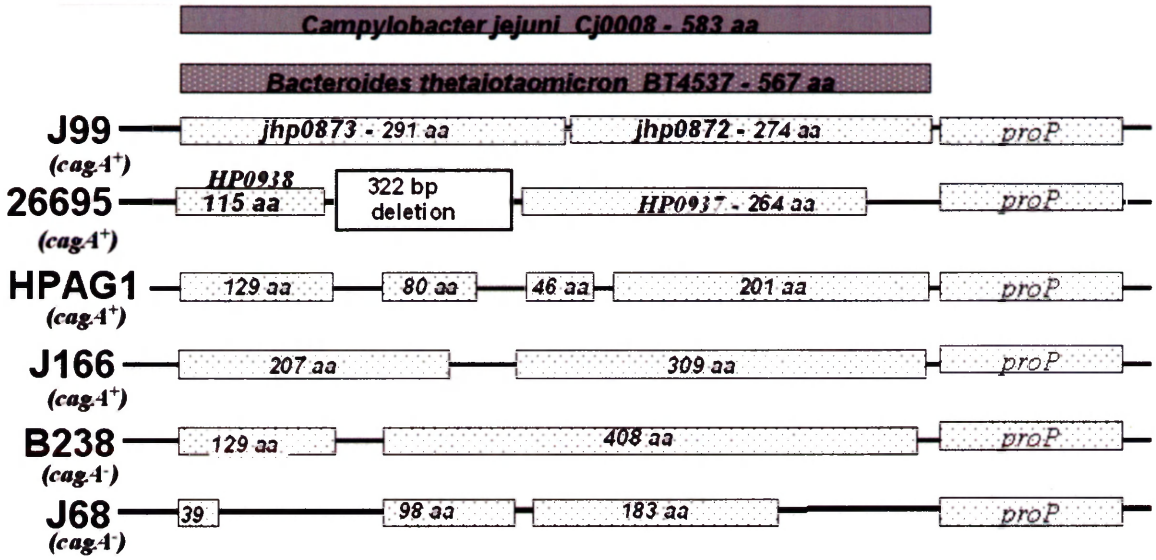
The resulting annotation of all orthologs of HP0937 and HP0938 are shown in Figure 8. The proposed ancestral gene that appears to still remain intact in the bacteria *Campylobacter jejuni* and *Bacteroides thetaiotaomicron* (closely and distantly related to *H. pylori*, respectively) is also shown for comparison. Comparison of these sequences suggest that HP0937 and HP0938 share significant sequence homology with many other bacterial species, however, the homologous sequences exist as a single gene of unknown function in all other species possessing the homolog. This suggests that HP0937 and HP0938 are, in fact, pseudogenes in *Helicobacter pylori*. Variation can be seen among these isolates in regard to the appearance of stop codons in this region, as well as a 322bp

deletion in the 26695 sequence, indicating that the genesis of this potential pseudogene was unique in each strain examined. We hypothesize that selection against this 'gene' developed relatively

recently. Alternatively, the simple absence of a functional constraint upon the gene product may have allowed this accumulation of mutations. The fact that their apparent patterns of pseudogenesis are so dissimilar indicates that the selective pressure occurred subsequent to the divergance of these strains from their last common ancestor. However, despite this, the sequence for each isolate is relatively well conserved over the 3' half of the locus, possibly suggesting that this sequence still may have a functional protein associated with it, despite the apparent loss of its 5' coding sequence.

We hypothesized that the function of the original gene product may have proven detrimental to *H. pylori* due to a change of host environments or the acquisition of virulence determinants that brought about a more pathogenic lifestyle. This has been shown to be the case in numerous bacteria, the best known being a deletional event characterized in *cadA* in *Shigella* [157]. *Helicobacter acinonychis*, a Helicobacter thought to have derived from *H. pylori* and found in the stomachs of large cats, shows severe degradation of its *vacA* gene [182]. *Neisseria meningitidis* and *Neisseria gonorrhoeae* both contain the locus containing the porin encoding *porA* gene, however, only *N. meningitidis* expresses the protein, due to the fact that *N. gonorrhoeae* strains contain multiple frameshift mutations at this locus including mutations in the promoter element that effectively prevent transcription from occuring [183]. In the current study, we demonstrated both *cag* PAI-positive and *cag* PAI-

Figure 8. Comparative Analysis of HP0938-O937 and its proposed ancestral homolog. The locus is shown here in four *cag*-PAI positive *H. pylori* strains and two *cag*-PAI negative *H. pylori* strains. Boxed regions indicate open reading frames predicted by the presence of start and stop codons, while lines indicate sequence between the ORFs. The homologous regions in *Bacteroides thetaiotaomicron* and *Campylobacter jejuni* represent the proposed ancestral gene. Despite the apparent genetic degradation in all *H. pylori* strains examined, there remains a large amount of conservation on the 3' end of this locus, indicating a potential selective purpose for its continuing presence in the genome.


Campylobacter jejuni Cj0008 - 583 aa
Bacteroides thetaiotaomicron BT4537 - 567 aa
J99
(cagA+)
jhp0873 - 291 aa
jhp0872 - 274 aa
proP
26695
(cagA+)
HP0938
115 aa
322 bp
deletion
HP0937 - 264 aa
proP
HPAG1
(cagA+)
129 aa
80 aa
46 aa
201 aa
proP
J166
(cagA+)
207 aa
309 aa
proP
B238
(cagA-)
129 aa
408 aa
proP
J68
(cagA-)
39
98 aa
183 aa
proP

negative *H. pylori* isolates contain early stop codons compared with the hypothetical, full-length, ancestral homolog present in numerous other bacteria. This contradicts our earlier hypothesis that the original function of the ancestral gene product was eliminated due to an incompatibility with a more virulent lifestyle. It does, however leave open the possibility that, rather than virulence, optimal colonization of the host stomach may have brought to bear the selective pressure that resulted in the strain-specific pseudogene remnants of the ancestral gene.

In order to test the hypothesis that colonization and/or virulence may be affected by the presence of the full-length ancestral homologue of HP0938-0937, we attempted to restore the ancestral gene's ability to encode a complete and functional product. *H. pylori* strain J99 provided the best chance of successfully repairing the ancestral gene, due to the fact that its inactivation appears to be the result of a single base pair switch, resulting in a premature stop codon (CAA to TAA) that interrupts the coding sequence approximately in the middle of the hypothetical protein.. We are currently undertaking to repair this single base pair using oligonucleotides that would change the single base pair and also introduce a novel restriction site into the intergenic region between HP0938 and HP0939, allowing for confirmation of our construct.

DISCUSSION

Identifying novel genes that may be involved in virulence and host colonization is of great importance in the study of bacterial pathogenesis. Previously identified and categorized virulence determinants are quite useful for co-localization studies that use them as markers while searching for other potential virulence determinants, especially in isolates obtained from patients with known pathologies. Both the genes HP1079 and HP1078 described here appear to be in linkage disequilibrium with a major virulence determinant in *Helicobacter pylori*, the *cag* Pathogenicity Island (*cag* PAI). These two genes have no homologs in any known bacterial species; however HP1079 does contain a highly conserved motif found widely among bacterial ATP hydrolyzing enzymes. We found that HP1079 was highly conserved among *H. pylori* isolates. We also showed it to be phase variable and the adjacent ORF, HP1078 appears to exist in at least two distinct allelic forms.

Regulation of gene expression throughout the colonization and infection of a pathogen is critically important. Phase variation provides a mechanism by which bacteria can produce populations containing a mixture of microbes possessing the same genes but expressing those genes either in the 'on' or 'off' positions. It is beneficial to a pathogen to have these contingency genes in both the 'on' and the 'off' phases in subpopulations simultaneously, especially in environments where the selective advantage of a contingency gene expression is often short-lived. Genes conferring enhanced pathogenic abilities to bacteria are often found to be regulated by phase variation [106, 184, 185]. A previous study from our lab demonstrated that HP1079 is present significantly more frequently in *H. pylori* isolates possessing *cagA* (PAI). In our current study, we have

documented a homopolymeric region in HP1079 where phase variation occurs due to slip-strand mispairing. This hypermutable sequence exists as a poly-adenosine tract approximately 403 bp downstream of the start codon. Two of the five isolates that we sequenced in this study were found to have a single adenine deletion in this region resulting in a frame shift mutation and the subsequent introduction of an in-frame stop codon well before the 'normal' stop codon.

We originally aimed to directly examine HP1079's potential role as a virulence determinant and investigate its possible involvement in host colonization and pathogenesis. A plasmid was constructed consisting of a truncated HP1079 gene interrupted with a Chloramphenicol Acetyl Transferase (CAT) gene to determine if the removal of this gene in a cag PAI-positive strain is detrimental to colonization/pathogenesis. We introduced our plasmid to *H. pylori* in the hopes that through natural transformation, the plasmid would be taken up and due to homologous recombination, our HP1079 construct would be effectively transferred to the bacterium. Transferring the bacteria to selection plates afterward would allow for the mutants to be isolated and PCR in addition to sequencing of the HP1079 locus would confirm our knock out. To date, this experiment has proven unsuccessful, although efforts to obtain this important mutant are continuing.

HP1079 has not previously been identified as a contingency gene. We hypothesize that the phase variation of HP1079 may indicate its involvement in the interaction with the host. Coupled with the fact that HP1079 exists in linkage disequilibrium with the *cag* PAI, we hypothesize that this gene product may be an uncharacterized factor involved in virulence. We continue to believe a mutant knockout

for HP1079 is achievable and that characterization of any associated phenotype with this mutant remains worthy of pursuit. Animal models such as Mongolian gerbils could be used to examine colonization and virulence effects in the knockout. In addition, placing HP1079/HP1078 genes into strains naturally lacking both HP1079 and HP1078 may allow us to examine the potential selective pressures at work within the genomes of less virulent strains of *H. pylori*.

The high degree of variability found between the two orthologs of HP1078 in the completely sequenced and annotated *H. pylori* strains was also investigated further in this study. The re-isolation of *H. pylori* from the original host of one of the completely sequenced strains (J99) allowed us the opportunity to examine any possible variability occurring within a six-year period in this particular strain [28]. After finding virtually no variation between the recent *H. pylori* isolates and the original, sequenced strain, we speculated that an evolutionary pressure may be present and acting to prevent significant genetic change from occurring in this gene. An alternative hypothesis would be that six years is an insufficient period of time for the accumulation of mutations, even in a gene whose product has been removed from functional constraints. Our subsequent PCR analyses and sequencing of the HP1079/HP1078 locus from several other *H. pylori* clinical isolates suggested the existence of a least two major allelic forms of the HP1078 gene. We attempted to determine whether either of these allelic types tended to be present more often in strains associated with more severe peptic ulcer disease. Currently, our findings indicate no such link between host clinical outcomes and HP1078 allelic type (data not shown).

When comparing isolates from African-Americans with those from Caucasians, we did find significantly more African-Americans possessed an HP1078 allele, however, this is likely due to the fact that significantly more of the isolates from African-American patients were *cag* PAI-positive (Fig. 5-6). A small pilot study comparing allelic variation in strains obtained from patients in continental Africa did find that 9 out of 10 strains tested positive for either allelic type of HP1078 and that allelic type B seemed to be the dominant allelic form present in that group (Table 2). The presence of multiple allelic forms often indicates a selective element present in the environment. In the case of pathogens, multiple allelic types may indicate an adaptation that confers increased survivability in differing host environments. Whether allowing for better defense against a host's immune response or higher binding affinity for a given cell type, allelic variation is often a sign of microbes adapting to their environment within their given host [105, 106, 186]. While we have currently been unable to correlate varying HP1078 allelic types with a specific selective pressure, the fact that HP1078 (as well as the adjacent HP1079) co-vary with the *cag* PAI indicates a potential role in virulence.

Understanding the various conditions and mechanisms involved in the emergence of bacterial pathogenesis is important if we wish to develop novel treatments for disease. From an evolutionary standpoint, tracking the emergence of a pathogen from a more free-living bacterium can show us commonalities among pathogens that may potentially lead to a better understanding of pathogenesis as a whole and the selective pressures that lead to the creation of more pathogenic microbes. Pseudogenes, while relatively rare in bacteria, can provide us with a partial snapshot of the evolutionary history of a developing microbe. Through careful analysis, previously protein-coding DNA can tell

us a great deal about an emerging pathogen; its rate/method of removal of foreign DNA, potential changes in its developing environment, co-existence and encounters with bacteriophages, and much more.

HP0937 and HP0938 appear to be the remnants of a single ancestral gene that continues to exist today as a single gene in numerous other bacterial species. Upon examining this genetic region among six different *H. pylori* clinical isolates, the region appears to suffer from dramatic genetic decay. What is even more striking is that no apparent grouping of the six strains could be created to explain the genetic damage over the course of time. This indicates to us that this region began degrading in numerous, unrelated strains independently, as none of the ancestral genes examined appear to be inactivated in the same way (Fig. 7). Assuming that the initial colonization of humans by *H. pylori* was a single evolutionary event, these findings indicate that this gene may have been inactivated only after the initial colonization of humans by *H. pylori*. Reasons for this lag in selection are not immediately apparent and further examination of this locus is necessary.

In order to approximate the rate of change in this pseudogene region, isolates can be examined over a period of time in order to identify conserved and plastic regions. An experimental approach similar to that carried out for the HP1079/HP1078 locus in J99 isolates taken from a patient six years after the initial microbes were isolated and characterized to examine the degradation of this locus. Our strain selection may be biased as each of the *H. pylori* clinical isolates used in this study was from North America. A more robust genomic analysis of this region using numerous strains from

distinct geographic regions would eliminate the possibility that what we have described here was only a localized event.

The fact that remnants of this apparently non-functional gene remain in the genome may suggest that either the inactivation of this gene was a rather recent evolutionary event, or that some form of selection continues to act on this genetic region, preventing its complete removal from the genome. This may be due to the proximity of critical sequences such as *nixA* or potentially HP1079. Unless a deletion event removes HP1078 without altering a sequence under positive selection, this event would be very rare. It may be possible that the products of the hypothetically fragmented ORFs are capable of functioning as dimers/trimers and continue to have a functional role in *H. pylori*. In order to address these questions, we must first generate some basic information concerning the hypothetical protein(s) itself.

Our current hypothesis is that the ancestral homolog of HP0937-0938 in *H. pylori* has undergone strong selection in order to eliminate the gene product from the proteome. If our knockout studies show no apparent deficiency in colonization or pathogenesis, then it would be helpful to determine if restoring the function of this pseudogene was detrimental to the bacterium and under what conditions. Successfully repairing the premature stop codon that causes the termination of HP0938 would allow us to restore the gene's functional protein. This restoration may not be that simple, however, due to the fact that other potential alterations in the predicted amino acid sequence may have accumulated and may prevent the translation of a fully functional product. In attempting to overcome this problem, it may be of value to replace this locus in *H. pylori* with a functional homolog from another bacterial species. *Campylobacter jejuni* would be a

fitting choice, given its significant homology with *Helicobacter pylori's* genome as a whole. It may also be beneficial to replace this locus with a functional homolog found in a free-living or commensal bacterium known to not cause disease in humans. If this gene is undergoing negative selection in human pathogens as a whole, then to successfully restore complete function, it may be necessary to acquire it from a non-pathogen.

In conclusion, we have found both HP1079 and HP1078 to exist in linkage disequilibrium with a major virulence determinant, the *cag* PAI, of *Helicobacter pylori*. HP1079 is a phase variable gene and HP1078 appears to exist in at least two allelic forms. We interpret these findings to suggest that these genes may play roles in virulence and host adaptation, respectively. In comparison, both HP0937 and HP0938 appear to exist in some form in all *H. pylori* strains sequenced in this study and their presence does not seem to co-vary with PAI status. It is worth noting, however, that given the apparent divergence observed in the genetic deterioration of this locus, the gene product that it originally encoded in an ancestral bacterium may be undergoing strong selection in *H. pylori*.

Works Cited

1. Marshall, B.J. and J.R. Warren, *Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration.* Lancet, 1984. **1**(8390): p. 1311-5.
2. Bardhan, P.K., *Epidemiological features of Helicobacter pylori infection in developing countries.* Clin Infect Dis, 1997. **25**(5): p. 973-8.
3. Bjorkholm, B., et al., *Mutation frequency and biological cost of antibiotic resistance in Helicobacter pylori.* Proc Natl Acad Sci U S A, 2001. **98**(25): p. 14607-12.
4. Hulten, K., et al., *Helicobacter pylori in the drinking water in Peru.* Gastroenterology, 1996. **110**(4): p. 1031-5.
5. Mazari-Hiriart, M., et al., *Helicobacter pylori and other enteric bacteria in freshwater environments in Mexico City.* Arch Med Res, 2001. **32**(5): p. 458-67.
6. Watson, C.L., et al., *Detection of Helicobacter pylori by PCR but not culture in water and biofilm samples from drinking water distribution systems in England.* J Appl Microbiol, 2004. **97**(4): p. 690-8.
7. Lu, Y., et al., *Isolation and genotyping of Helicobacter pylori from untreated municipal wastewater.* Appl Environ Microbiol, 2002. **68**(3): p. 1436-9.
8. Honda, K., et al., *High risk of Helicobacter pylori infection in young Japanese dentists.* J Gastroenterol Hepatol, 2001. **16**(8): p. 862-5.
9. Sinha, S.K., et al., *The incidence of Helicobacter pylori acquisition in children of a Canadian First Nations community and the potential for parent-to-child transmission.* Helicobacter, 2004. **9**(1): p. 59-68.
10. Solnick, J.V., et al., *Determination of the infectious dose of Helicobacter pylori during primary and secondary infection in rhesus monkeys (Macaca mulatta).* Infect Immun, 2001. **69**(11): p. 6887-92.
11. Ernst, P.B. and B.D. Gold, *The disease spectrum of Helicobacter pylori: the immunopathogenesis of gastroduodenal ulcer and gastric cancer.* Annu Rev Microbiol, 2000. **54**: p. 615-40.
12. Ashorn, M., *What are the specific features of Helicobacter pylori gastritis in children?* Ann Med, 1995. **27**(5): p. 617-20.

13. Whitney, A.E., et al., *Increased macrophage infiltration of gastric mucosa in Helicobacter pylori-infected children.* Dig Dis Sci, 2000. **45**(7): p. 1337-42.
14. Robinson, K., R.H. Argent, and J.C. Atherton, *The inflammatory and immune response to Helicobacter pylori infection.* Best Pract Res Clin Gastroenterol, 2007. **21**(2): p. 237-59.
15. Negrini, R., et al., *Helicobacter pylori infection induces antibodies cross-reacting with human gastric mucosa.* Gastroenterology, 1991. **101**(2): p. 437-45.
16. Kuipers, E.J., J.C. Thijs, and H.P. Festen, *The prevalence of Helicobacter pylori in peptic ulcer disease.* Aliment Pharmacol Ther, 1995. **9 Suppl 2**: p. 59-69.
17. Byrd, J.C., et al., *Inhibition of gastric mucin synthesis by Helicobacter pylori.* Gastroenterology, 2000. **118**(6): p. 1072-9.
18. Teixeira, A., et al., *Expression of mucins (MUC1, MUC2, MUC5AC, and MUC6) and type 1 Lewis antigens in cases with and without Helicobacter pylori colonization in metaplastic glands of the human stomach.* J Pathol, 2002. **197**(1): p. 37-43.
19. Matsuzwa, M., et al., *Helicobacter pylori infection up-regulates gland mucous cell-type mucins in gastric pyloric mucosa.* Helicobacter, 2003. **8**(6): p. 594-600.
20. Fitzgerald, J.R. and J.M. Musser, *Evolutionary genomics of pathogenic bacteria.* Trends Microbiol, 2001. **9**(11): p. 547-53.
21. Bamford, K.B., et al., *Helicobacter pylori: comparison of DNA fingerprints provides evidence for intrafamilial infection.* Gut, 1993. **34**(10): p. 1348-50.
22. Fraser, A.G., et al., *DNA fingerprints of Helicobacter pylori before and after treatment with omeprazole.* J Clin Pathol, 1992. **45**(12): p. 1062-5.
23. Kivi, M., et al., *Concordance of Helicobacter pylori strains within families.* J Clin Microbiol, 2003. **41**(12): p. 5604-8.
24. Disotell, T.R., *Discovering human history from stomach bacteria.* Genome Biol, 2003. **4**(5): p. 213.
25. Falush, D., et al., *Traces of human migrations in Helicobacter pylori populations.* Science, 2003. **299**(5612): p. 1582-5.
26. Kersulyte, D., et al., *Differences in genotypes of Helicobacter pylori from different human populations.* J Bacteriol, 2000. **182**(11): p. 3210-8.

27. Bjorkholm, B., et al., *Comparison of genetic divergence and fitness between two subclones of Helicobacter pylori.* Infect Immun, 2001. **69**(12): p. 7832-8.
28. Israel, D.A., et al., *Helicobacter pylori genetic diversity within the gastric niche of a single human host.* Proc Natl Acad Sci U S A, 2001. **98**(25): p. 14625-30.
29. Lundin, A., et al., *Slow genetic divergence of Helicobacter pylori strains during long-term colonization.* Infect Immun, 2005. **73**(8): p. 4818-22.
30. Salama, N., et al., *A whole-genome microarray reveals genetic diversity among Helicobacter pylori strains.* Proc Natl Acad Sci U S A, 2000. **97**(26): p. 14668-73.
31. Censini, S., et al., *cag, a pathogenicity island of Helicobacter pylori, encodes type I-specific and disease-associated virulence factors.* Proc Natl Acad Sci U S A, 1996. **93**(25): p. 14648-53.
32. Cover, T.L. and M.J. Blaser, *Purification and characterization of the vacuolating toxin from Helicobacter pylori.* J Biol Chem, 1992. **267**(15): p. 10570-5.
33. Terry, C.E., et al., *Genomic Comparison of cag pathogenicity island (PAI)-positive and -negative Helicobacter pylori strains: identification of novel markers for cag PAI-positive strains.* Infect Immun, 2005. **73**(6): p. 3794-8.
34. Santos, A., et al., *New pathogenicity marker found in the plasticity region of the Helicobacter pylori genome.* J Clin Microbiol, 2003. **41**(4): p. 1651-5.
35. Kavermann, H., et al., *Identification and characterization of Helicobacter pylori genes essential for gastric colonization.* J Exp Med, 2003. **197**(7): p. 813-22.
36. Whittam, T.S. and A.C. Bumbaugh, *Inferences from whole-genome sequences of bacterial pathogens.* Curr Opin Genet Dev, 2002. **12**(6): p. 719-25.
37. Suerbaum, S., et al., *Free recombination within Helicobacter pylori.* Proc Natl Acad Sci U S A, 1998. **95**(21): p. 12619-24.
38. Garcia-Vallve, S., P.J. Janssen, and C.A. Ouzounis, *Genetic variation between Helicobacter pylori strains: gene acquisition or loss?* Trends Microbiol, 2002. **10**(10): p. 445-7.
39. Hacker, J., U. Hentschel, and U. Dobrindt, *Prokaryotic chromosomes and disease.* Science, 2003. **301**(5634): p. 790-3.

40. Daubin, V., E. Lerat, and G. Perriere, *The source of laterally transferred genes in bacterial genomes.* Genome Biol, 2003. **4**(9): p. R57.

41. Navarre, W.W., et al., *Selective silencing of foreign DNA with low GC content by the H-NS protein in Salmonella.* Science, 2006. **313**(5784): p. 236-8.

42. Hacker, J., et al., *Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extraintestinal Escherichia coli isolates.* Microb Pathog, 1990. **8**(3): p. 213-25.

43. Wren, B.W., *The yersiniae--a model genus to study the rapid evolution of bacterial pathogens.* Nat Rev Microbiol, 2003. **1**(1): p. 55-64.

44. Gal-Mor, O. and B.B. Finlay, *Pathogenicity islands: a molecular toolbox for bacterial virulence.* Cell Microbiol, 2006. **8**(11): p. 1707-19.

45. Hacker, J. and J.B. Kaper, *Pathogenicity islands and the evolution of microbes.* Annu Rev Microbiol, 2000. **54**: p. 641-79.

46. Klumpp, J. and T.M. Fuchs, *Identification of novel genes in genomic islands that contribute to Salmonella typhimurium replication in macrophages.* Microbiology, 2007. **153**(Pt 4): p. 1207-20.

47. Novick, R.P. and A. Subedi, *The SaPIs: mobile pathogenicity islands of Staphylococcus.* Chem Immunol Allergy, 2007. **93**: p. 42-57.

48. Scott, D.R., et al., *Mechanisms of acid resistance due to the urease system of Helicobacter pylori.* Gastroenterology, 2002. **123**(1): p. 187-95.

49. Ha, N.C., et al., *Supramolecular assembly and acid resistance of Helicobacter pylori urease.* Nat Struct Biol, 2001. **8**(6): p. 505-9.

50. Kabir, S., *The current status of Helicobacter pylori vaccines: a review.* Helicobacter, 2007. **12**(2): p. 89-102.

51. Eaton, K.A. and S. Krakowka, *Effect of gastric pH on urease-dependent colonization of gnotobiotic piglets by Helicobacter pylori.* Infect Immun, 1994. **62**(9): p. 3604-7.

52. Tsuda, M., et al., *A urease-negative mutant of Helicobacter pylori constructed by allelic exchange mutagenesis lacks the ability to colonize the nude mouse stomach.* Infect Immun, 1994. **62**(8): p. 3586-9.

53. Andrutis, K.A., et al., *Inability of an isogenic urease-negative mutant stain of Helicobacter mustelae to colonize the ferret stomach.* Infect Immun, 1995. **63**(9): p. 3722-5.

54. Kreiss, C., et al., *Safety of oral immunisation with recombinant urease in patients with Helicobacter pylori infection.* Lancet, 1996. **347**(9015): p. 1630-1.

55. Michetti, P., et al., *Oral immunization with urease and Escherichia coli heat-labile enterotoxin is safe and immunogenic in Helicobacter pylori-infected adults.* Gastroenterology, 1999. **116**(4): p. 804-12.

56. Tsujii, M., et al., *Mechanism for ammonia-induced promotion of gastric carcinogenesis in rats.* Carcinogenesis, 1995. **16**(3): p. 563-6.

57. Forman, D., *Helicobacter pylori and gastric cancer.* Scand J Gastroenterol Suppl, 1996. **220**: p. 23-6.

58. Graham, D.Y., et al., *Iatrogenic Campylobacter pylori infection is a cause of epidemic achlorhydria.* Am J Gastroenterol, 1988. **83**(9): p. 974-80.

59. Smoot, D.T., *How does Helicobacter pylori cause mucosal damage? Direct mechanisms.* Gastroenterology, 1997. **113**(6 Suppl): p. S31-4; discussion S50.

60. Takashima, M., et al., *Effects of Helicobacter pylori infection on gastric acid secretion and serum gastrin levels in Mongolian gerbils.* Gut, 2001. **48**(6): p. 765-73.

61. Windle, H.J., D. Kelleher, and J.E. Crabtree, *Childhood Helicobacter pylori infection and growth impairment in developing countries: a vicious cycle?* Pediatrics, 2007. **119**(3): p. e754-9.

62. O'Toole, P.W., M.C. Lane, and S. Porwollik, *Helicobacter pylori motility.* Microbes Infect, 2000. **2**(10): p. 1207-14.

63. Eaton, K.A., et al., *Colonization of gnotobiotic piglets by Helicobacter pylori deficient in two flagellin genes.* Infect Immun, 1996. **64**(7): p. 2445-8.

64. Croxen, M.A., et al., *The Helicobacter pylori chemotaxis receptor TlpB (HP0103) is required for pH taxis and for colonization of the gastric mucosa.* J Bacteriol, 2006. **188**(7): p. 2656-65.

65. Schreiber, S., et al., *The spatial orientation of Helicobacter pylori in the gastric mucus.* Proc Natl Acad Sci U S A, 2004. **101**(14): p. 5024-9.

66. Cooke, C.L., J.L. Huff, and J.V. Solnick, *The role of genome diversity and immune evasion in persistent infection with Helicobacter pylori.* FEMS Immunol Med Microbiol, 2005. **45**(1): p. 11-23.

67. Moran, A.P., *Lipopolysaccharide in bacterial chronic infection: Insights from Helicobacter pylori lipopolysaccharide and lipid A.* Int J Med Microbiol, 2007.

68. Andersen-Nissen, E., et al., *Evasion of Toll-like receptor 5 by flagellated bacteria.* Proc Natl Acad Sci U S A, 2005. **102**(26): p. 9247-52.

69. Goodwin, C.S., et al., *Unusual cellular fatty acids and distinctive ultrastructure in a new spiral bacterium (Campylobacter pyloridis) from the human gastric mucosa.* J Med Microbiol, 1985. **19**(2): p. 257-67.

70. Kostrzynska, M., et al., *Identification, characterization, and spatial localization of two flagellin species in Helicobacter pylori flagella.* J Bacteriol, 1991. **173**(3): p. 937-46.

71. Geis, G., et al., *Ultrastructure and chemical analysis of Campylobacter pylori flagella.* J Clin Microbiol, 1989. **27**(3): p. 436-41.

72. de Robertis, E. and C.M. Franchi, *Electron microscope observation on the fine structure of bacterial flagella.* Experimental Cell Research, 1951. **2**: p. 295-298.

73. Seidler, R.J. and M.P. Starr, *Structure of the flagellum of Bdellovibrio bacteriovorus.* J Bacteriol, 1968. **95**(5): p. 1952-5.

74. Fuerst, J.A. and A.C. Hayward, *The sheathed flagellum of Pseudomonas stizolobii.* J Gen Microbiol, 1969. **58**(2): p. 239-45.

75. Follett, E.A. and J. Gordon, *An Electron Microscope Study of Vibrio Flagella.* J Gen Microbiol, 1963. **32**: p. 235-9.

76. Hranitzky, K.W., et al., *Characterization of a flagellar sheath protein of Vibrio cholerae.* Infect Immun, 1980. **27**(2): p. 597-603.

77. Geis, G., et al., *Ultrastructure and biochemical studies of the flagellar sheath of Helicobacter pylori.* J Med Microbiol, 1993. **38**(5): p. 371-7.

78. Luke, C.J. and C.W. Penn, *Identification of a 29 kDa flagellar sheath protein in Helicobacter pylori using a murine monoclonal antibody.* Microbiology, 1995. **141 (Pt 3)**: p. 597-604.

79. Jones, A.C., et al., *A flagellar sheath protein of Helicobacter pylori is identical to HpaA, a putative N-acetylneuraminyllactose-binding hemagglutinin, but is not an adhesin for AGS cells.* J Bacteriol, 1997. **179**(17): p. 5643-7.

80. Luke, C.J., et al., *Identification of flagellar and associated polypeptides of Helicobacter (formerly Campylobacter) pylori.* FEMS Microbiol Lett, 1990. **59**(1-2): p. 225-30.

81. Doig, P. and T.J. Trust, *Identification of surface-exposed outer membrane antigens of Helicobacter pylori.* Infect Immun, 1994. **62**(10): p. 4526-33.

82. Skene, C., et al., *Helicobacter pylori flagella: antigenic profile and protective immunity.* FEMS Immunol Med Microbiol, 2007.

83. Lundstrom, A.M., et al., *HpaA shows variable surface localization but the gene expression is similar in different Helicobacter pylori strains.* Microb Pathog, 2001. **31**(5): p. 243-53.

84. Gay, N.J. and F.J. Keith, *Drosophila Toll and IL-1 receptor.* Nature, 1991. **351**(6325): p. 355-6.

85. Medzhitov, R., P. Preston-Hurlburt, and C.A. Janeway, Jr., *A human homologue of the Drosophila Toll protein signals activation of adaptive immunity.* Nature, 1997. **388**(6640): p. 394-7.

86. Rock, F.L., et al., *A family of human receptors structurally related to Drosophila Toll.* Proc Natl Acad Sci U S A, 1998. **95**(2): p. 588-93.

87. Trinchieri, G. and A. Sher, *Cooperation of Toll-like receptor signals in innate immune defence.* Nat Rev Immunol, 2007. **7**(3): p. 179-90.

88. Kirschning, C.J., et al., *Human toll-like receptor 2 confers responsiveness to bacterial lipopolysaccharide.* J Exp Med, 1998. **188**(11): p. 2091-7.

89. Yang, R.B., et al., *Toll-like receptor-2 mediates lipopolysaccharide-induced cellular signalling.* Nature, 1998. **395**(6699): p. 284-8.

90. Chow, J.C., et al., *Toll-like receptor-4 mediates lipopolysaccharide-induced signal transduction.* J Biol Chem, 1999. **274**(16): p. 10689-92.

91. Hemmi, H., et al., *A Toll-like receptor recognizes bacterial DNA.* Nature, 2000. **408**(6813): p. 740-5.

92. Hayashi, F., et al., *The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5.* Nature, 2001. **410**(6832): p. 1099-103.

93. Smith, K.D., et al., *Toll-like receptor 5 recognizes a conserved site on flagellin required for protofilament formation and bacterial motility.* Nat Immunol, 2003. **4**(12): p. 1247-53.

94. Miyake, K., *Innate recognition of lipopolysaccharide by Toll-like receptor 4-MD-2.* Trends Microbiol, 2004. **12**(4): p. 186-92.

95. Caroff, M., et al., *Structural and functional analyses of bacterial lipopolysaccharides.* Microbes Infect, 2002. **4**(9): p. 915-26.

96. Reaves, T.A., A.C. Chin, and C.A. Parkos, *Neutrophil transepithelial migration: role of toll-like receptors in mucosal inflammation.* Mem Inst Oswaldo Cruz, 2005. **100 Suppl 1**: p. 191-8.

97. Bosshart, H. and M. Heinzelmann, *Targeting bacterial endotoxin: two sides of a coin.* Ann N Y Acad Sci, 2007. **1096**: p. 1-17.

98. Appelmelk, B.J., et al., *Potential role of molecular mimicry between Helicobacter pylori lipopolysaccharide and host Lewis blood group antigens in autoimmunity.* Infect Immun, 1996. **64**(6): p. 2031-40.

99. Monteiro, M.A., et al., *Expression of histo-blood group antigens by lipopolysaccharides of Helicobacter pylori strains from asian hosts: the propensity to express type 1 blood-group antigens.* Glycobiology, 2000. **10**(7): p. 701-13.

100. Kirkland, T., et al., *Helicobacter pylori lipopolysaccharide can activate 70Z/3 cells via CD14.* Infect Immun, 1997. **65**(2): p. 604-8.

101. Lepper, P.M., et al., *Lipopolysaccharides from Helicobacter pylori can act as antagonists for Toll-like receptor 4.* Cell Microbiol, 2005. **7**(4): p. 519-28.

102. Moran, A.P., *Helicobacter pylori lipopolysaccharide-mediated gastric and extragastric pathology.* J Physiol Pharmacol, 1999. **50**(5): p. 787-805.

103. Wang, G., et al., *Lewis antigens in Helicobacter pylori: biosynthesis and phase variation.* Mol Microbiol, 2000. **36**(6): p. 1187-96.

104. Harper, M., et al., *Pasteurella multocida expresses two LPS glycoforms simultaneously but only a single form is required for virulence: identification of two acceptor specific heptosyl I transferases.* Infect Immun, 2007.

105. Khamri, W., et al., *Variations in Helicobacter pylori lipopolysaccharide to evade the innate immune component surfactant protein D.* Infect Immun, 2005. **73**(11): p. 7677-86.

106. Salaun, L., S. Ayraud, and N.J. Saunders, *Phase variation mediated niche adaptation during prolonged experimental murine infection with Helicobacter pylori.* Microbiology, 2005. **151**(Pt 3): p. 917-23.

107. Salaun, L. and N.J. Saunders, *Population-associated differences between the phase variable LPS biosynthetic genes of Helicobacter pylori.* BMC Microbiol, 2006. **6**: p. 79.

108. Raetz, C.R. and C. Whitfield, *Lipopolysaccharide endotoxins.* Annu Rev Biochem, 2002. **71**: p. 635-700.

109. Stead, C., et al., *A novel 3-deoxy-D-manno-octulosonic acid (Kdo) hydrolase that removes the outer Kdo sugar of Helicobacter pylori lipopolysaccharide.* J Bacteriol, 2005. **187**(10): p. 3374-83.

110. Ogawa, T., et al., *Endotoxic and immunobiological activities of a chemically synthesized lipid A of Helicobacter pylori strain 206-1.* FEMS Immunol Med Microbiol, 2003. **36**(1-2): p. 1-7.

111. Ogawa, T., et al., *Immunobiological activities of chemically defined lipid A from Helicobacter pylori LPS in comparison with Porphyromonas gingivalis lipid A and Escherichia coli-type synthetic lipid A (compound 506).* Vaccine, 1997. **15**(15): p. 1598-605.

112. Suerbaum, S. and M. Achtman, *Evolution of Helicobacter pylori: the role of recombination.* Trends Microbiol, 1999. **7**(5): p. 182-4.

113. Gunn, M.C., et al., *The significance of cagA and vacA subtypes of Helicobacter pylori in the pathogenesis of inflammation and peptic ulceration.* J Clin Pathol, 1998. **51**(10): p. 761-4.

114. Covacci, A. and R. Rappuoli, *Helicobacter pylori: after the genomes, back to biology.* J Exp Med, 2003. **197**(7): p. 807-11.

115. Segal, E.D., et al., *Induction of host signal transduction pathways by Helicobacter pylori.* Proc Natl Acad Sci U S A, 1997. **94**(14): p. 7595-9.

116. Sharma, S.A., et al., *Activation of IL-8 gene expression by Helicobacter pylori is regulated by transcription factor nuclear factor-kappa B in gastric epithelial cells.* J Immunol, 1998. **160**(5): p. 2401-7.

117. Ando, T., et al., *A Helicobacter pylori restriction endonuclease-replacing gene, hrgA, is associated with gastric cancer in Asian strains.* Cancer Res, 2002. **62**(8): p. 2385-9.

118. Cao, P. and T.L. Cover, *Two different families of hopQ alleles in Helicobacter pylori.* J Clin Microbiol, 2002. **40**(12): p. 4504-11.

119. Occhialini, A., et al., *Distribution of open reading frames of plasticity region of strain J99 in Helicobacter pylori strains isolated from gastric carcinoma and gastritis patients in Costa Rica.* Infect Immun, 2000. **68**(11): p. 6240-9.

120. Cover, T.L. and S.R. Blanke, *Helicobacter pylori VacA, a paradigm for toxin multifunctionality.* Nat Rev Microbiol, 2005. **3**(4): p. 320-32.

121. Atherton, J.C., et al., *Mosaicism in vacuolating cytotoxin alleles of Helicobacter pylori. Association of specific vacA types with cytotoxin production and peptic ulceration.* J Biol Chem, 1995. **270**(30): p. 17771-7.

122. Rudi, J., et al., *Diversity of Helicobacter pylori vacA and cagA genes and relationship to VacA and CagA protein expression, cytotoxin production, and associated diseases.* J Clin Microbiol, 1998. **36**(4): p. 944-8.

123. Ilver, D., et al., *Helicobacter pylori adhesin binding fucosylated histo-blood group antigens revealed by retagging.* Science, 1998. **279**(5349): p. 373-7.

124. Gerhard, M., et al., *Clinical relevance of the Helicobacter pylori gene for blood-group antigen-binding adhesin.* Proc Natl Acad Sci U S A, 1999. **96**(22): p. 12778-83.

125. Colbeck, J.C., et al., *Genotypic profile of the outer membrane proteins BabA and BabB in clinical isolates of Helicobacter pylori.* Infect Immun, 2006. **74**(7): p. 4375-8.

126. Hennig, E.E., J.M. Allen, and T.L. Cover, *Multiple chromosomal loci for the babA gene in Helicobacter pylori.* Infect Immun, 2006. **74**(5): p. 3046-51.

127. Mahdavi, J., et al., *Helicobacter pylori SabA adhesin in persistent infection and chronic inflammation.* Science, 2002. **297**(5581): p. 573-8.

128. Backstrom, A., et al., *Metastability of Helicobacter pylori bab adhesin genes and dynamics in Lewis b antigen binding.* Proc Natl Acad Sci U S A, 2004. **101**(48): p. 16923-8.

129. Unemo, M., et al., *The sialic acid binding SabA adhesin of Helicobacter pylori is essential for nonopsonic activation of human neutrophils.* J Biol Chem, 2005. **280**(15): p. 15390-7.

130. Petersson, C., et al., *Helicobacter pylori SabA adhesin evokes a strong inflammatory response in human neutrophils which is down-regulated by the neutrophil-activating protein.* Med Microbiol Immunol (Berl), 2006.

131. Yamaoka, Y., et al., *Role of interferon-stimulated responsive element-like element in interleukin-8 promoter in Helicobacter pylori infection.* Gastroenterology, 2004. **126**(4): p. 1030-43.

132. Kudo, T., et al., *Regulation of RANTES promoter activation in gastric epithelial cells infected with Helicobacter pylori.* Infect Immun, 2005. **73**(11): p. 7602-12.

133. Yamaoka, Y., D.H. Kwon, and D.Y. Graham, *A M(r) 34,000 proinflammatory outer membrane protein (oipA) of Helicobacter pylori.* Proc Natl Acad Sci U S A, 2000. **97**(13): p. 7533-8.

134. Yamaoka, Y., et al., *Helicobacter pylori outer membrane proteins and gastroduodenal disease.* Gut, 2006. **55**(6): p. 775-81.

135. Peek, R.M., Jr., et al., *Adherence to gastric epithelial cells induces expression of a Helicobacter pylori gene, iceA, that is associated with clinical outcome.* Proc Assoc Am Physicians, 1998. **110**(6): p. 531-44.

136. Tomb, J.F., et al., *The complete genome sequence of the gastric pathogen Helicobacter pylori.* Nature, 1997. **388**(6642): p. 539-47.

137. Lu, H., et al., *Duodenal ulcer promoting gene of Helicobacter pylori.* Gastroenterology, 2005. **128**(4): p. 833-48.

138. Garcia-Medina, R., et al., *Pseudomonas aeruginosa acquires biofilm-like properties within airway epithelial cells.* Infect Immun, 2005. **73**(12): p. 8298-305.

139. Faruque, S.M., et al., *Transmissibility of cholera: in vivo-formed biofilms and their relationship to infectivity and persistence in the environment.* Proc Natl Acad Sci U S A, 2006. **103**(16): p. 6350-5.

140. Brady, R.A., et al., *Identification of Staphylococcus aureus proteins recognized by the antibody-mediated immune response to a biofilm infection.* Infect Immun, 2006. **74**(6): p. 3415-26.

141. Joshua, G.W., et al., *A Caenorhabditis elegans model of Yersinia infection: biofilm formation on a biotic surface.* Microbiology, 2003. **149**(Pt 11): p. 3221-9.

142. Cerca, N., et al., *Comparative antibody-mediated phagocytosis of Staphylococcus epidermidis cells grown in a biofilm or in the planktonic state.* Infect Immun, 2006. **74**(8): p. 4849-55.

143. Nichols, W.W., et al., *The penetration of antibiotics into aggregates of mucoid and non-mucoid Pseudomonas aeruginosa.* J Gen Microbiol, 1989. **135**(5): p. 1291-303.

144. Stark, R.M., et al., *Biofilm formation by Helicobacter pylori.* Lett Appl Microbiol, 1999. **28**(2): p. 121-6.

145. Cole, S.P., et al., *Characterization of monospecies biofilm formation by Helicobacter pylori.* J Bacteriol, 2004. **186**(10): p. 3124-32.

146. Azevedo, N.F., et al., *Shear stress, temperature, and inoculation concentration influence the adhesion of water-stressed Helicobacter pylori to stainless steel 304 and polypropylene.* Appl Environ Microbiol, 2006. **72**(4): p. 2936-41.

147. Coticchia, J.M., et al., *Presence and density of Helicobacter pylori biofilms in human gastric mucosa in patients with peptic ulcer disease.* J Gastrointest Surg, 2006. **10**(6): p. 883-9.

148. Ochman, H., J.G. Lawrence, and E.A. Groisman, *Lateral gene transfer and the nature of bacterial innovation.* Nature, 2000. **405**(6784): p. 299-304.

149. Bennett, P.M., *Genome plasticity: insertion sequence elements, transposons and integrons, and DNA rearrangement.* Methods Mol Biol, 2004. **266**: p. 71-113.

150. Brussow, H., C. Canchaya, and W.D. Hardt, *Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion.* Microbiol Mol Biol Rev, 2004. **68**(3): p. 560-602.

151. Lawrence, J.G., R.W. Hendrix, and S. Casjens, *Where are the pseudogenes in bacterial genomes?* Trends Microbiol, 2001. **9**(11): p. 535-40.

152. Bergthorsson, U. and H. Ochman, *Distribution of chromosome length variation in natural isolates of Escherichia coli.* Mol Biol Evol, 1998. **15**(1): p. 6-16.

153. Wernegreen, J.J., et al., *Decoupling of genome size and sequence divergence in a symbiotic bacterium.* J Bacteriol, 2000. **182**(13): p. 3867-9.

154. Kunin, V. and C.A. Ouzounis, *The balance of driving forces during genome evolution in prokaryotes.* Genome Res, 2003. **13**(7): p. 1589-94.

155. Parkhill, J., et al., *Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica.* Nat Genet, 2003. **35**(1): p. 32-40.

156. Garnier, T., et al., *The complete genome sequence of Mycobacterium bovis.* Proc Natl Acad Sci U S A, 2003. **100**(13): p. 7877-82.

157. Maurelli, A.T., et al., *"Black holes" and bacterial pathogenicity: a large genomic deletion that enhances the virulence of Shigella spp. and enteroinvasive Escherichia coli.* Proc Natl Acad Sci U S A, 1998. **95**(7): p. 3943-8.

158. Day, W.A., Jr., R.E. Fernandez, and A.T. Maurelli, *Pathoadaptive mutations that enhance virulence: genetic organization of the cadA regions of Shigella spp.* Infect Immun, 2001. **69**(12): p. 7471-80.

159. Moore, R.A., et al., *Contribution of gene loss to the pathogenic evolution of Burkholderia pseudomallei and Burkholderia mallei.* Infect Immun, 2004. **72**(7): p. 4172-87.

160. Cummings, C.A., et al., *Bordetella species are distinguished by patterns of substantial gene loss and host adaptation.* J Bacteriol, 2004. **186**(5): p. 1484-92.

161. Pedulla, M.L., et al., *Origins of highly mosaic mycobacteriophage genomes.* Cell, 2003. **113**(2): p. 171-82.

162. Weinbauer, M.G. and F. Rassoulzadegan, *Are viruses driving microbial diversification and diversity?* Environ Microbiol, 2004. **6**(1): p. 1-11.

163. Meeks, J.C., et al., *An overview of the genome of Nostoc punctiforme, a multicellular, symbiotic cyanobacterium.* Photosynth Res, 2001. **70**(1): p. 85-106.

164. Lerat, E. and H. Ochman, *Recognizing the pseudogenes in bacterial genomes.* Nucleic Acids Res, 2005. **33**(10): p. 3125-32.

165. Wick, L.M., et al., *Evolution of genomic content in the stepwise emergence of Escherichia coli O157:H7.* J Bacteriol, 2005. **187**(5): p. 1783-91.

166. Heintschel von Heinegg, E., H.P. Nalik, and E.N. Schmid, *Characterisation of a Helicobacter pylori phage (HP1).* J Med Microbiol, 1993. **38**(4): p. 245-9.

167. Andersson, J.O. and S.G. Andersson, *Pseudogenes, junk DNA, and the dynamics of Rickettsia genomes.* Mol Biol Evol, 2001. **18**(5): p. 829-39.

168. Jordan, I.K., et al., *Essential genes are more evolutionarily conserved than are nonessential genes in bacteria.* Genome Res, 2002. **12**(6): p. 962-8.

169. Day, M., *Bacterial sensitivity to bacteriophage in the aquatic environment.* Sci Prog, 2004. **87**(Pt 3): p. 179-91.

170. Gorski, A. and B. Weber-Dabrowska, *The potential role of endogenous bacteriophages in controlling invading pathogens.* Cell Mol Life Sci, 2005. **62**(5): p. 511-9.

171. Ochman, H. and L.M. Davalos, *The nature and dynamics of bacterial genomes.* Science, 2006. **311**(5768): p. 1730-3.

172. Moran, N.A. and G.R. Plague, *Genomic changes following host restriction in bacteria.* Curr Opin Genet Dev, 2004. **14**(6): p. 627-33.

173. Lerat, E. and H. Ochman, *Psi-Phi: exploring the outer limits of bacterial pseudogenes.* Genome Res, 2004. **14**(11): p. 2273-8.

174. Casadevall, A., *Cards of Virulence and the Global Virulome for Humans*, in *Microbe*. 2006. p. 359-364.

175. Breitbart, M. and F. Rohwer, *Here a virus, there a virus, everywhere the same virus?* Trends Microbiol, 2005. **13**(6): p. 278-84.

176. Filippini, M., et al., *Infection paradox: high abundance but low impact of freshwater benthic viruses.* Appl Environ Microbiol, 2006. **72**(7): p. 4893-8.

177. Boneca, I.G., et al., *A revised annotation and comparative analysis of Helicobacter pylori genomes.* Nucleic Acids Res, 2003. **31**(6): p. 1704-14.

178. Glazebrook, J. and F.M. Ausubel, *Isolation of phytoalexin-deficient mutants of Arabidopsis thaliana and characterization of their interactions with bacterial pathogens.* Proc Natl Acad Sci U S A, 1994. **91**(19): p. 8955-9.

179. Aras, R.A., et al., *Plasticity of repetitive DNA sequences within a bacterial (Type IV) secretion system component.* J Exp Med, 2003. **198**(9): p. 1349-60.

180. Alm, R.A. and T.J. Trust, *Analysis of the genetic diversity of Helicobacter pylori: the tale of two genomes.* J Mol Med, 1999. **77**(12): p. 834-46.

181. McNulty, S.L., et al., *Novel 180- and 480-base-pair insertions in African and African-American strains of Helicobacter pylori.* J Clin Microbiol, 2004. **42**(12): p. 5658-63.

182. Eppinger, M., et al., *Who ate whom? Adaptive helicobacter genomic changes that accompanied a host jump from early humans to large felines.* PLoS Genet, 2006. **2**(7): p. e120.

183. Feavers, I.M. and M.C. Maiden, *A gonococcal porA pseudogene: implications for understanding the evolution and pathogenicity of Neisseria gonorrhoeae.* Mol Microbiol, 1998. **30**(3): p. 647-56.

184. Stern, A. and T.F. Meyer, *Common mechanism controlling phase and antigenic variation in pathogenic neisseriae.* Mol Microbiol, 1987. **1**(1): p. 5-12.

185. de Vries, N., et al., *Transcriptional phase variation of a type III restriction-modification system in Helicobacter pylori.* J Bacteriol, 2002. **184**(23): p. 6615-23.

186. Bergman, M., et al., *Helicobacter pylori phase variation, immune modulation and gastric autoimmunity.* Nat Rev Microbiol, 2006. **4**(2): p. 151-9.