Helicobacter pylori: Comparative genomics and structure-function analysis of the flagellum biogenesis protein HP0958

A Thesis Presented in Partial Fulfilment of the Requirements for the Degree of

Doctor of Philosophy

by

Ceara de Lacy Clancy, B.Sc.

School of Microbiology

National University of Ireland, Cork

Supervisor:

Prof. Paul W. O'Toole

Head of School of Microbiology:

Prof. Gerald Fitzgerald

February 2014

Table of Contents

List of Tables

List of Figures

List of Appendices

Declaration

I hereby declare that the content of this thesis is the result of my own work and has not been submitted for another degree either at University College Cork or elsewhere.

_____

Ceara de Lacy Clancy

This thesis is dedicated to my parents.

“I know I was born and I know that I’ll die, the in between is mine.”

Pearl Jam

Abstract

Helicobacter pylori is a gastric pathogen which infects ~50% of the global population and can lead to the development of gastritis, gastric and duodenal ulcers and carcinoma. Genome sequencing of *H. pylori* revealed high levels of genetic variability; this pathogen is known for its adaptability due to mechanisms including phase variation, recombination and horizontal gene transfer. Motility is essential for efficient colonisation by *H. pylori*. The flagellum is a complex nanomachine which has been studied in detail in *E. coli* and *Salmonella*. In *H. pylori*, key differences have been identified in the regulation of flagellum biogenesis, warranting further investigation.

In this study, the genomes of two *H. pylori* strains (CCUG 17874 and P79) were sequenced and published as draft genome sequences. Comparative studies identified the potential role of restriction modification systems and the *comB* locus in transformation efficiency differences between these strains. Core genome analysis of 43 *H. pylori* strains including 17874 and P79 defined a more refined core genome for the species than previously published. Comparative analysis of the genome sequences of strains isolated from individuals suffering from *H. pylori*-related diseases resulted in the identification of "disease-specific" genes.

Structure-function analysis of the essential motility protein HP0958 was performed to elucidate its role during flagellum assembly in *H. pylori*. The previously reported HP0958-FliH interaction could not be substantiated in this study and appears to be a false positive. Site-directed mutagenesis confirmed that the coiled-coil domain of HP0958 is involved in the interaction with RpoN (74-284), while the Zn-finger domain is required for direct interaction with the full length *flaA* mRNA transcript. Complementation of a non-motile *hp0958*-null derivative strain of P79 with site-directed mutant alleles of *hp0958* resulted in cells producing flagellar-type extrusions from non-polar positions. Thus, HP0958 may have a novel function in spatial localisation of flagella in *H. pylori*.

Abbreviations

All abbreviations and units used in this thesis and not specified in this list are standard International System of Units.

aa	Amino acid
ABC	ATPase-binding cassette
ABI	Applied Biosystems
ABL	Abelson murine leukemia viral oncogene homolog 1
ACT	Artemis Comparison Tool
Amp	Ampicillin
APS	Ammonium persulphate
ATP	Adenosine triposphate
BabA	Blood group antigen binding protein A
BHI	Brain heart infusion
cag	Cytotoxin-associated gene
CBA	Columbia base agar
cDNA	Complementary DNA
Cm	Chloramphenicol
COG	Cluster of orthologous group of proteins
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclase
dNTP	Deoxynucleoside triphosphate
DTT	DL-dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
ELISA	Enzyme-linked immunosorbent assay
Em	Erythromycin
EMSA	Electrophoretic migration shift assay
FAK	Focal adhesion kinase
FBS	Foetal bovine serum
GI	Gastrointestinal
hr(s)	Hour(s)

Kan	Kanamycin
kDa	Kilo Dalton
IL	Interleukin
IPTG	Isopropyl β-D-thiogalactopyranoside
LB	Luria Bertani
LPS	Lipopolysaccharide
MAP	Mitogen-activated protein
min	Minute(s)
MOPS	3-(N-morpholino) propanesulphonic acid
mRNA	Messenger RNA
NCBI	National Center for Biotechnology Information
NGS	Next generation sequencing
OD	Optical density
o/n	Overnight
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PAI	Pathogenicity island
PAR1/MARK	Partitioning-defective 1/microtubule affinity-regulating kinase
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PFGE	Pulse field gel electrophoresis
PZ	Plasticity zone
qRT-PCR	Quantitative Real-time PCR
Rif	Rifampicin
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Revolution per minute
RT	Reverse transcription
s	Second(s)
SabA	Sialic acid-binding protein A
SAM	S-adenosyl-methionine
SDS	Sodium dodecyl sulphate
SHP-2	SRC homology 2 domain-containing phosphatase
SMase	Sphingomyelinase

SOE	Splicing by overlap extension (by PCR)
SSC	Saline sodium citrate buffer
T1SS	Type I secretion system
T2SS	Type II secretion system
T3SS	Type III secretion system
T4SS	Type IV secretion system
T5SS	Type V secretion system
T6SS	Type VI secretion system
Tat	Twin-arginine translocation (system)
TE	Tris EDTA
Tet	Tetracycline
TEMED	N,N,N',N' tetramethylethylenediamine
TIGR	The Institute for Genomic Research
Tris	Tris(hydroxylmethyl)methylamine
tRNA	Transfer RNA
U	Unit (enzymatic activity unit)
UV	Ultraviolet (radiation)
vol	Volume
vac	Vacuolating cytotoxin B gene
w/v	Weight/volume

In addition, the conventional one-letter codes for amino-acids, deoxyribonucleosides and ribonucleosides were applied:

Amino-acids: A, R, N, D, C, E, Q, G, H, I, L, K, M, F, P, S, T, W, Y, V for alanine, arginine, asparagine, aspartic acid, cysteine, glutamate, glutamine, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, praline, serine, threonine, tryptophan, tyrosine and valine, respectively.

Deoxyribonucleosides: A, C, G, T for deoxyadenosine, deoxycytidine, deoxyguanosine and deoxythymidine, respectively.

Ribonucleosides: A, C, G, U for adenine, cytidine, guanine and uridine, respectively.

Chapter 1
Literature Review

1 *Helicobacter pylori*

1.1 Discovery of *H. pylori*

Helicobacter pylori is a Gram negative micro-aerophilic member of the ε-proteobacteria, order *Campylobacterales*, family *Helicobacteriaceae* (Goodwin *et al.*, 1989). The spiral cells of this organism have been observed in samples from animals and humans since the late 19th century but were not isolated until relatively recently due to their fastidious growth requirements (Marshall and Warren, 1984; Warren and Marshall, 1983). *H. pylori* is a human gastric pathogen, isolated by Marshall and Warren in 1982 (Marshall and Warren, 1984), which colonises the mucosal lining of the stomach. It is motile because of multiple polar sheathed flagella, which are vital for colonisation and persistence in the host (Eaton *et al.*, 1992; Yoshiyama and Nakazawa, 2000).

H. pylori was first identified as a member of the *Campylobacter* genus, originally named *Campylobacter pyloridis* (Goodwin *et al.*, 1986). However, when differences in 5S and 16S rRNA, cellular fatty acid composition and ultrastructure were identified, it was reclassified under the new genus, *Helicobacter* (Goodwin *et al.*, 1989). In 2002, Fox reported 24 named *Helicobacter* members (Fox, 2002) and today there are over 200 members of the genus listed (NCBI, 2013), with new species being isolated regularly, *e.g. H. macacae* (Fox *et al.*, 2007) and new *H. pylori* strains (Blanchard *et al.*, 2013). *Helicobacter pylori* was the first species to have the genome of more than one strain sequenced, with that of *H. pylori* 26695 completed in 1997 (Tomb *et al.*, 1997) followed by J99, facilitating comparative analyses (Alm and Trust, 1999).

The genus can be sub-divided into gastric and enterohepatic *Helicobacter* species. The former primarily occupy the antrum of the stomach, *e.g. H. heilmannii* and *H. felis*, while the latter occupy the intestinal crypts of their host, *e.g. H. pullorum* and *H. cinaedi* (Fox, 2002; Rossi and Hänninen, 2012; Smet *et al.*, 2011). *Helicobacter* species have been isolated from 142 vertebrate host species where they are associated with a wide range of disease types (Table 1) (Smet *et al.*, 2011).

Animal models used to study *H. pylori* include mouse, gnotobiotic piglet, Mongolian gerbil and guinea pig (Kusters *et al.*, 2006). Recently, the teleost fish *Danio rerio* (zebrafish) has been successfully used as a model organism for

investigation of the *H. pylori* CagA virulence factor (Neal *et al.*, 2013). A novel ex-*vivo* three-dimensional system, termed an "organoid", has been developed from gastric stem cells, an exciting alternative to conventional 2-dimensional mammalian tissue culture systems (Wroblewski *et al.*, 2013).

Table 1 Features of *H. pylori* species

Species	Mammalian hosts	Pathology or clinical presentation	Reference
H. pylori	Human, primate	Gastritis, peptic ulcer disease, gastric adenocarcinoma, MALT lymphoma	(Blaser, 1990; Dorer *et al.*, 2009; Nagini, 2012)
H. felis	Cat, dog, mouse, human	Gastritis in natural host; may cause peptic ulcers or gastric adenocarcinoma in mouse	(Fox *et al.*, 2002; Haesebrouck *et al.*, 2009; Lee *et al.*, 1988; Trebesius *et al.*, 2001)
H. mustelae	Ferret	Gastritis, peptic ulcer disease, gastric adenocarcinoma, MALT lymphoma	(Fox, 2002, 1994; Fox *et al.*, 1997)
H. acinonychis	Cheetah, tiger	Gastritis, peptic ulcer disease	(Cattoli *et al.*, 2000; Eaton *et al.*, 1993)
H. heilmannii	Human, dog, cat, monkey, cheetah, rat	Gastritis, dyspeptic symptoms, MALT lymphoma	(Andersen *et al.*, 1999; Stolte *et al.*, 997; Trebesius *et al.*, 2001)
H. hepaticus	Mouse, other rodents	Proliferative typhlocolitis, hepatitis, hepatocellular carcinoma	(Fox *et al.*, 1994; Ward *et al.*, 1994)
H. anseris	Goose		(Fox *et al.*, 2006)
H. suis	Pig, human	Gastritis, dyspepsia	(De Groote *et al.*, 1999; Joosten *et al.*, 2013)
H. bilis	Mouse, human	Bacteremia, cellulitis	(Fox *et al.*, 1995; Turvey *et al.*, 2012)
H. cinaedi	Human, monkey	Bacteremia, cellulitis	(Fox *et al.*, 2001; Kikuchi *et al.*, 2012)

1.2 Morphology and Physiology

Helicobacter pylori cells are generally 2 - 5 µm long and 0.5 - 1 µm wide (Figure 1). *H. pylori* is motile via its 2 - 6 unipolar, sheathed flagella (Geis *et al.*, 1993; Yoshiyama and Nakazawa, 2000). On solid media, cells form small translucent colonies of ~3 mm (Dunn *et al.*, 1997). The flagella are ~3 µm in length and often possess a bulb at the distal end (Goodwin *et al.*, 1985). It is hypothesised that the sheath may aid host colonization by protecting the flagellar filament from degradation by the acidic human stomach (Jones and Curry, 1989; Luket and Penn, 1995). Flagella-associated autotransporter protein, FaaA, is a VacA-like protein. It has been shown to localise to the flagellar sheath. An isogenic *faaA* mutant exhibits decreased motility and an impaired ability to colonise the stomach of mice, indicating its important role in motility and host colonisation (Radin *et al.*, 2013).

The bacterium is spiral *in vivo*, but assumes a rod-shaped or coccoid form when cultured *in vitro* (Benaissa *et al.*, 1996). Conflicting arguments suggested this coccoid form could indicate cell death, while others indicated it may be a dormant stage still capable of infection (Benaissa *et al.*, 1996; Catrenich and Makin, 1991; Kusters *et al.*, 1997). It has since been shown that the coccoid form can be sub-divided into two types: a viable coccoid form and a non-viable, degenerative coccoid form (Azevedo *et al.*, 2007; Saitoa *et al.*, 2003; Willén *et al.*, 2000). Induced coccoid forms of *H. pylori* are still capable of expression of *cagE* and *babA*, indicating that cells with this morphology are still possibly capable of infection (Poursina *et al.*, 2013).

Motility is a key feature of this organism, which is necessary for efficient colonisation of the host (Ottemann and Lowenthal, 2002). Motility and chemotaxis play a dual role in infection, because motile but *che*-negative derivative strains of *H. pylori* exhibit reduced host colonisation (Foynes *et al.*, 2000; Terry *et al.*, 2005). *H. pylori* can swim with curvilinear velocities of ~25 µm/s (Karim *et al.*, 1998). Motility appears more rapid than that of rod-shaped *E. coli* in a viscous solution indicating its ability to swim through the viscous mucosal lining of the stomach (Hazell *et al.*, 1986). Celli *et al.* showed that the mechanism by which *H. pylori* can move so rapidly through this mucus relates to the viscoelastic properties of mucin, a major component of the gastric mucosa (Section 2.4).

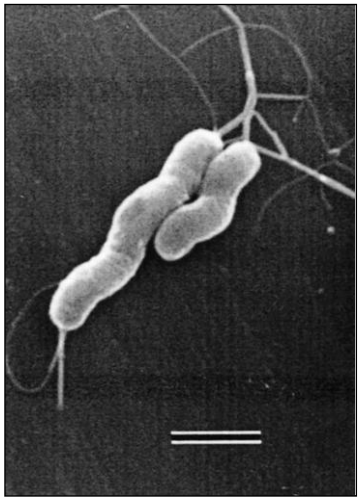


Figure 1 *Helicobacter pylori* spiral cells with polar sheathed flagella.

Electron microscopy image of *H. pylori* (Yoshiyama and Nakazawa, 2000).

H. pylori is catalase, oxidase and urease positive (Goodwin *et al.*, 1989; Kusters *et al.*, 2006). Superoxide dismutase and catalase production have a protective role for the bacterium against oxidation by phagocytes (Handa *et al.*, 2010). As *H. pylori* is acid-sensitive, the production of urease is a key virulence factor which aids infection and colonisation of the acidic human stomach (Hazell *et al.*, 1986). *H. pylori* is capable of biofilm formation, a feature used by pathogenic bacteria to aid in infection and survival in the host (Yonezawa *et al.*, 2010). Its biofilm extracellular polymeric substance (EPS) matrix includes polysaccharides and extracellular DNA, possibly providing an environment which promotes inter-strain recombination events (Grande *et al.*, 2011, 2012).

H. pylori is a fastidious organism which requires complex medium for growth. It is routinely grown on Columbia Agar Base or in Brucella broth or Brain Heart Infusion broth. Supplements include horse blood, activated charcoal, cornstarch, β-cyclodextrins or foetal calf serum (Buck *et al.*, 1987; Kusters *et al.*, 2006; Morgan *et al.*, 1987). In the presence of antibiotics used for primary culture on selective agar, inhibited growth of the bacteria can be circumvented through the addition of ferrous sulphate, sodium pyruvate, and mucin (Jiang and Doyle, 2000). Several amino acids are generally required for growth: leucine, valine, phenylalanine, methionine, arginine, and histidine (Nedenskov, 1994). Some amino acids can also serve as an

energy source: serine, alanine, proline and aspartate (Tomb *et al.*, 1997). It grows well in the pH range 5.5 – 8 (Morgan *et al.*, 1987).

Optimal culture conditions include a microaerobic environment of 5 - 15% O_2 and 5% CO_2 (Goodwin and Armstrong, 1990) and a temperature of 33 - 40.5°C (Goodwin *et al.*, 1986). *H. pylori* is a capnophile, *i.e.* it thrives in the presence of CO_2, which has been found to be more sensitive to aerobic environments when at low cell densities (Bury-Moné *et al.*, 2006). Thin-layered liquid culture of *H. pylori* facilitates efficient gas transfer and hence growth in a microaerobic environment (Joo *et al.*, 2010). Microaerobic conditions can be achieved using hypoxia chamber, CO_2-regulated incubator and CampyGen gas pack.

1.3 Genetics

There are now over 200 named members of the *Helicobacter* genus (Smet *et al.*, 2011). Today, the complete genomes of 52 *H. pylori* strains have been sequenced and annotated; these are available on the NCBI Genome web resource (NCBI, 2013). The draft genome sequences of a further 228 strains are also available, though some lack annotation. Typically, the complete genome of *H pylori* has an average size of 1.6 - 1.7 Mbp, a GC content of 35 - 39%; some strains possess bacteriophage DNA and approximately 50% carry cryptic plasmids (Alm and Trust, 1999; Baltrus *et al.*, 2009; Farnbacher *et al.*, 2010; Penfold *et al.*, 1988; Uchiyama *et al.*, 2012). Current sequencing information has identified that *H. pylori* has an open pan-genome, with ~1,200 core genes, and a coding density of 89 - 92% (Farnbacher *et al.*, 2010; Fischer *et al.*, 2010; Lara-Ramírez *et al.*, 2011). The gene content can differ by >10% between two given strains of *H. pylori* (Farnbacher *et al.*, 2010; Fischer *et al.*, 2010). Identical strains of *H. pylori* have only been isolated from family members, indicating intrafamilial transmission and highlighting the genetic variability of this organism (Achtman *et al.*, 1999; Linz *et al.*, 2007; Raymond *et al.*, 2008).

Comparative analysis of the genomes of *H. pylori* 26695 and J99 in 1997 indicated that *H. pylori* genomes display high levels of genetic recombination (Alm and Trust, 1999). The identification of transposable elements, repeat sequences and a large number of single nucleotide polymorphisms (SNPs) throughout the genome indicated a species with a high level of genomic plasticity. *H pylori* genomes have characteristic regions of high variability known as "plasticity zones" and pathogenicity islands (PAI) which encode key virulence genes *e.g. cagA* (see below)

and a large number of genes of unknown function. A large proportion of the genetic variation between strains can be localised to these regions which have a different GC content to the rest of the genome (Alm *et al.,* 1999; Kersulyte *et al.,* 2009). The *cag* PAI contains genes which encode components of a type IV secretion system, facilitating injection of this strain-specific virulence factor into host cells (TFSS) (Akopyants *et al.,* 1998; Censini *et al.,* 1996; Duncan *et al.,* 2013; Furuta *et al.,* 2011). Sequential sequencing of the genomes of *H. pylori* isolates from chronically infected subjects revealed genome evolution during infection. The high rate of recombination occurred at non-random sites throughout the genome; Kennemann *et al.* suggest that *H. pylori* initially imports long DNA fragments and these are subsequently fragmented and distributed to different locations (Kennemann *et al.,* 2011). Genes with a high recombination rate include those encoding outer membrane proteins and proteins involved in lipopolysaccharide synthesis (Yahara *et al.,* 2012).

Helicobacter pylori is naturally competent for the uptake of exogenous DNA (Nedenskov-Sorensen *et al.,* 1990). *H. pylori* is unique amongst naturally competent bacteria as it is the only known species which does not use pilus proteins during transformation. The proposed mechanism by which natural transformation is regulated involves the Type IV secretion system, ComB and is localised to cell poles (Hofreuter *et al.,* 2001; Stingl *et al.,* 2010). ComB facilitates DNA uptake through the outer membrane, while ComEC is required for the passage of DNA through the inner membrane (Stingl *et al.,* 2010). *H. pylori* can also transfer plasmids and chromosomal DNA between cells by conjugation (Fischer *et al.,* 2010; Heuermann and Haas, 1998; Kuipers *et al.,* 1998). Recently, the ComB system has also been identified as playing a role in DNaseI-resistant plasmid transfer. A novel T4SS-independent pathway termed the alternate DNaseI-Resistant pathway (ADR) has been identified, highlighting the important role of horizontal gene transfer in the genome flux of *H. pylori* (Rohrer *et al.,* 2012). Bacteriophage also contribute to this genetic variability through DNA transduction (Luo *et al.,* 2012).

Restriction modification (RM) systems are a protective strategy used by bacteria to prevent invasion of foreign DNA through the activity of restriction endonucleases. *H. pylori* strains possess RM systems that are strain-specific (Alm & Trust, 1999; Ando *et al.,* 2000; Tomb *et al.,* 1997). RM systems can be classified as types I - IV. Type II systems are the best understood of these, involving restriction endonuclease/DNA methyltransferase pairs of enzymes that have opposing

intracellular activities to cleave/methylate DNA at specific recognition sites, respectively (Xu *et al.*, 2000). There is a complex interplay of factors which ultimately determine the capacity of a given strain of *H. pylori* for horizontal gene transfer *e.g. nucT, dprB* and *ruvC* (Humbert and Salama, 2008; Humbert *et al.*, 2011).

.

2 Pathogenesis of *Helicobacter pylori*

2.1 Epidemiology

Helicobacter pylori has had an intimate association with humans for ~100,000 years (Moodley *et al.*, 2012). The movement of anatomically modern humans from Africa ~60,000 years ago matches the divergence of *H. pylori*, thereby indicating that *H. pylori* migrated throughout the globe with its human host (Linz *et al.*, 2007). *H. pylori* can be subdivided into 7 populations: hpAfrica1, hpAfrica2, hpNEAfrica, hpAsia2, hpEastAsia, hpSahul and hpEurope (Moodley *et al.*, 2009, 2012).

Although there is a high global incidence of *Helicobacter pylori*, infection has been shown to vary locally due to factors including geographical location, age, race, and socioeconomic status (Khalifa *et al.*, 2010). Levels of infection are generally higher in developing countries, with infection occurring often in early childhood (Figure 2) (Perez-Perez *et al.*, 2004; Pounder and Ng, 1995). Industrialised countries show lower rates of infection in children (Fiedorik *et al.*, 1991). In Western countries *e.g.* Germany and USA, the prevalence of infection is low in children and higher amongst those above the age of 50 (Prinz *et al.*, 2006). In China and Japan, the proportion of infected individuals is high (~60% of young members of the population and 80 - 90% of older members) (Dorji *et al.*, 2013; Inoue and Tsugane, 2005; Prinz *et al.*, 2006). Black and Hispanic ethnicities are associated with higher risk of infection, though different ethnicities can be broadly linked with differing social classes and hygiene standards (reviewed in (Brown, 2000)).

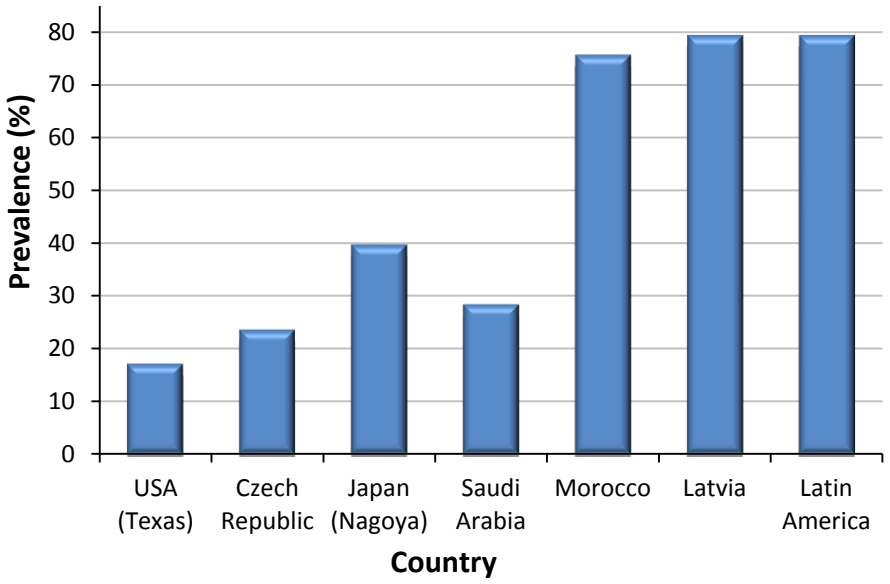


Figure 2 Prevalence of *H. pylori* infection in studies published during 2012.

Bar chart representation of data overviewed by (Calvet *et al.*, 2013).

2.2 Transmission

Approximately half the worldwide population harbours *H. pylori* (Hunt *et al.*, 2011; The Eurogast Study Group, 1993) and hence, eradication of this pathogenic bacteria is of global concern. However, because *H. pylori* can persist in the host for years before symptoms develop, it is often difficult to determine the point of infection, likely during childhood (Brown, 2000; Malaty *et al.*, 2002; Salama *et al.*, 2013). *H. pylori* is generally accepted to be acquired from close personal contact *i.e.* between family members (Fialho *et al.*, 2010; Raymond *et al.*, 2008).

There are several proposed routes of transmission of *H. pylori*: oral-oral, faecal-oral, gastro-oral and iatrogenic. The oral-oral route of transmission has been substantiated, with several studies isolating the pathogen from saliva, dental plaques and the oesophagus (Cellini *et al.*, 2010; Silva *et al.*, 2010; Zou and Li, 2011), while others suggest a link between pre-mastication of young infants' food and infection (Clemens *et al.*, 1996; Frenck and Clemens, 2003). The faecal-oral route is possibly another dominant mode of transmission, as DNA and viable *H. pylori* cells have been isolated from the faeces of infected hosts (Momtaz *et al.*, 2012; Parsonnet *et al.*, 1999). Improvement in sanitation standards in the United States achieved in the latter

half of the 19th century correlated with a reduction in the transmission of rates of *H. pylori*, supporting the likelihood of a predominantly oral/faecal-oral transmission route (Rupnow *et al.*, 2000).

Iatrogenic transmission has been found to occur through use of compromised medical equipment *e.g.* endoscopes (Brown, 2000). Transmission associated with gastroesophageal reflux and contact with vomitus has also been reported (Parsonnet *et al.*, 1999). Water has been suggested as a reservoir for *H. pylori* in the environment, both in freshwater streams and off-shore marine waters (reviewed in (Bellack *et al.*, 2006), (Twing *et al.*, 2011)). *H. pylori* has been isolated from a variety of non-human hosts including primates and domestic cats (Fox, 1995; Handt *et al.*, 1994). Other non-*pylori Helicobacters* have been isolated from humans, though this is not their primary host *e.g. H. suis* and *H. pullorum* (Table 1) (Haesebrouck *et al.*, 2009; Joosten *et al.*, 2013). Thus, the possibility of zoonotic transmission of this gastric pathogen, perhaps through close human contact, must also be considered.

2.3 Infection and Inflammation

Unlike the alkaline lumen of the bowel, the extremely acidic (pH 1 - 2) gastric lumen is a much more hostile environment for bacteria. Therefore, gastric colonisers such as *H. pylori* find their niche in the thick mucosal lining of the stomach and next to the epithelial cell surface (Hazell *et al.*, 1986). Urease secretion, motility through the use of its multiple polar flagella and a range of virulence factors facilitate *H. pylori* persistence and infection *e.g.* catalase and superoxide dismutase (Section 2.4, Figure 3).

Control of the local pH of the bacteria's environment is key to the survival of the bacteria in the stomach long enough to establish infection. Secretion of urease alters the pH of the gastric lumen by hydrolysing urea to produce ammonia and carbon dioxide. *H. pylori* activates cytoplasmic urease through a pH-controlled channel, UreI (Weeks, 2000). Unperturbed by this otherwise hostile environment, the motile bacteria can then swim to, and interact with, host epithelial cells, thereby eliciting inflammation (Ottemann and Lowenthal, 2002). *H. pylori* localises in the gastric antrum, where there are few acid-producing parietal cells (McNulty and Watson, 1984). It is also found deep in the mucosal layer of the corpus (Kuipers *et al.*, 1995).

H. pylori motility is inhibited by active pepsin, a peptidase which is active at low pH levels. Postprandial occurrences associated with bouts of luminal neutralisation could provide an opportunity for *H. pylori* infection (Schreiber *et al.*, 2005). This may help to explain why infection tends to occur during childhood as opposed to adulthood. In infants, this neutralising effect lasts for ~1 hr, while in adults it is as short as a few minutes. Thus, in infants and young children there is an extended postprandial period during which *H. pylori* can swim from the lumen to the gastric mucus and epithelial cells (Agunod *et al.*, 1969; Bücker *et al.*, 2012; Mitchell *et al.*, 2001).

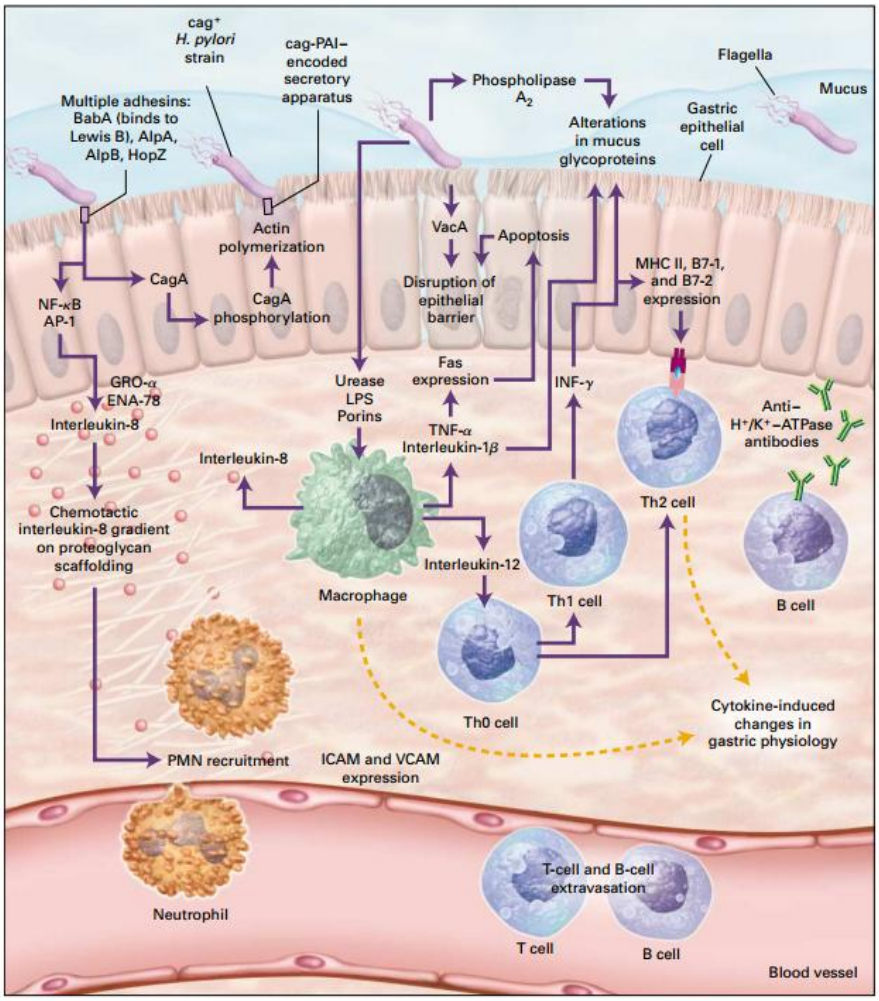


Figure 3 Pathogen-host interaction during infection.

Overview of the various points of bacterial-host interaction during *H. pylori* infection and host responses (Suerbaum & Michetti 2002).

Infection induces a host response which ultimately leads to inflammation and tissue damage, but not destruction of the invading pathogen. Upon reaching the

epithelial cells, host recognition of *H. pylori* peptidoglycan by NOD1 receptors initiates a cascade which activates NFκB (Viala *et al.*, 2004). Injection of the *H. pylori* CagA effector protein into host cells illicits an NFκB and Fas-mediated pro-inflammatory response (Lamb *et al.*, 2009). Recognition of *H. pylori* triggers a T helper type 1 (Th1) mucosal cell response which may induce gastritis by up-regulation of IL18, interferon-γ (INF-γ) and caspase 1 (Ghosh *et al.*, 2002; Tomita *et al.*, 2001). Infection is also associated with increased levels of caspase 3 which induces apoptosis of host cells, contributing to the development of gastric ulcers and cancers (Shimada *et al.*, 2008). Chemotaxis contributes to apoptosis induction through regulation of Th17 cells. Che$^-$ mutant strains of *H. pylori* trigger less inflammation and apoptosis, despite colonising to a similar extent when compared with wild type infections of a mouse model (Rolig *et al.*, 2011). Recently, a novel virulence factor, JHP940, has been described which activates an NFκB response through its activity as a eukaryotic-type Ser/Thr kinase (Kim *et al.*, 2010). *H. pylori* mediates life-long low-level inflammation in the host by altering these immune response pathways, leading to a number of clinical conditions.

2.4 Virulence Factors

A variety of virulence factors contribute to the pathogenicity of *Helicobacter pylori* (Backert and Clyne, 2011). These include urease, motility, *cag* PAI, VacA, adhesins and lipopolysaccharide (LPS). Urease production and motility are key virulence factors which work in tandem to enable *Helicobacter pylori* infection and persistence in the host. Urease is a nickel-containing metalloenzyme composed of two subunits, UreA and UreB. The active form of urease hydrolyzes urea to produce ammonia and carbon dioxide, which raises the pH. *H. pylori* is not an acidophile, and hence, activity of this enzyme is essential for survival in the gastric lumen which has a pH of ~2. Cytoplasmic urease allows the bacteria to maintain a neutral internal pH (Stingl *et al.*, 2001). There is also evidence of altruistic autolysis of *H. pylori* making urease available for outer membrane association (Phadnis *et al.*, 1996). Urease-deficient *H. pylori* are unable to colonise the stomachs of gnotobiotic piglets. Surprisingly, urease-deficient *H. pylori* are out-competed by urease-producing *H. pylori* in co-infection experiments of achlorhydric piglets, indicating another role for this enzyme beyond pH neutralisation (Eaton and Krakowka, 1994).

It was suggested that urease has a dual function as an adhesin, but this was later disproven (Clyne *et al.*, 1996). A study using the gerbil model of infection indicated that *H. pylori* spatial orientation in the gastric mucosa is dependent on pH gradient (Bahari *et al.*, 1982; Schreiber *et al.*, 2004). Ammonia and bicarbonate produced by the enzyme may also cause pathological effects including cytotoxicity and suppression of host bactericidal activity (Kuwahara *et al.*, 2000). Urease promotes survival of *H. pylori* engulfed by macrophages and inhibits opsonisation (Rokita *et al.*, 1998; Schwartz and Allen, 2006). It also indirectly disrupts the tight junctions of gastric epithelial cells, contributing to virulence (Wroblewski and Peek, 2011; Wroblewski *et al.*, 2009).

Urease activity, chemotaxis and motility are closely associated in their contribution to the virulence of *H. pylori*. Motility is essential for *H. pylori* infection and persistence in the host (Eaton *et al.*, 1992; Ottemann and Lowenthal, 2002). Motility enables the bacterium to travel to, and remain in, the host gastric mucosal layer and to interact with host epithelial cells through adhesins and other factors (see above). A number of flagellar proteins (FliQ, FliM and FliS) participate in adherence to AGS cells (Zhang *et al.*, 2002). Chemotactic sensing enables the bacteria to swim away from an acidic environment, *i.e.* the lumen, through membrane-bound or cytoplasmic chemoreceptors *e.g.* TlpD (Croxen *et al.*, 2006). In *Salmonella enterica* serovar Typhimurium the phosphorylated form of CheY interacts with flagellar switch protein FliM to initiate clockwise flagellar rotation resulting in tumbling and change of direction (Lowenthal *et al.*, 2009). Bacterial urease activity raises the local pH of the mucus, which is associated with a reduction in mucin viscosity. This altered rheology allows the bacteria to move more freely through the mucosal layer (Celli *et al.*, 2009). Cell shape contributes to *H. pylori* motility and colonisation ability. Cells lacking the characteristic helical twist of *H. pylori* display impaired motility in gel-like mucin at low pH and reduced colonisation capabilities (Sycuro *et al.*, 2010, 2012). *H. pylori* is known to interact with gastric mucins, glycosylated extracellular proteins which function in homeostasis and host protection (Van de Bovenkamp *et al.*, 2003). *H. pylori* reduces the rate of mucin production in a mouse model (Navabi *et al.*, 2013). An *in vitro* study using human gastric cancer cell lines also found that urease and flagellin can alter the expression profile of mucins (Perrais *et al.*, 2013). The flagellar-type secretion system is not exclusively used for flagellum biogenesis. Virulence proteins have been identified which are secreted through the

flagellar lumen *e.g.* YplA in *Yersinia enterocolitica* and CiaB in *Campylobacter* (Christensen *et al.*, 2009; Young *et al.*, 1999).Therefore, the importance of flagella and motility as an important virulence and colonisation factor should not be underestimated.

H. pylori strains are subdivided into two subsets on the basis of their Cytotoxin-associated gene A (*cagA*) status: Class I strains are *cagA* positive and are associated with increased virulence, Class II are *cagA* negative (Covacci *et al.*, 1993; Xiang *et al.*, 1995). CagA is a highly immunogenic protein which is injected into host epithelial cells *via* a Type IV secretion system (Covacci *et al.*, 1993). The *cagA* gene is present on a 40 kb PAI alongside genes for approximately 30 proteins (Figure 4) (Censini *et al.*, 1996). This includes the genes encoding a complete T4SS (Kutter *et al.*, 2008). The content of this PAI varies considerably between strains as it quite unstable and is subject to frequent inversion and deletion events (Akopyants *et al.*, 1998; Kauser *et al.*, 2004). Other genes in this PAI encode proteins with homology to several virulence-associated proteins, *e.g.* conjugative plasmids and heat shock proteins, thus highlighting its multifaceted role in *H. pylori* virulence (Akopyants *et al.*, 1998) (Akopyants *et al.*, 1998).

CagA is the first identified bacterial oncoprotein that functions in mammalian hosts (Ohnishi *et al.*, 2008). CagA-positive strains of *H. pylori* are associated with increased inflammation and development of more severe conditions such as gastric adenocarcinoma and B cell mucosa-associated lymphoid tissue (MALT) lymphoma (Murata-Kamiya, 2011; Wroblewski *et al.*, 2010). Exposed CagA on the bacterial surface triggers externalisation of host plasma membrane phosphatidylserine to which CagA binds and initiates entry into host cells (Murata-Kamiya *et al.*, 2010). CagA undergoes tyrosine phosphorylation by Src kinase or Abelson murine leukemia viral oncogene homolog 1 (ABL) kinase (Poppe *et al.*, 2007; Selbach *et al.*, 2002). Phosphorylated CagA can then bind to and activate SRC homology 2 domain-containing phosphatase (SHP-2) (Higashi *et al.*, 2002). This activity causes elongation of cells known as the hummingbird phenotype (Segal *et al.*, 1999). CagA interacts with PAR1/MARK to disrupt tight junctions and induce loss of cell polarity (Saadat *et al.*, 2007). It has also been found to destabilise the tight junction complex E-cadherin/β-catenin and promote cell migration (Bagnoli *et al.*, 2005; Murata-Kamiya *et al.*, 2007). Furthermore, CagA increases IL-8 expression, enhancing the inflammatory response to infection (Crabtree *et al.*, 1994).

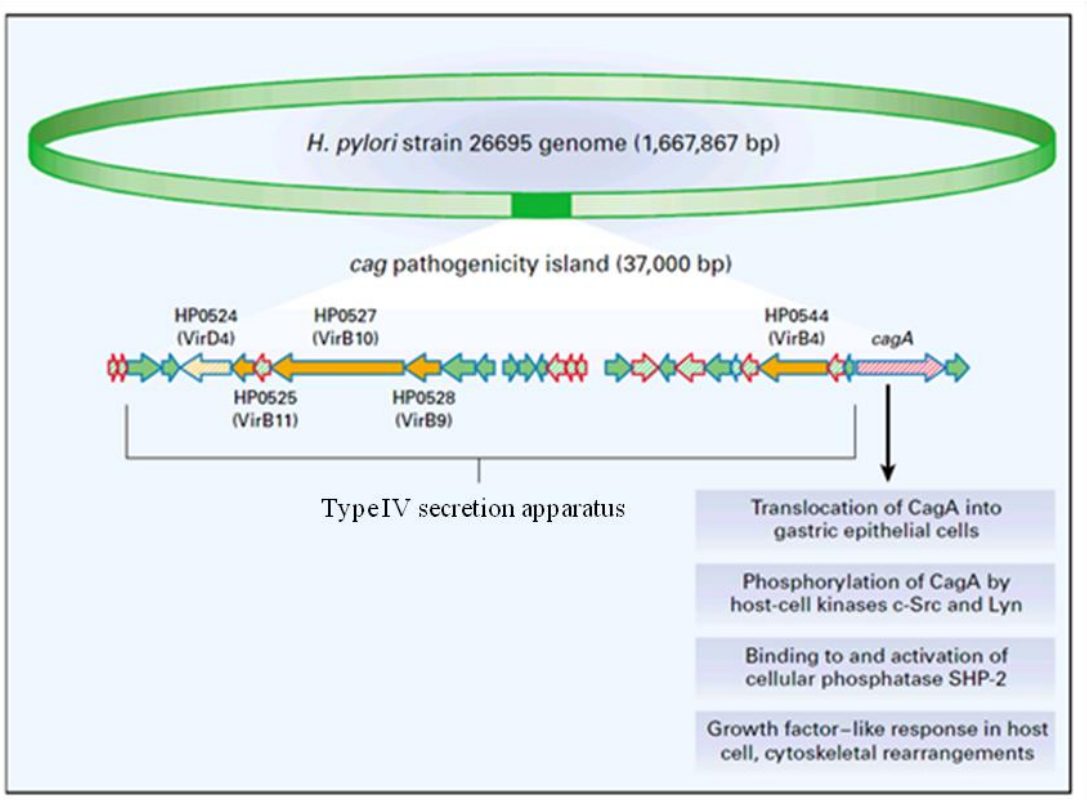


Figure 4 The *cag* pathogenicity island.

Overview of the components of the *cag* PAI in *H. pylori* adapted from (Suerbaum & Michetti, 2002).

Vacuolating cytotoxin (VacA) is a multifactorial secreted cytotoxin which has been extensively studied in *H. pylori*. VacA was identified soon after *H. pylori* was first isolated as a toxin which induces vacuolation of host epithelial cells (Leunk *et al.*, 1988). All strains possess the *vacA* gene but it is subject to extreme diversity. The preprotoxin is ~193 kDa and contains three variable regions which are the site of polymorphisms: signal sequence (s), mid (m) and intermediate (i) regions (Atherton *et al.*, 1995; Rhead *et al.*, 2007). The *vacA* allele harboured by a given strain is a determinant for pathogenicity. The s1/m1 genotype was considered the most virulent form but evidence suggests the i-region variant may be a better predictor of more severe clinical outcomes (Chung *et al.*, 2010; Miehlke *et al.*, 2000; Rhead *et al.*, 2007).

The pre-protoxin consists of an N-terminal signal sequence for transport across the inner membrane, a passenger domain and a C-terminal auto-transporter domain

which enables Type 5 secretion. The passenger domain is composed of p33 and p55 subunits which are cleaved to produce an 88 kDa toxin (Lupetti *et al.*, 1996; Telford *et al.*, 1994). p33 has a role in cell binding, while p55 is central to vacuolation, membrane depolarisation and internalisation (Torres *et al.*, 2005). The p33 subunit targets VacA to the mitochondrial inner membrane, resulting in apoptosis as indicated by the release of cytochrome c and activation of caspase 3 (Galmiche *et al.*, 2000; Willhite and Blanke, 2004). VacA forms endosomal membrane anion channels which inhibit antigen presentation, procathepsin D maturation and destruction of epidermal growth factor (EGF) (Molinari *et al.*, 1998; Satin *et al.*, 1997; Tombola *et al.*, 1999). This potent cytotoxin also inhibits T-cell proliferation and activation, enabling *H. pylori* to alter and evade the host adaptive immune response to infection (Gebert *et al.*, 2003; Sundrud *et al.*, 2004). Salama *et al.* showed that wild type *H. pylori* out-compete *vacA*-negative mutants in a murine gastric model, indicating an additional role in initial host colonisation for this toxin (Salama *et al.*, 2001).

LPS is a component of the outer membrane and the flagellar sheath (Jones and Curry, 1989). It is composed of lipid A, core oligosaccharide and an O-antigen domain. In *H. pylori*, LPS exerts a low level of pro-inflammatory activity compared with LPS of other enterobacteria, including *E. coli* and *Salmonella* (Muotiala *et al.*, 1992). Altered lipid A backbone phosphorylation allows LPS to evade TLR2 recognition, contributing to persistence of this pathogen (Cullen *et al.*, 2011). *H. pylori* exerts molecular mimicry through the O-chain of LPS which can present structures similar to host Lewis blood group antigens (Wirth *et al.*, 1997). Lewis X-induced autoantibodies are associated with colonisation ability and adhesion of *H. pylori* to AGS cells (Sheu *et al.*, 2007). LPS possibly exhibits its endotoxic activity through up-regulation of inducible nitric oxide expression which leads to impairment of host DNA repair machinery, as studied in human colon carcinoma cell lines (Cavallo *et al.*, 2011). A link between LPS production and flagellar assembly has been identified in *C. jejuni*. Cj0256 is a phosphoethanolamine transferase which post-translationally modifies both lipid A and flagellar rod protein, FlgG. Cj0256 is essential for motility in *C. jejuni*, as deletion of this gene results in aflagellate cells (Cullen and Trent, 2010).

H. pylori peptidoglycan is transported to host epithelial cells either by the *cag*PAI T4SS or bacterial outer membrane vesicles (Kaparakis *et al.*, 2010; Viala *et al.*, 2004). There, it binds to NOD1 receptors, triggering an NFκB/IL-8-mediated

inflammatory response (Girardin *et al.*, 2003; Viala *et al.*, 2004). Many *H. pylori* adhesins are surface proteins which bind host epithelial glycoproteins. Blood group antigen-binding protein (BabA) is an adhesin that binds Lewis-b blood group antigen (Borén *et al.*, 1993). Presence of the *babA2* gene has been suggested as a marker of increased risk to *H. pylori*-associated diseases including duodenal ulcer and adenocarcinoma (Gerhard *et al.*, 1999; Mizushima *et al.*, 2001). Sialic acid-binding adhesin (SabA) binds sialated Lewis-X (Mahdavi *et al.*, 2002). This adhesin facilitates binding of *H. pylori* to mucosal epithelial cells where it then illicits activation of neutrophils and inflammation (Petersson *et al.*, 2006; Unemo *et al.*, 2005). This strong response can be dampened by neutrophil activating protein (HP-NAP) to aid *H. pylori* persistence (Unemo *et al.*, 2005; Wang *et al.*, 2006). Outer inflammatory protein A (OipA) is an adhesin which illicits a pro-inflammatory host response through the activation of Il-8 (Yamaoka *et al.*, 2000). Additionally, OipA is involved in focal adhesion kinase (FAK) phosphorylation and activation, affecting actin stress fiber formation and cell motility (Tabassam *et al.*, 2008).

Catalase, superoxide dismutase and HP-NAP also function to neutralise reactive oxygen species, an oxidative stress response contributing to virulence (Bauerfeind *et al.*, 1997; Seyler *et al.*, 2001; Wang *et al.*, 2006). Duodenal ulcer promoting protein A (DupA) was identified in 2005 as an adhesin associated with IL-8 and IL-12 induction and inflammation in mononuclear cells (Hussein *et al.*, 2010; Lu *et al.*, 2005). *H. pylori* produces a pore-forming cytolysin orthologue, TlyA, which confers haemolytic activity (Lata *et al.*, 2014; Martino *et al.*, 2001). The outer membrane phospholipase A (PldA) has also been shown to mediate haemolysis and contribute to host colonisation (Dorrell *et al.*, 1999; Sitaraman *et al.*, 2012). Sphingomyelinase (SMase) activity results in haemolysis of blood lymphocytes as well as activation of mitogen-activated protein (MAP) kinases and apoptosis in AGS cells (Chan *et al.*, 2000; Tseng *et al.*, 2004). The capacity of *H. pylori* for horizontal gene transfer *in vitro* is indicative of its ability to adapt to and persist in its host (Blaser and Atherton, 2004; Nedenskov-Sorensen *et al.*, 1990). Toller *et al.* hypothesize that double-strand DNA breakages induced by *H. pylori* infection may be a contributing factor in the development of gastric carcinoma. Persistent infection may lead to mutations generated during DNA repair, thereby increasing the risk for cancer development (Toller *et al.*, 2011). Thus, *H. pylori* has a diverse armoury of virulence factors at its

disposal to aid in host infection and persistence while evading removal by the host immune system.

2.5 Disease

The association between *Helicobacter pylori* and chronic gastritis and peptic ulceration was noted by Marshall and Warren (Marshall and Warren, 1984; Warren and Marshall, 1983). In fact, Marshall ingested *H. pylori* to confirm Koch's postulates identifying *H. pylori* as a causative agent of gastritis (Marshall *et al.*, 1985). It is now accepted that almost all subjects infected with *H. pylori* develop chronic gastritis (Kusters *et al.*, 2006). However, only a minority develop the more severe pathological effects of infection, including peptic ulcer, non-ulcer dyspepsia, gastric carcinoma and MALT lymphoma (Figure 5). An association has also been made between *H. pylori* infection and sudden infant death syndrome (Kerr, 2000).

In most cases, acute gastritis does not progress further, but it can develop into gastric/duodenal ulcer disease (Cave and Goddard, 1999). *H. pylori* infection is also associated with dyspepsia (Harvey *et al.*, 2010). Gastric ulcers occur in the region of the stomach where the corpus mucosa meets the antrum mucosa, whereas duodenal ulcers are associated with the duodenal bulb (Kusters *et al.*, 2006). *H. pylori* is a major causative agent of gastric ulcers. Complications can arise from ulceration including bleeding and stricture formation.

The more severe disease types associated with *H. pylori* infection include gastric adenocarcinoma and MALT lymphoma. In 1994, *Helicobacter pylori* was classified as a human carcinogen (International Agency for Research on Cancer, 1994). Atrophic gastritis and intestinal metaplasia can result from chronic gastritis-associated inflammation (Kuipers *et al.*, 1995). These features can increase the risk of development of gastric cancer (Uemura *et al.*, 2001). Gastric cancer is the third most common form of cancer in men, and fifth in women (Jemal *et al.*, 2011; Society, 2011). Approximately 50% of gastric cancer occurrences can be attributed to *H. pylori* infection, illustrating the global burden of *H. pylori* infection (Parkin, 2006). In a Swedish cohort, *H. pylori* was associated with ~70% of all noncardia adenocarcinomas (Ekström *et al.*, 2001). Clinical studies show that a high percentage of MALT lymphoma patients test positive for *H. pylori* and its eradication has been linked with disease remission (Eidt *et al.*, 1994; Fischbach *et al.*, 2009).

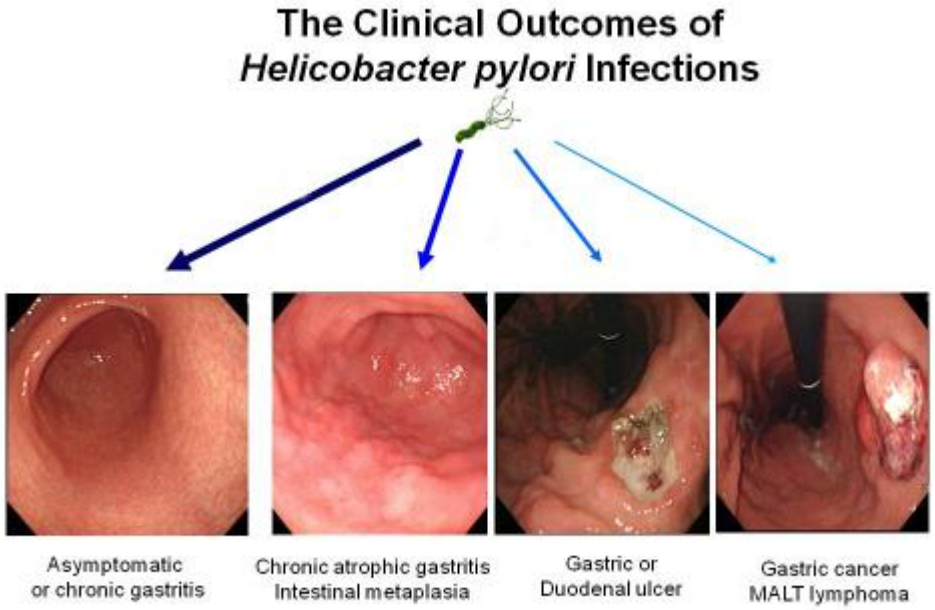


Figure 5 Clinical outcomes of *H. pylori* infection.

Adapted from (Kim *et al.*, 2011).

2.6 Diagnostics

A broad range of diagnostic tests have been established for *H. pylori* infection. These can be classified as either invasive (endoscopy, histology) or non-invasive (serological tests, urine, stool antigen assays, breath *etc.*). Results of two tests are sometimes required to confirm diagnosis- urea breath testing and histological examination of tissue are most commonly used (Calvet *et al.*, 2009; Graham *et al.*, 2008).

Culture of *H. pylori* is not only useful as a confirmatory test, but also allows determination of antibiotic susceptibilities of the strain(s) isolated to aid in treatment strategy selection. Urease-based biopsy tests available range from the CLOtest, which takes up to 24 hr, to the PyloriTek strip which takes 1 hr to give a result (Cutler *et al.*, 1995; Laine *et al.*, 1996). However, non-invasive alternative urease breath tests are now commonly used in which labelled urea is ingested and the resulting bicarbonate, exhaled as CO_2, is detected (Koletzko *et al.*, 1995; Minoli *et al.*, 1998). The rapid urease test has > 98% sensitivity, while $^{13}C/C^{14}$ urea breath test identifies the presence of *H. pylori* with sensitivity > 95% (WGO, 2010). PCR offers

specificity, high sensitivity and may be used to detect *H. pylori* non-invasively from bodily fluids *e.g.* saliva; however, PCR has yet to be optimised and standardised for this purpose. PCR-based detection of *vacA* intermediate region has been developed which could be used to rapidly screen patients at elevated risk for gastric cancer development (Ferreira *et al.*, 2012). Quantitative real-time (qRT) PCR and fluorescence qPCR are recently emerging detection methods which may provide an alternative to the traditional detection methods (Ou *et al.*, 2013; Saez *et al.*, 2012). Recent comparative studies suggest histology may be the most robust detection method (Choi *et al.*, 2012; Tian *et al.*, 2012). A recent study involving enzyme-linked immunosorbent assay (ELISA) of stool samples of children highlights another method warranting development as a non-invasive diagnostic of *H. pylori* infection (Leal *et al.*, 2011).

2.7 Treatment

Consensus guidelines recommend treatment to eradicate *H. pylori* in patients presenting symptoms of infection (Malfertheiner *et al.*, 2012). Growing *H. pylori* antibiotic resistance presents a challenge for treatment of infection due to capacity for mutation and gene acquisition facilitated by the extreme genome plasticity of *H. pylori* (outlined above) and the common use of antibiotic therapies to treat bacterial infections (Mégraud, 2004). Failed dual therapy results in the development of *H. pylori* dual resistance, indicating that emerging resistance is associated with previous antibiotic treatments (Heep *et al.*, 2000). Clarythromycin-based triple therapy using two antibiotics and a proton pump inhibitor is the standard first line treatment for infection (Bazzoli *et al.*, 1993). However, growing resistance to clarithromycin, and levofloxacin, means this first line of treatment must be reviewed (Mégraud, 2012).

Quadruple bismuth-based therapies is one approach to overcome the issue of resistance (Malfertheiner *et al.*, 2012). Tailored treatment is an attractive solution, made more feasible with the development of molecular-based susceptibility screening such as GenoType HelicoDR (Cambau *et al.*, 2009). An effective vaccine in humans has yet to be described that affords sustainable protection from *H. pylori* infection (Koch *et al.*, 2013). Furthermore, due to the difficulties in identifying the point of infection, when to vaccinate could be vital to the success of vaccination schemes. OipA, LPS and urease A immunogens have had some success in protection

from *H. pylori* using murine models (Altman *et al.*, 2012; Chen *et al.*, 2012; Guo *et al.*, 2012).

3 Bacterial Secretion Systems

The secretion of bacterial proteins is of particular importance for understanding pathogenic bacteria. A complete secretion system enables the delivery of toxins, enzymes and other virulence factors during infection, while adhesins can be presented on cell surfaces to mediate interactions. Non-pathogenic bacteria also utilise protein secretion to enhance survival *e.g.* the secretion of sortase by *Lactobacillus salivarius* subspecies *salivarius* strain UCC118 (van Pijkeren *et al.*, 2006). The mechanisms for secretion differ between Gram positive and Gram negative organisms. Each of the currently known secretion systems (I - VII) are described below.

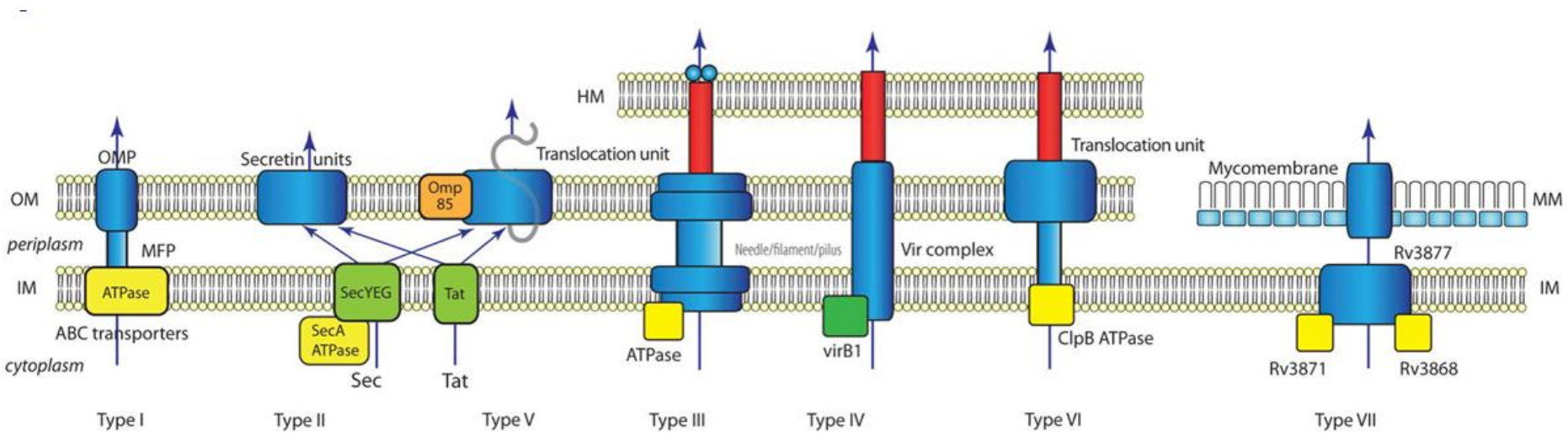


Figure 6 Generalised diagram of the known bacterial secretion systems.

Graphical representation of bacterial protein secretion systems spanning the inner membrane (IM), outer membrane (OM), host membrane (HM) and mycomembrane (MM). OMP: outer membrane protein; MFP: membrane fusion protein (Tseng *et al.*, 2009)

3.1 Type I Secretion System

The Type I secretion system (T1SS) is found in both Gram positive and Gram negative bacteria. It facilitates the secretion of a wide variety of proteins including toxins, proteases, lipases and bacteriocins. Bacteria use ABC transporters as a mechanism for multidrug resistance, which can account for > 10% of the transporters encoded for on the genome (Paulsen, 2003). The T1SS is composed of an inner membrane ATPase-binding cassette (ABC) transporter protein, a membrane fusion protein and an outer membrane pore-forming protein (Omori and Idei, 2003). The mechanism of secretion by this pathway is Sec-independent (Figure 6). Proteins exported in this way have a characteristic carboxy-terminal signal sequence, which is not cleaved during secretion (Duong *et al.*, 1996). Secretion of the *E. coli* haemolysin, Hly, is the model system for the T1SS (Mackman *et al.*, 1986). *Lactococcus lactis* is a Gram positive organism with 40 putative multidrug resistance proteins secreted by the T1SS including LmrCD (Lubelski *et al.*, 2007). *H. pylori* strains possess a number of ABC transporters *e.g.* NixA (Hendricks and Mobley, 1997).

3.2 Type II Secretion System

The Type II secretion system (T2SS) is a more complex two-step process used by Gram negative bacteria to secrete a variety of proteins, mainly enzymes. This system is encoded by 12 - 16 genes, generally found together in one operon, and has been studied in detail in *E. coli*, *Pseudomonas aeruginosa* and *Legionella pneumophila* (Douzi *et al.*, 2012; Jyot *et al.*, 2011; Rossier *et al.*, 2004). The first step in this process is transport across the inner membrane into the periplasm either by the Sec or twin-arginine translocation (Tat) pathways (Pugsley, 1993; Voulhoux *et al.*, 2001).

The Sec secretion system is a general export system used by Gram positive bacteria to transport proteins across its single membrane to the extracellular milieu. In Gram negative bacteria, Sec secretion translocates proteins to the periplasm in three stages: protein sorting, translocation and release/maturation (Papanikou *et al.*, 2007). Sec secretion requires the recognition of an amino-terminal leader sequence on proteins targeted for secretion. These pre-proteins are brought to a channel-forming translocase complex at the inner membrane composed of SecY, SecE and SecG which drives protein export through activity of the SecA ATPase (van den

Berg *et al.*, 2004). Once in the periplasm, the signal sequence is cleaved by a signal peptidase and protein folding is initiated to produce the mature form of the protein (Mogensen and Otzen, 2005; Paetzel *et al.*, 2002).

The Tat pathway facilitates the transfer of folded proteins across the inner membrane and has a role in diverse cellular functions including cell division, quorum sensing and cell motility (Ding and Christie, 2003; Palmer and Berks, 2012; Stanley *et al.*, 2001; Stevenson *et al.*, 2007). Proteins are targeted to the Tat system for export through recognition of an amino-terminal signal sequence containing a twin-arginine motif (Chaddock *et al.*, 1995). The export apparatus located in the bacterial inner membrane consists of TatA and TatC protein families (Palmer and Berks, 2012). The TatABC complex binds target proteins and export is driven by proton motive force (Bageshwar and Musser, 2007; Mould and Robinson, 1991). The crystal structure of TatC, the core component of the pathway, from *Aquifex aeolicus* has recently been published (Rollauer *et al.*, 2012). TatC recognises the signal sequence of target proteins and recruits other Tat export proteins to initiate export. The TatBC complex of *E. coli* has been modelled from transmission electron microscopy (TEM) data (Tarry *et al.*, 2009). TatA forms a homopolymeric pore in the inner membrane which has a variable diameter thereby allowing export of proteins of varying sizes (Gohlke *et al.*, 2005).

The T2SS subsequently translocates target exoproteins across the outer membrane in a process involving an outer-membrane complex and a pesudopilus structure (Korotkov *et al.*, 2012). The pseudopilus is composed of a multimer of the major pseudopilin, secretin and four other pseudopilins. The assembled structure resembles the Type IV secretion system (Figure 6) (Durand *et al.*, 2003; Sauvonnet *et al.*, 2000). There is also some similarity with components of the archaeal flagellum, *e.g.* the pre-flagellin peptidase, FlaK in *Methanococcus* shares homology with prepilin peptidases (Peabody, 2003). The current hypothesis for the mechanism of secretion is that binding of the target protein to the periplasmic domain of secretins triggers the ATPase-driven extension of the pesudopilus which pushes the exoprotein through the channel (Korotkov *et al.*, 2012).

3.3 Type III Secretion System

The Type III secretion system (T3SS) is one of the most complex secretion systems known, composed of > 20 proteins (Cornelis, 2006). In Gram negative

bacteria, the T3SS translocates target proteins in a one-step process which is independent of the Sec pathway. The T3SS mediates bacterial/host interactions through injection of effector molecules into host cells and plays a role in colonisation and pathogenesis (Galán and Collmer, 1999; Rosqvist *et al.*, 1994). It produces an injectisome, so-called due to its needle-like structure (Figure 6). There are seven families of injectisomes, which have evolved independently of the bacteria in which they are found *i.e.* there is evidence of horizontal gene transfer events in their evolution (Troisfontaines and Cornelis, 2005). Injectisomes are produced by *Yersinia*, *Pseudomonas*, *Shigella*, and *E. coli* (Cornelis, 2006). The T3SS allows the translocation of a wide variety of effector molecules and is largely considered to be a virulence factor, though it has been found in some non-pathogenic bacteria. There is evidence that the injectisome itself can cause host cell damage and death, without the translocation of effector proteins (Hauser, 2009).

The T3SS generally requires host-cell contact to activate protein export which is guided by chaperones (Pettersson *et al.*, 1996). It is composed of a multi-ring basal structure which spans the bacterial membranes, and a protruding needle-like filament which delivers effector proteins. A membrane-associated ATPase in the cytoplasm is essential for protein export and it has homology to F_1-ATPase subunits (Woestyn *et al.*, 1994; Zarivach *et al.*, 2007). At the distal end of the needle, there is a tip complex which acts as a platform for translocators which induce pore formation in the host cell (Moraes *et al.*, 2008; Mueller *et al.*, 2008). Effector proteins are then translocated in an unfolded state to the host through the hollow lumen of the injectisome (Akeda and Galán, 2005).

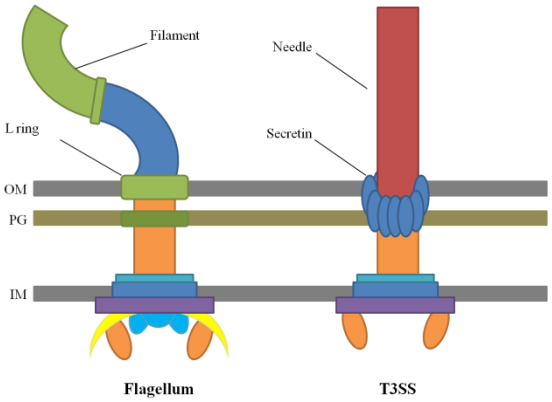


Figure 7 The bacterial flagellum and the Gram negative injectisome.

Schematic comparison of the components of these Type III secretion apparatus. Adapted from (Saier, 2004).

The T3SS shares close similarity at the level of individual components, and how they are assembled, to the bacterial flagellar export apparatus (Figure 7). Phylogenetic analyses had indicated that these systems evolved independently of each other from a common ancestor (Gophna *et al.*, 2003). However, a recent re-evaluation of the evolutionary relationship between these two systems incorporating the large body of genome sequencing information now available, seemed to indicate that the T3SS evolved from the bacterial flagellum (Abby and Rocha, 2012). While the primary function of the flagellar export apparatus is the sequential export of components of the flagellar apparatus (Section 4), there are documented cases where it is also used for the export of non-flagellar proteins *e.g.* YplA of *Yersinia enterocolitica* and CiaC of *Campylobacter jejuni* (Christensen *et al.,* 2009; Young *et al.*, 1999). Similar to flagellar biogenesis, construction of the T3SS is a hierarchical process which is tightly regulated. A number of components display sequence similarity with their flagellar counterparts. For instance, Spa32, which controls needle length of injectisomes produced in *Shigella flexneri* is homologous with FliK, the flagellar hook-length control protein (Magdalena *et al.*, 2002).

3.4 Type IV Secretion System

The Type IV secretion system (T4SS) is found in both Gram positive and Gram negative bacteria. It can be sub-divided into three categories based on function: bacterial conjugation, DNA transfer and effector protein translocation. The T4SS is unique in its ability to transfer both DNA and proteins to bacterial, plant and animal cells *i.e.* interkingdom transfer (Fronzes *et al.*, 2009). *Agrobacterium tumefaciens* C8 is the model system of T4S. The VirB/D system is composed of 12 proteins which assemble to produce a cytoplasmic/inner membrane complex, a channel which spans bacterial and host cell membranes and an external pilus structure (Figure 6). Using this system, *A. tumefaciens* delivers oncogenic DNA-protein complexes to host cells which can lead to the development disease in plants (Fronzes *et al.*, 2009). Conjugation is a DNA transfer mechanism which requires cell-cell contact. Conjugative plasmids generally contain genes with functions in areas such as antibiotic resistance and stress-response, which promote environmental adaptation and genetic diversity (Wallden *et al.*, 2010). T4SS produce pili which enable a "slingshot"-type crawling motility as seen in *Pseudomonas aeruginosa* (Jin *et al.*,

2011). *Bordetella pertussis* and *Legionella pneumophila* are examples of human pathogens which effectively use the T4SS to illicit disease (Wallden *et al.*, 2010).

In *H. pylori*, the CagA virulence protein is exported by the T4SS, the machinery for which is encoded on a pathogenicity island (Section 2.4). In addition to conjugative transfer of genetic material, *H. pylori* encodes genes for the ComB system which facilitates the transfer of DNA independently of conjugation (Fischer *et al.*, 2010; Hofreuter *et al.*, 2001). Each of these mechanisms contributes to the characteristic extreme genetic plasticity of *H. pylori*.

3.5 Type V Secretion System

The Type V secretion system (T5SS) is a two-step process describing three distinct mechanisms of protein export: the autotransporter (Va) system, the two-partner secretion (Vb) pathway and the oligomeric coiled-coil adhesin (Vc) system (Henderson *et al.*, 2004). The T5SS is the simplest of the secretion systems and is the most commonly employed system for protein secretion by Gram negative bacteria (Figure 6). Effector proteins secreted by the Va system are composed of three structural domains, the first of which is an N-terminal signal sequence recognised by the Sec system for transport across the inner membrane. The exoprotein also contains a central passenger domain and a C-terminal translocation β-barrel domain which forms a pore that enables the protein to cross the outer membrane and be secreted into the extracellular milieu (Pohlner *et al.*, 1987). After secretion, the protein undergoes auto-proteolysis to cleave the helper domain *e.g.* VacA in *H. pylori*. However, there are some exceptions as not all effector proteins have autocatalytic protease activity *e.g.* the Hsr surface protein ring of *H. mustelae* remains tethered to the outer membrane (O'Toole *et al.*, 1994). Autotransporters are widely found in ε-proteobacteria and are associated with virulence as effector proteins often include adhesins, enzymes and toxins (Henderson *et al.*, 2004).

The two-partner secretion pathway is similar to the Va system. An N-terminal signal sequence on the passenger protein facilitates transport to the periplasm. The key difference is that the translocation unit is translated as a separate protein in this system (Jacob-Dubuisson *et al.*, 2001). The Vc system describes adhesins which are composed of 6 distinct domains, the archetype of which is YadA in *Yersinia* (Hoiczyk *et al.*, 2000). An N-terminal signal sequence is followed by domains designated as: head-D, neck-D, stalk-D, linking-R and C-terminal β-barrel domains.

The lollipop-shaped effector proteins are exposed on the surface of the outer membrane where they are anchored by the C-terminal domain.

Relatively recently, Salacha *et al.* described a novel T5SS in *P. aeruginosa* which they named the Vd system. This is a hybrid of the autotransporter and two-partner systems in which the C-terminus translocation domain of the protein for export, PlpD, more closely resembles a translocation unit of the Vb system (Salacha *et al.*, 2010). An inverse mechanism describes the most recently identified type Ve system. This family shares closest similarity to the Va autotransporters, however, the β-barrel translocation domain is at the N-terminus, while the C-terminus is exposed on the outer membrane surface (Oberhettinger *et al.*, 2012). Intimin of enteropathogenic *E. coli* and invasin of *Yersinia*, both virulence factors, were the first described members of this group.

3.6 Type VI Secretion System

The Type VI secretion system (T6SS) was first described in *Vibrio cholera* (Pukatzki *et al.*, 2006). It has since been identified in the genomes of more than 80 of Gram negative bacteria, including both pathogenic and non-pathogenic species such as *Yersinia pestis* and *Burkholderia pseudomallei* (Boyer *et al.*, 2009). *Vibrio cholera* uses the T6SS to export toxins not only to eukaryotic host cells, but also to other bacteria, providing a competitive advantage in its environment (MacIntyre *et al.*, 2010).

The 15 - 25 genes in the T6SS locus encode effector proteins, structural components, chaperones and ATPases to power secretion by this Sec-independent system (Pukatzki *et al.*, 2009). While many structural proteins have been studied, little is known about the other components. Haemolysin A coregulated protein (Hcp) is secreted by all functional T6SS to form homohexamer rings. As these rings can be stacked to produce a nanotube structure, it is possible that Hcp is the building block of a core channel through which effector proteins can be transported (Ballister *et al.*, 2008). Interestingly, Hcp requires the secretion of VgrG proteins in *Vibrio cholera* which interact to form a complex; these have sequence similarity with bacteriophage T4 tailspike proteins which puncture host cells and hence may have a key role T6SS function (Pukatzki *et al.*, 2007). Threonine phosphorylation at a post-translational level has been identified as playing a role in T6SS regulation (Mougous *et al.*, 2007). More work in this area is needed to elucidate the regulation of this secretion system.

3.7 Type VII Secretion System

The Type VII secretion system (T7SS) is a specialised system used by *Mycobacterium tuberculosis*, the causative agent of tuberculosis in humans (Figure 6). This system is adapted to the unusual cell envelope of this Gram positive pathogen (Brennan and Nikaido, 1995). The T7SS includes five types, ESX 1 - 5, the first of which is the archetype for the system (Stanley *et al.*, 2003). Although the exact nature of the role the T7SS plays in *M. tuberculosis* virulence is unknown, disruption of this pathway has been shown to attenuate virulence (Abdallah *et al.*, 2007; Stanley *et al.*, 2003). Effector proteins secreted by this system possess a recently identified C-terminal signal sequence and include T cell antigens (Daleke *et al.*, 2012). Components of this system include chaperones, membrane-spanning proteins and ATPases, though structural information is lacking (Abdallah *et al.*, 2007). Proteins for secretion by the ESX 1 system form a heterodimeric complex which is targeted for secretion (Renshaw *et al.*, 2005). T7SS homologues have been identified in other Gram positive bacteria including *S. aureus* and *Bacillus* spp..

4 Composition and Organization of the Bacterial Flagellum

The bacterial flagellum is an ancient and complex nanomachine which facilitates motility. It is an important feature of *H. pylori* because motility is an essential colonization factor (Eaton *et al.*, 1992). Phylogenetic analysis suggests that the flagellum evolved from a single gene that was duplicated and underwent mutations, leading to new functions (Liu and Ochman, 2007). The best studied models for bacterial flagellum biogenesis are those in *E. coli* and *Salmonella enterica*. The *H. pylori* flagellum largely resembles these models, with some differences, the details of which will be discussed in this section.

4.1 Morphology of the flagellum in *Enterobacteriaceae*

The flagellar superstructure is composed of four sub-sections: basal body, export apparatus, hook and filament. The basal body is composed of three rings: an inner membrane (MS) ring, a periplasmic (P) ring and an outer membrane (L) ring (Figure 8). These are connected by the cylindrical rod and the structure serves to anchor the flagellum in the bacterial cell membrane (Macnab, 2004). FliF, FlgI and FlgH compose the MS, P and L rings, respectively. These proteins, involved in early stages of flagellum biogenesis, are assembled using the Sec secretion system (Jones *et al.*, 1989). The rod is composed of a number of proteins: an MS-ring rod junction protein (FliE), transmission shaft proteins (FlgB/C/F/G) and a rod capping protein (FlgJ) (Homma *et al.*, 1990). The rod proteins are exported by the flagellar T3SS.

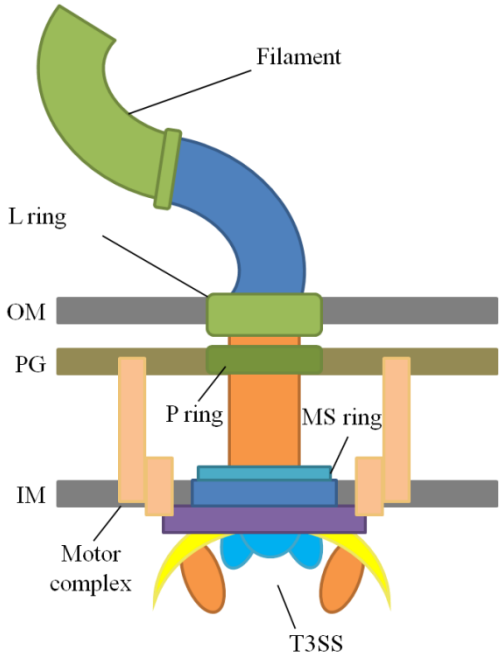


Figure 8 Generalised diagram of the bacterial flagellum.

The basal body and export apparatus are contained in the bacterial membrane and cytoplasm, while the rod, hook and filament form a hollow tube through which flagellar components are exported (Pallen *et al.*, 2006).

The flagellar motor, composed of stator and rotor elements, interacts with the basal body to generate torque which results in propulsion (Lloyd *et al.*, 1996). MotA and B proteins of the stator assemble as integral membrane studs in the peptidoglycan layer, while the rotor (FliG multimer) is non-covalently attached to the cytoplasmic side of the MS ring through interactions with FliF (Braun *et al.*, 1999; Francis *et al.*, 1992). Directional movement is achieved by flagellar rotation that is either clockwise (tumbling) or anticlockwise (swimming). The switch complex which controls rotation is composed of FliM, FliN and FliG proteins which assemble a C ring complex around the MS ring (Francis *et al.*, 1994; Yamaguchi *et al.*, 1986). MotA and MotB studs in the periplasm interact with the rotor and C ring (Braun *et al.*, 1999; Thomas *et al.*, 1999). Thus, the bacterial flagellum contains a motor with the capacity for controlled rotation towards stimuli/away from repellents, mediated by chemotactic response regulators *e.g.* CheY (Foynes *et al.*, 2000).

The hook is known as a universal joint as it links the rod to the filament (propeller component of the flagellum). It is a flexible helical assembly of ~120 FlgE monomers which allows multiple polar flagella rotate as a coordinated bundle

(Macnab, 1977; Makishima *et al.*, 2001). Assembly of the short, curved hook requires a capping protein (FlgD) to guide assembly (Ohnishi *et al.*, 1994). Hook associated proteins, FlgK (HAP1) and FlgL (HAP3), assemble at the junction between the hook and filament where they act as structural adapters (Hirano *et al.*, 1994). FliK tightly controls the invariant length of the hook, which is 55 nm in *Salmonella* (Hirano *et al.*, 1994).

The filament is a long, thin helical structure. In *Salmonella*, the filament is composed entirely of FliC which assembles as 11 protofilaments which can be modified through supercoiling when alternating between swimming and tumbling modes of motility (Samatey *et al.*, 2001). The filament of *H. pylori* is composed of two subunits: a major flagellin, FlaA, and minor flagellin, FlaB (Leying *et al.*, 1992; Suerbaum *et al.*, 1993). Assembly of the filament is guided by the filament capping protein, FliD (HAP2) which is essential for motility (Ikeda *et al.*, 1987; Kim *et al.*, 1999). These flagellins have only 58% sequence identity and their expression is alternately regulated (see below). FlaB incorporates into the filament in a hook-proximal position (Kostrzynska *et al.*, 1991). While *flaA*-null mutants are completely non-motile, *flaB* mutants retain the motility phenotype (Suerbaum *et al.*, 1993).

Structural proteins of the flagellum are translocated by the export apparatus through a narrow central channel using the flagellar Type III secretion system (T3SS) and assemble to extend the growing flagellum. The export apparatus is composed of soluble proteins (FliH, FliI, FliJ), located in the cytoplasm, and integral membrane proteins (FlhA, FlhB, FliO, FliP, FliQ and FliR) likely located in the MS ring (Figure 9) (Minamino and Macnab, 1999). Localisation of the export apparatus at the MS ring is mediated, in part, by the interaction of FlhA with FliF (Kihara *et al.*, 2001). The localisation of FliP and FliR to the basal body supports the hypothesis that the export apparatus is found in the central pore of the MS ring (Fan *et al.*, 1997). Interactions between export apparatus chaperones and the proteins of the C ring indicate a role for the C ring in docking (González-Pedrajo *et al.*, 2006). The membrane-bound components of the export apparatus form a proton-driven export gate where proteins are unfolded and translocated across the membrane (Minamino and Namba, 2008; Minamino *et al.*, 2009). The soluble components of the export apparatus function to bind and deliver proteins to the export gate for efficient flagellum assembly (Minamino and Namba, 2008).

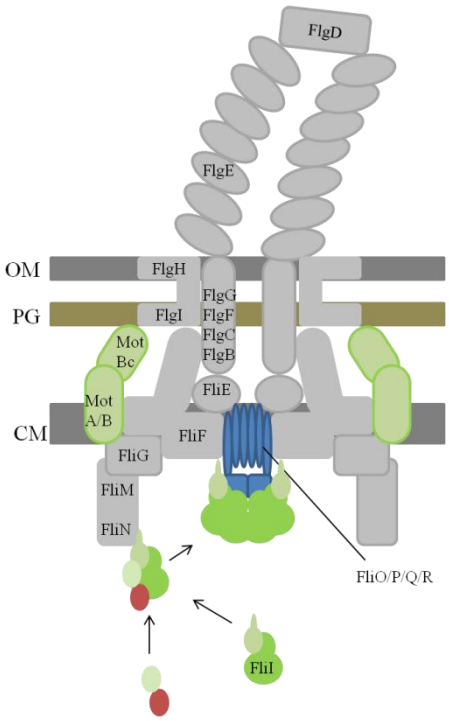


Figure 9 Schematic overview of the bacterial flagellar export apparatus.

FliH, FliI and FliJ are soluble components of the export apparatus (coloured green). FlhA, FlhB, FliO, FliP, FliQ and FliR are integral membrane components (coloured blue). Proteins coloured red represent substrates for export coupled to a chaperone. Adapted from (Minamino *et al.*, 2008).

In *H. pylori*, FliI is an essential protein in flagellum construction while a *fliI*-null mutant in *Salmonella* exhibits only reduced motility (Jenks *et al.*, 1997; Minamino and Namba, 2008). FliH regulates FliI activity to prevent hydrolysis of ATP when FliI is not involved in protein export (González-Pedrajo *et al.*, 2002; Minamino and MacNab, 2000; Minamino *et al.*, 2001). When in complex, FliH binds the N-terminus of FliI and inhibits the ATPase activity but also promotes docking (Lane *et al.*, 2006; Minamino and MacNab, 2000). The stable FliH-FliI heterodimer docks at the export apparatus through interactions with the FlhA-FlhB complex (Minamino *et al.*, 2003). The FliH-FliI complex also binds the C ring through FliH N-terminal interactions with FliN (McMurry *et al.*, 2006).

FliI is homologous to the α and β subunits of F_0F_1-ATPase and functions in protein docking at the export gate (Fan and Macnab, 1996). FliI assembles as a hexameric pore at the export apparatus where protein unfolding and export follow (Claret *et al.*, 2008; Kazetani *et al.*, 2009). FliJ is an essential component of the

export apparatus which is involved in chaperone recycling (Evans *et al.*, 2006). In *Salmonella*, it promotes the ATPase activity of FliI and interacts with FlhA to facilitate docking of the FliH-FliI complex (Ibuki *et al.*, 2013; Minamino *et al.*, 2010).

As flagellum biosynthesis is a hierarchical process, control of the substrate-specificity switch is critical to prevent premature export of filament components before completion of the hook. FlhB is an integral membrane protein which interacts with FliK to control the switch from export of early (rod/hook) to late (major flagellin) flagellar proteins (Williams *et al.*, 1996). FlhB is located at the cytoplasmic face of the export apparatus and contains a number of transmembrane helices (Zhu *et al.*, 2002). The substrate-specificity of FlhB depends on the conformational state of the protein, which is mediated by autolytic cleavage of its carboxy terminus (Ferris *et al.*, 2005). FliK is the hook length control protein, which triggers the cleavage of FlhB once the hook has reached its full length (Erhardt *et al.*, 2010; Moriya *et al.*, 2006). The mechanism by which FliK determines the hook length is termed the molecular ruler theory (Erhardt *et al.*, 2010). FliK is secreted intermittently through the growing flagellum during assembly of the hook-basal body complex. Interaction of the FliK N-terminus with hook subunits and the hook cap causes a pause in secretion when the FliK C-terminus can interact with $FlhB_c$ once the hook is long enough.

4.2 Flagellum Assembly in *Enterobacteriaceae*

Flagellum assembly occurs as a sequential, tightly regulated process whereby the cell proximal components *i.e.* basal body, are assembled first and followed by the more distal components in a sequential manner (Figure 10). The first component of the flagellum to assemble is the MS ring subunit, FliF, mediated by the Sec secretion system (Ueno *et al.*, 1994). The export apparatus and substrate switch complex assemble around the MS ring in an independent process which does not require other flagellar proteins (Kubori *et al.*, 1997). The C-terminal peptidoglycan-binding motif of MotB dimers may be responsible for targeting of the flagellar rotor to the stator (Kojima *et al.*, 2008).

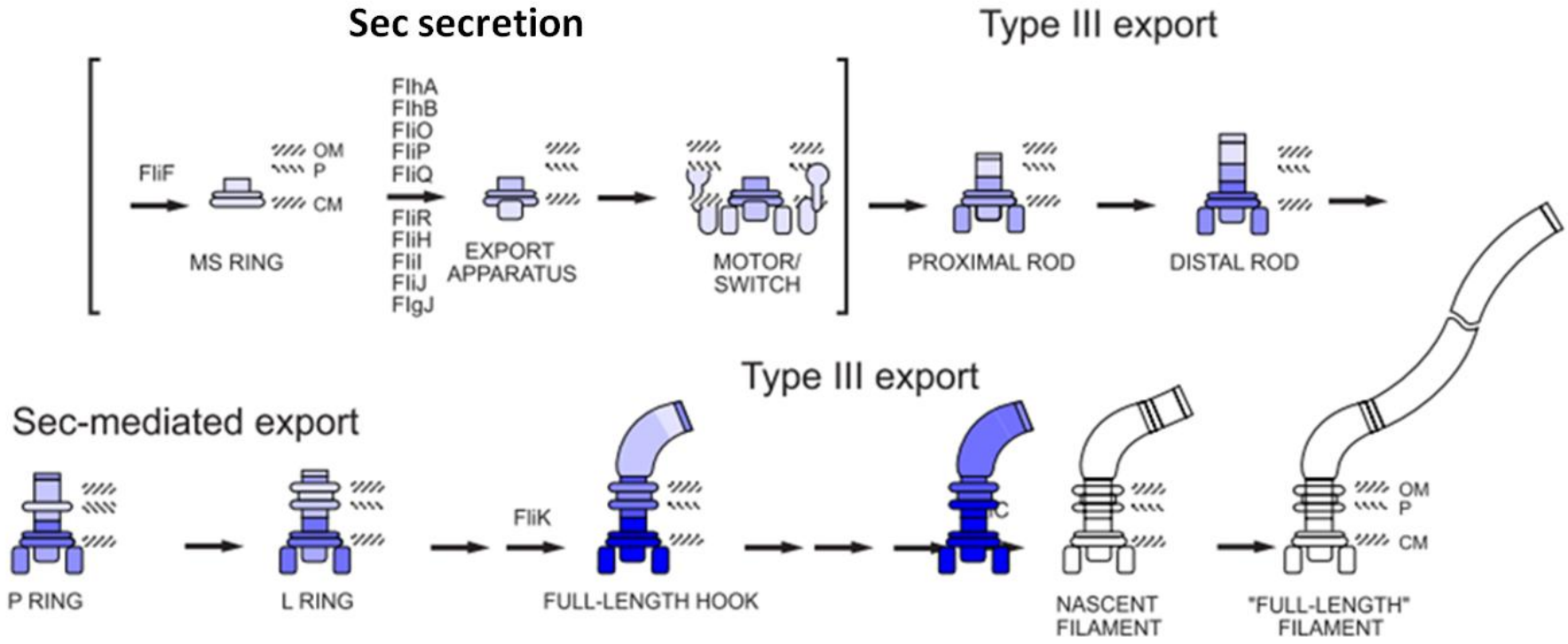


Figure 10 Ordered flagellum synthesis in *Salmonella.*

Sequential assembly of the bacterial flagellum where bracketed steps denote proteins expressed before a functional export apparatus is assembled *i.e.* proteins exported by the Sec secretion system. Assembly of the periplasmic and outer membrane rings requires Sec-mediated protein export. Hook and filament proteins are secreted by the flagellar Type III export apparatus. Adapted from (Macnab, 2004).

Assembly of the proximal rod components *i.e.* MS ring rod/junction protein, FlgB/C/F is guided by the rod capping protein, FlgJ (Hirano *et al.*, 2001). Distal rod formation is followed by Sec-mediated assembly of the P- and L-rings (Figure 10). It has been proposed that the hook capping protein (FlgD) attaches to the distal growing end of the hook by its N-terminus where it promotes FlgE polymerisation, while the C-terminus functions to block and prevent secretion of unincorporated monomers into the extracellular milieu (Moriya *et al.*, 2011). As mentioned above, FliK acts as a molecular ruler to control flagellar hook length.

In *Salmonella*, FlgM is secreted into the extracellular milieu upon completion of the hook, releasing FliA suppression (Gillen and Hughes, 1991; Hughes *et al.*, 1993). In *H. pylori*, FlgM is not secreted after the substrate-specificity switch. It is hypothesised that FlgM may instead shuttle between FliA and $FlhA_c$ in the cytoplasm (Rust *et al.*, 2009). This initiates expression of the late flagellar genes which then assemble to form the junction, filament and capping proteins. FliD promotes growth of the filament at the distal end and is retained in the final structure (Yonekura, 2000). The filament can contain up to 20,000 subunits (Macnab, 2003). The entropy of polymerisation of flagellin subunits at the tip of the growing filament is sufficient to recruit another subunit from the export apparatus in a chain mechanism of filament extension (Evans *et al.*, 2013).

4.3 Regulation of Flagellar Assembly

Since over 40 proteins are involved in flagellum assembly, the process must be tightly regulated to avoid improper assembly and/or energy wastage (Chevance and Hughes, 2008). Hierarchical flagellar assembly of *Enterobacteriaceae* coupled with sequential transcriptional activation of flagellar genes maintains tight control of this complex process (Chilcott and Hughes, 2000). Flagellar genes can be divided into three classes which represent the early, middle and late genes. Three RNA polymerase sigma factors control the gene expression of these classes (McCarter, 2006).

In *Salmonella*, Class I genes are located in the *flhDC* operon. The gene products of this operon form a complex, $FlhD_4C_2$. Activation/inactivation of $FlhD_4C_2$ is dependent upon an array of environmental stimuli including osmolarity and catabolic repression, as well as bacterial growth phase (Prüss and Matsumura, 1997; Shin and Park, 1995; Soutourina *et al.*, 1999). In *E. coli*, $FlhD_4C_2$ is post-transcriptionally

regulated by the global regulator, CsrA (Wei *et al.*, 2001). $FlhD_4C_2$ also plays a role in other cell processes which are not related to bacterial motility (Stafford *et al.*, 2005).

The primary function of $FlhD_4C_2$ lies in its role as the "master regulator" of flagellar gene expression. $FlhD_4C_2$ mediates σ^{70} RNA polymerase transcriptional activation of the Class II (middle) flagellar genes which include components of the basal body and hook (Liu *et al.*, 1995). Included in the genes transcribed by σ^{70} is another RNA polymerase sigma factor, σ^{28} (Kutsukake *et al.*, 1990). The alternative sigma factor (σ^{28}) in turn controls the expression of late flagellar genes *i.e.* the flagellar motor and filament subunits. FlgM is a negative regulator of σ^{28} and is secreted upon completion of the hook (Gillen and Hughes, 1991; Ohnishi *et al.*, 1992). Secretion of FlgM, and hence release of σ^{28}, triggers transcription of the late flagellar genes (Kutsukake, 1994).

4.4 Flagellar chaperones in *Enterobacteriaceae*

There are a number of cytoplasmic chaperones which play an important role in flagellum biosynthesis. Chaperones protect their substrate from degradation in the cell before its function is required and target proteins to the export apparatus during flagellum synthesis. The chaperone-substrate complexes dock at the export apparatus ATPase where they are secreted through the central lumen of the growing flagellum by proton motive force (Thomas *et al.*, 2004). Chaperone-substrate complexes bind $FlhA_c$ at the export gate with different affinities, potentially favouring the export of the hook-filament junction substrates prior to filament formation (Kinoshita *et al.*, 2013).

As described in Section 4.2, the FliH-FliI heterodimer is a chaperone-substrate complex involved in regulating the docking and export of substrates *via* the flagellar T3SS. This complex binds FliJ, a chaperone which is instrumental in efficient substrate export through the T3SS. FliJ is responsible for recycling of chaperones for the minor filament subunits, FlgK, FlgL and FliD, but not the major subunit, FliC (Evans *et al.*, 2006). FliS is a cytoplasmic chaperone which stabilises FliC before assembly and assists flagellin export during filament extension (Auvray *et al.*, 2001). FliS prevents aggregation and premature polymerisation of FliC subunits in the cytoplasm before export. In *H. pylori*, HP1076 has been identified as a co-chaperone which promotes the correct folding and activity of FliS (Lam *et al.*, 2010).

In *Salmonella*, FlgM secretion is suppressed by all members of the *fliD* operon (FliD/S/T) (Yokoseki *et al.*, 1996). FlgN regulates the translation of FlgM as well as the export of the hook-filament junction proteins, FlgK and FlgL (Fraser *et al.*, 1999). In *Salmonella*, Flk prevents the premature secretion of FlgM; however, no such homologue has been identified in *H. pylori* (Aldridge *et al.*, 2006a). FliA has an additional role as a chaperone which promotes FlgM secretion (Aldridge *et al.*, 2006b). FliT is a chaperone which guides the filament capping protein FliD to the export gate for export (Fraser *et al.*, 1999). It also functions as a regulator of flagellar gene expression through interactions with $FlhD_4C_2$ (Yamamoto and Kutsukake, 2006). This interaction disrupts the ability of the FlhDC complex to bind the Class II promoter and therefore prevents expression of the middle flagellar genes.

4.5 Regulation of flagellum biogenesis in *H. pylori*

The composition of the flagellum of *H. pylori* closely resembles that of the extensively studied model organisms, *Salmonella* and *E. coli* (Lertsethtakarn *et al.*, 2011). However, there are notable deviations in both structure and regulation of assembly (Figure 11). At genome level, there is a clear difference in the organisation of flagellar genes in *H. pylori* and that of *E. coli* and *S. typhimurium*. The latter contain a number of distinct operons, while in *H. pylori* flagellar genes are scattered throughout the genome in multicistronic operons (Danielli *et al.*, 2010; Tomb *et al.*, 1997).

One major deviation from the model of flagellum biogenesis is the lack of an $FlhD_4C_2$ homologue in *H. pylori*. This is a clear indication that *H. pylori* flagellar gene expression is alternatively regulated. Hierarchical assembly is coupled to ordered flagellar gene expression, which is controlled by three RNA polymerase sigma factors (σ^{80}, σ^{54}, σ^{28}) (Alm *et al.*, 1999; Beier and Frank, 2000; Josenhans *et al.*, 2007; Niehus *et al.*, 2004). RpoD, or σ^{80}, controls the expression of Class I genes which include regulators and components of the basal body. In *H. pylori*, there is an additional component of the MS ring, FliY, with sequence similarity to FliN (Lowenthal *et al.*, 2009). Notably, the Class I genes include a two-component system: histidine kinase, HP0244, and its response regulator, FlgR (Spohn and Scarlato, 1999). Together, these function as enhancers of RpoN activity, the sigma factor controlling expression of the Class II genes.

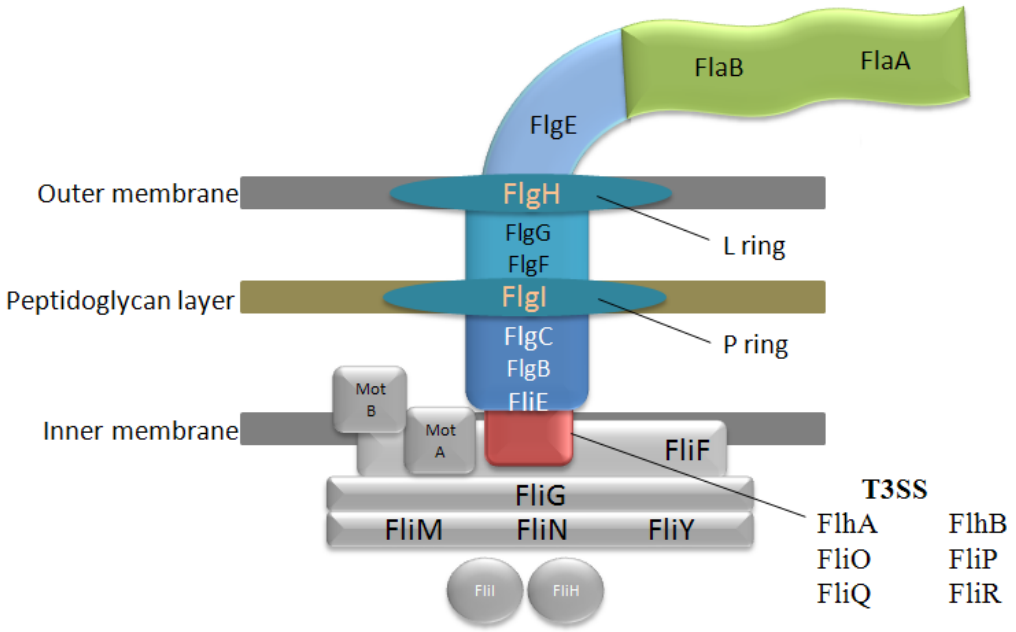


Figure 11 Overview of the known flagellar components of the *H. pylori* flagellum.

Major structural elements of the flagellum are colour-coded based and largely match the transcriptional regulatory class: components of the export apparatus (Class I) are coloured red; rod and hook proteins (Class II) are coloured blue; and filament proteins are coloured green. FlaB is an exception as it is a Class II gene. Proteins in grey either have unknown transcriptional regulation, or are not regulated within the flagellar transcriptional hierarchy. Adapted from (Lertsethtakarn *et al.*, 2011).

A transcriptional checkpoint has been identified at an early stage of flagellum assembly; mutational inactivation of early flagellar genes results in reduced transcription of Class II and Class III genes (Allan *et al.*, 2000). RpoN is a sigma factor which triggers the expression of middle flagellar genes including the rod capping protein, hook, and minor flagellin, *flaB* (Niehus *et al.*, 2004). In addition to RpoN activation, FlgR also represses premature production of the major filament protein, FlaA (Jagannathan *et al.*, 2001). An intermediate class of flagellar genes is regulated by both RpoN and FliA (σ^{28}) (Niehus *et al.*, 2004). Included in this class are components of the export apparatus, chaperones and early filament structural proteins. FliA controls the expression of late (Class III) flagellar genes which include that for the major flagellin, FlaA.

As in *Salmonella*, FlgM binds and inhibits FliA to prevent early Class III gene expression (Josenhans *et al.*, 2002). In *H. pylori*, however, FlgM is not secreted upon completion of the hook; instead it is now known to be predominantly cytoplasmic. This may indicate that the switch in expression in the case of *H. pylori* may require a different stimulus (Rust *et al.*, 2009). FlhA and FlhF have been suggested as the *H. pylori* alternative to flagellar master regulators (Niehus *et al.*, 2004). FlhA-FlgM interaction illicits a negative feedback control mechanism on the expression of Class II genes. FlhF is a GTPase which is involved in control of RpoN expression, the details of which remain unclear (Balaban *et al.*, 2009; Lertsethtakarn *et al.*, 2011; Niehus *et al.*, 2004). These proteins are central to the regulation of flagellar assembly in *H. pylori*, as illustrated by a double knock-out mutant which was aflagellate and non-motile (Niehus *et al.*, 2004).

A yeast two-hybrid study investigating the protein-protein interaction map of *H. pylori* strain 26695 identified a number of potential interaction partners of RpoN (Rain *et al.*, 2001). A protein from this subset, HP0958, was later identified as a chaperone of RpoN (Pereira and Hoover, 2005; Ryan *et al.*, 2005a). The role of HP0958 in flagellum biogenesis is discussed in more detail in Section 4.6. FlhB and FliK are involved in the substrate-specificity switch between export of rod/hook to filament proteins, but the mechanism is complex and currently unknown (Ryan *et al.*, 2005b; Smith *et al.*, 2009).

FlaA levels are controlled by a number of diverse mechanisms, from DNA supercoiling to posttranslational regulation. In addition to the mechanisms described above, *flaA* expression is also regulated by growth phase. LuxS-based quorum sensing has been shown to affect *flaA* expression whereby low cell density is associated with low *flaA* transcription while at higher cell densities, *flaA* transcription rate increases (Loh *et al.*, 2004). The spacer length between a promoter and transcriptional start site can affect the strength of expression. In *H. pylori*, the normal spacer region for σ^{28} promoters is 14 - 15 bp (Josenhans *et al.*, 2002). The *flaA* promoter spacer has a length of 13 bp which is important for growth phase dependent alterations in DNA supercoiling (Ye *et al.*, 2007). Relaxation of supercoiling resulted in reduced *flaA* transcription, whereas increased supercoiling increased *flaA* transcription levels. Interestingly, the RpoN chaperone, HP0958, also contributes to FlaA regulation, but at a posttranscriptional level (Douillard *et al.*,

2008). Therefore, the regulation of flagellum biogenesis in *H. pylori* differs from the model systems, with many details remaining unknown at present.

4.6 HP0958

HP0958 was characterised as a hypothetical protein of unknown function in the genome of *H. pylori* strain 26695 (Tomb *et al.*, 1997). HP0958 is well conserved within the *Helicobacter* genus and orthologues can also be found in some ε-proteobacteria, but are absent in *E. coli* and *Salmonella* (Figure 12).

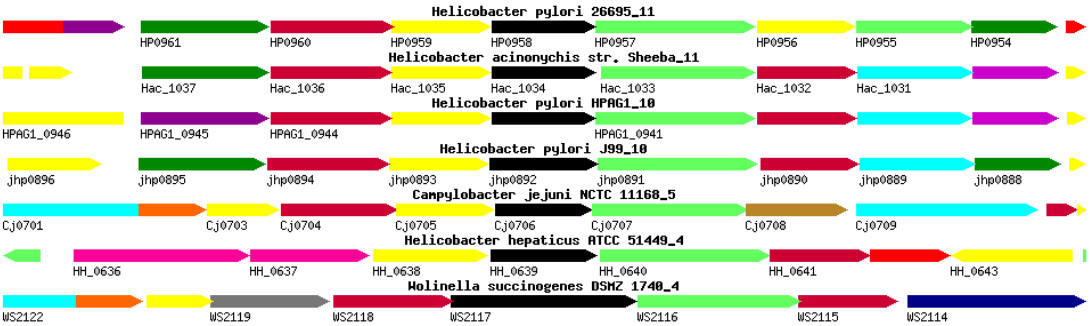


Figure 12 Conservation of HP0958 in ε-proteobacteria.

Genome region of selected strains containing HP0958 orthologues (coloured black). Image generated and adapted from TIGR CMR Genome Region Comparison online tool (Peterson *et al.*, 2001).

In 2001, the interaction network of the *H. pylori* strain 26695 proteome was published; the data can be viewed on the Hybrigenics PIMRider® platform where PIM Biological Scores indicate the confidence for predicted interaction sets (PIMRider®, Rain *et al.*, 2001). This study revealed HP0958 as a potential novel flagellar-associated protein due to predicted interactions with the flagellum biosynthesis proteins FliH and RpoN. Subsequent studies were undertaken by our group and others to characterise the predicted role of HP0958 in flagellum assembly.

Knock-out studies generated *hp0958*-null mutants which were completely aflagellate and non-motile (Pereira and Hoover, 2005; Ryan *et al.*, 2005a). Mutant *H. pylori* strains lacking HP0958 had altered transcription levels of both Class II and Class III flagellar genes (Ryan *et al.*, 2005). Further investigation has shown that HP0958 plays multiple roles in the flagellar regulatory process. HP0958 is not only a chaperone that stabilises the Class II sigma factor, RpoN, but it also interacts with

the major flagellin mRNA transcript (Douillard *et al.*, 2008). In the absence of HP0958, *H. pylori* mutants had increased *flaA* transcription but decreased levels of FlaA protein (Douillard *et al.*, 2008). Therefore, HP0958 is an essential component of flagellum biogenesis in *H. pylori* and studying it may yield insights into the different mechanism by which assembly is regulated in *H. pylori* compared to the model systems.

A hypothesis for the mechanism by which HP0958 influences flagellum assembly was proposed by Douillard *et al.* (Figure 13) whereby HP0958 interacts with FliH, potentially to guide the *flaA* mRNA transcript to the export apparatus in advance of *flaA* translation and assembly of the filament. In this model, RpoN is less stable in the absence of HP0958, resulting in no Class II expression. When HP0958 binds RpoN, the sigma factor is stabilised and can initiate transcription of the middle flagellar genes. During the switch in substrate specificity from rod/hook to filament subunits, HP0958 interacts with the *flaA* mRNA transcript to destabilize it, in order to prevent premature secretion of the major flagellin. Interaction with FliH guides the HP0958-*flaA* mRNA complex to the export apparatus where translation and export of FlaA subunits can begin (Douillard *et al.*, 2008).

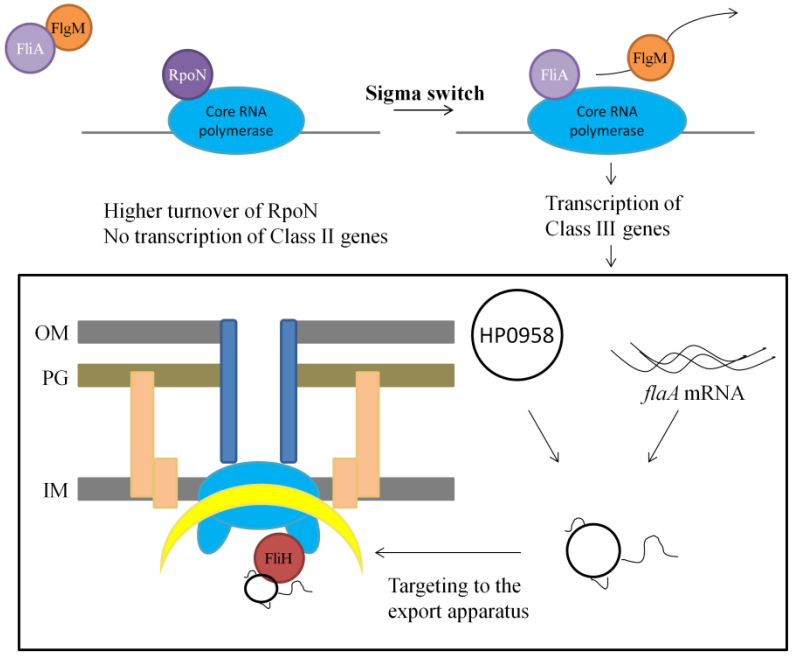


Figure 13 Proposed model for the role of HP0958 in flagellum biogenesis.

Adapted from (Douillard *et al.*, 2008).

The crystal structure of HP0958 was solved in 2010, providing insights into the mechanism by which it can interact with flagellar proteins and RNA (Caly *et al.*, 2010). It revealed that HP0958 consists of two domains: an N-terminal, anti-parallel coiled-coil and a C-terminal zinc-finger domain (Figure 14).

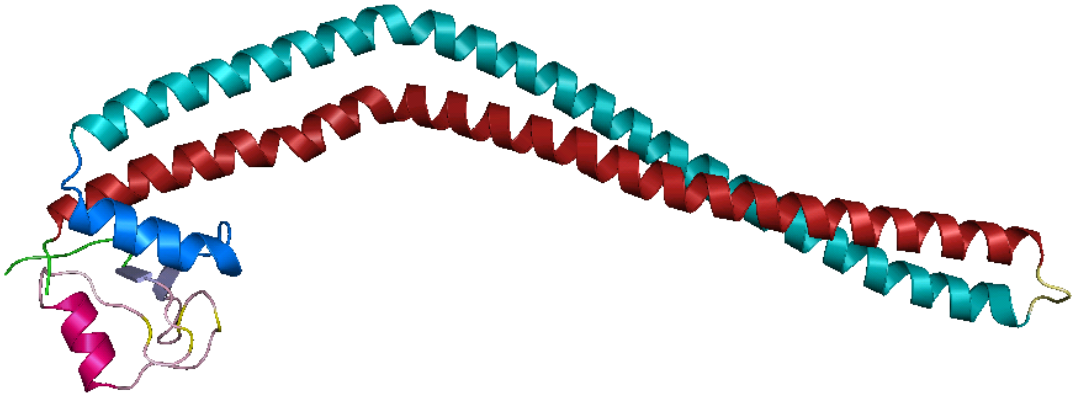


Figure 14 Structural domains of HP0958.

Image generated in Pymol highlighting the secondary structural elements of HP0958: helix 1 (red); helix 2 (cyan); helix 3 (blue); helix 4 (pink); and two beta strands in the Zn-ribbon domain (mauve).The four cystine residues of the Zn-finger are yellow.

The coiled-coil domain reveals a highly elongated, kinked, anti-parallel structure. The motif consists of two α-helices supercoiled around one another. The residues are arranged in a heptad repeat (**a-g**) where residues at **a** and **d** are generally apolar, *e.g.* leucine, valine and isoleucine, and hydrophobic. The 4-3 hydrophobic repeat along with charged residues at **e** and **g** contribute to the stability of the structure (Oakley and Hollenbeck, 2001; Parry *et al.*, 2008). Helix 3 (residues 173-185) separates the coiled-coil and Zn-finger and interacts with both domains. Discontinuities in the heptad sequence are called stutters, stammers and skips (Brown *et al.*, 1996; Lupas and Gruber, 2005). The coiled-coil kinks strongly in helix 1 at residue Arg29 and helix 2 kinks at Glu143 as a result of a stammer in the HP0958 structure (Caly *et al.*, 2010). Hydrophobic and positively charged residues have a preference to form coiled-coils: alanine, glutamic acid, lysine, leucine and arginine (Gromiha and Parry, 2004). Hydrophobicity, salt bridges (between residues Ile121 and Glu52; Lys85 and Glu89; and Lys73 and Glu96) and hydrogen bonding contribute to the stability of this motif (Caly *et al.*, 2010).

The Zn-ribbon domain (residues 174-238) contains a large number of solvent-exposed aromatic (Phe178, Tyr179, Trp185 and Tyr211) and positively charged (Arg181, Arg184, Arg205, Lys209) amino acids (Caly *et al.*, 2010). An abundance of these types of residues is associated with involvement in protein-RNA interactions (Ellis *et al.*, 2007; Jones *et al.*, 2001). In HP0958, the Zn-finger domain contains the consensus sequence CXGCX20CPHCGR (where X is any amino acid) involving four cystines co-ordinating one zinc ion (Caly *et al.*, 2010). Aromatic residues in zinc ribbon domains tend to form aromatic stacking interactions with nucleic acid bases. Positively charged residues can interact with the phosphate group of nucleic acids, as well as with other proteins (Gamsjaeger *et al.*, 2007; Laity *et al.*, 2001).

The elucidation of protein structures has contributed to our understanding of their function. Flagellum biogenesis models have been aided by emerging structural analyses of proteins from various organisms that compose the hook/basal body complex, the export apparatus and the filament. The crystal structure of the FliC chaperone in complex with its co-chaperone, HP1076, revealed a hydrophobic binding interface distinct from the FliS-FliC binding site (Lam *et al.*, 2010). The structure of FliT revealed that this chaperone interacts with the FlhDC complex, FliI and FliJ through its C-terminal helices indicating a conformational change in FliT is responsible for the switch in binding preference (Imada *et al.*, 2010). Therefore, structure-function analysis of HP0958 could provide key insights into interactions during flagellum biogenesis.

5 Aims of this Study

Bacterial motility through the use of flagella has been extensively studied in the enteric model systems *E. coli* and *Salmonella*. Motility in *H. pylori* is a key feature of pathogenesis and is essential for colonisation of its human host. While the composition of the flagellum in *H. pylori* closely mirrors that of the model systems, there are a number of differences in the regulation of assembly. Therefore, there is a need to further investigate the mechanism by which *H. pylori* controls the complex, hierarchical process of flagellar assembly.

A key feature of *H. pylori* is its extreme genetic plasticity. With the relatively recent upsurgence in genome sequencing, subtle differences between bacterial strains at the gene level can be identified. For instance, an additional level of regulation of *H. pylori* motility is phase-variation (Josenhans *et al.*, 2000). Today, there is a large volume of genome sequence information from *H. pylori* strains. However, *H. pylori* CCUG 17874 (the highly motile type-strain for the species) is frequently used in motility studies, and its genome has not been sequenced. This strain is not readily transformable, which is a barrier to the use of this strain for motility studies involving genetic manipulations. On the other hand, another motile strain, P79, is readily transformable.

HP0958 is an essential component of flagellar biogenesis (Ryan *et al.*, 2005a). The crystal structure revealed two domains, an N-terminal coiled-coil, and a C-terminal Zn-finger (Caly *et al.*, 2010). Initial knock-out studies as well as structure-function analysis of HP0958 provided clues as to how it regulates flagellum construction (Caly *et al.*, 2010; Pereira and Hoover, 2005; Ryan *et al.*, 2005a). HP0958 is a chaperone of RpoN and is predicted to interact with the ATPase inhibitor, FliH; however, how HP0958 forms these interactions at a structural level is unknown. Point-mutation of HP0958 indicated that the Zn-finger may be prominent in the HP0958-*flaA* mRNA interaction. A hypothesis for the function of HP0958 during flagellum biogenesis (outlined in Section 4.6) included HP0958 targeting the *flaA* transcript to the export apparatus through an interaction with FliH.

Therefore, the global objective of this study was to expand the current understanding of flagellum biogenesis and regulation in *H. pylori*. Additionally, genome comparative analyses was performed in the hope that it would provide

insights into the determinants for natural competence, as well as a broader definition of the core genome of *H. pylori*.

The aims of this study were:

- to sequence the genomes of *H. pylori* strains CCUG 17874 and P79
- to perform comparative genomics on these strains and the currently sequenced, publically available genomes of other *H. pylori* strains
- to define the interacting regions within the HP0958, RpoN and FliH
- to investigate the potential role of HP0958 in switching between expression of the σ^{54} and the σ^{28} regulons during flagellum biogenesis
- to determine the effect of expressing site-directed mutant derivatives of HP0958 upon flagellum biosynthesis.

The results of this study were collated into a short publication (a genome announcement) and an expanded comparative analysis (both described in Chapter 2), and a detailed analysis of the role of HP0958 in flagellum biogenesis (Chapter 3).

6 References

Abby SS, Rocha EPC. 2012. The non-flagellar type III secretion system evolved from the bacterial flagellum and diversified into host-cell adapted systems. *PLoS Genet.* 8: e1002983.

Abdallah AM, Gey van Pittius NC, Champion P a D, Cox J, *et al.* 2007. Type VII secretion--mycobacteria show the way. *Nat. Rev. Microbiol.* 5: 883–91.

Achtman M, Azuma T, Berg DE, Ito Y, *et al.* 1999. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol. Microbiol*. 32: 459–70.

Agunod M, Yamaguchi N, Lopez R, Luhby L, Glass GBJ. 1969. Correlative study of hydrochloric acid, pepsin, and intrinsic factor secretion in newborns and infants. *Am. J. Dig. Dis.* 14: 400–414.

Akeda Y, Galán JE. 2005. Chaperone release and unfolding of substrates in type III secretion. *Nature* 437: 911–5.

Akopyants NS, Clifton SW, Kersulyte D, Crabtree JE, *et al.* 1998. Analyses of the cag pathogenicity island of *Helicobacter pylori*. *Mol. Microbiol*. 28: 37–53.

Aldridge P, Karlinsey JE, Becker E, Chevance FF V, Hughes KT. 2006a. Flk prevents premature secretion of the anti-sigma factor FlgM into the periplasm. *Mol. Microbiol.* 60: 630–43.

Aldridge PD, Karlinsey JE, Aldridge C, Birchall C, *et al.* 2006b. The flagellar-specific transcription factor , sigma 28 , is the Type III secretion chaperone factor FlgM. *Genes Dev.* 20: 2315–2326.

Allan E, Dorrell N, Foynes S, Anyim M, Wren BW. 2000. Mutational analysis of genes encoding the early flagellar components of *Helicobacter pylori*: evidence for transcriptional regulation of flagellin A biosynthesis. *J. Bacteriol.* 182: 5274–5277.

Alm RA, Ling LL, Moir DT, King BL, *et al.* 1999. Genomic sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 397: 186–190.

Alm RA, Trust TJ. 1999. Analysis of the genetic diversity of *Helicobacter pylori*: the tale of two genomes. *J. Mol. Med*. (Berl). 77: 834–46.

Altman E, Chandan V, Harrison B a, Veloso-Pita R, *et al.* 2012. Design and immunological properties of *Helicobacter pylori* glycoconjugates based on a truncated lipopolysaccharide lacking Lewis antigen and comprising an α-1,6-glucan chain. *Vaccine* 30: 7332–41.

Andersen LP, Boye K, Blom J, Holck S, *et al.* 1999. Characterization of a culturable "*Gastrospirillum hominis"* (*Helicobacter heilmannii*) strain isolated from human gastric mucosa. *J. Clin. Microbiol.* 37: 1069.

Atherton JC, Cao P, Peek RMJ, Tummuru MKR, *et al.* 1995. Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. *J. Biol. Chem.* 270: 17771–17777.

Auvray F, Thomas J, Fraser GM, Hughes C. 2001. Flagellin polymerisation control by a cytosolic export chaperone. *J. Mol. Biol.* 308: 221–229.

Azevedo NF, Almeida C, Cerqueira L, Dias S, *et al.* 2007. Coccoid form of *Helicobacter pylori* as a morphological manifestation of cell adaptation to the environment. *Appl. Environ. Microbiol.* 73: 3423–7.

Backert S, Clyne M. 2011. Pathogenesis of *Helicobacter pylori* infection. *Helicobacter* 16: 19–25.

Bageshwar UK, Musser SM. 2007. Two electrical potential-dependent steps are required for transport by the *Escherichia coli* Tat machinery. *J. Cell Biol.* 179: 87–99.

Bagnoli F, Buti L, Tompkins L, Covacci A, Amieva MR. 2005. *Helicobacter pylori* CagA induces a transition from polarized to invasive phenotypes in MDCK cells. *Proc. Natl. Acad. Sci. U. S. A.* 102: 16339–44.

Bahari HM, Ross IN, Turnberg L a. 1982. Demonstration of a pH gradient across the mucus layer on the surface of human gastric mucosa in vitro. *Gut* 23: 513–6.

Balaban M, Joslin SN, Hendrixson DR. 2009. FlhF and its GTPase activity are required for distinct processes in flagellar gene regulation and biosynthesis in *Campylobacter jejuni*. *J. Bacteriol.* 191: 6602–11.

Ballister ER, Lai AH, Zuckermann RN, Cheng Y, Mougous JD. 2008. In vitro self-assembly of tailorable nanotubes from a simple protein building block. *Proc. Natl. Acad. Sci. U. S. A.* 105.

Baltrus DA, Amieva MR, Covacci A, Lowe TM, *et al.* 2009. The Complete genome sequence of *Helicobacter pylori* strain G27. *J. Bacteriol.* 191: 447–448.

Bauerfeind P, Garner R, Dunn BE, Mobley HL. 1997. Synthesis and activity of *Helicobacter pylori* urease and catalase at low pH. *Gut* 40: 25–30.

Bazzoli F, Zagari RM, Fossi S, Pozzato P, *et al.* 1993. Efficacy and tolerability of a short-term low-dose triple therapy for eradication of *Helicobacter pylori*. *Gastroenterology* 104: 40A.

Beier D, Frank R. 2000. Molecular characterization of two-component systems of *Helicobacter pylori*. *J. Bacteriol.* 182: 2068–76.

Bellack NR, Koehoorn MW, MacNab YC, Morshed MG. 2006. A conceptual model of water's role as a reservoir in *Helicobacter pylori* transmission: a review of the evidence. *Epidemiol. Infect.* 134: 439–49.

Benaissa M, Babin P, Quellard N, Pezennec L, *et al.* 1996. Changes in *Helicobacter pylori* ultrastructure and antigens during conversion from the bacillary to the coccoid form. *Infect. Immun.* 64: 2331–5.

Berg B van den, Clemons WMJ, Collinson I, Modis Y, *et al.* 2004. X-ray structure of a protein-conducting channel. *Nature* 427: 36–44.

Blanchard TG, Czinn SJ, Correa P, Nakazawa T, *et al.* 2013. Genome sequences of 65 *Helicobacter pylori* strains isolated from asymptomatic individuals and patients with gastric cancer, peptic ulcer disease, or gastritis. *Pathog. Dis.* 68: 39–43.

Blaser MJ. 1990. *Helicobacter pylori* and the pathogenesis of gastroguodenal inflammation. *J. Infect. Dis.* 161: 626–633.

Blaser MJ, Atherton JC. 2004. *Helicobacter pylori* persistence: biology and disease. *J. Clin. Invest.* 113: 321–333.

Borén T, Falk P, Roth K a, Larson G, Normark S. 1993. Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens. *Science* 262: 1892–5.

Bovenkamp JHB Van de, Mahdavi J, Korteland-Van Male AM, Büller H A, *et al.* 2003. The MUC5AC glycoprotein is the primary receptor for *Helicobacter pylori* in the human stomach. *Helicobacter* 8: 521–32.

Boyer F, Fichant G, Berthod J, Vandenbrouck Y, Attree I. 2009. Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? *BMC Genomics* 10: 104.

Braun TF, Poulson S, Gully JB, Empey C, *et al.* 1999. Function of Proline Residues of MotA in Torque Generation by the Flagellar Motor of *Escherichia coli*. *J. Bacteriol.* 181: 3542-3551.

Brennan PJ, Nikaido H. 1995. The envelope of mycobacteria. *Annu. Rev. Biochem.* 64: 29–63.

Brown JH, Cohen C, Parry DA. 1996. Heptad breaks in alpha-helical coiled coils: stutters and stammers. *Proteins* 26: 134–145.

Brown LM. 2000. *Helicobacter pylori*: epidemiology and routes of transmission. *Epidemiol. Rev.* 22: 283–97.

Buck GE, Smith JS, Buckt GE, Smith JS. 1987. Medium supplementation for growth of Medium Supplementation for Growth of *Campylobacter pyloridis*. *J. Clin. Microbiol.* 25: 597-599.

Bücker R, Azevedo-Vethacke M, Groll C, Garten D, *et al.* 2012. *Helicobacter pylori* colonization critically depends on postprandial gastric conditions. *Sci. Rep.* 2: 1-10.

Bury-Moné S, Kaakoush NO, Asencio C, Mégraud F, *et al.* 2006. Is *Helicobacter pylori* a true microaerophile? *Helicobacter* 11: 296–303.

Calvet X, Ramírez Lázaro M-J, Lehours P, Mégraud F. 2013. Diagnosis and epidemiology of *Helicobacter pylori* infection. *Helicobacter* 18: 5–11.

Calvet X, Sánchez-Delgado J, Montserrat A, Lario S, *et al.* 2009. Accuracy of diagnostic tests for *Helicobacter pylori*: a reappraisal. *Clin. Infect. Dis.* 48: 1385–91.

Caly DL, O'Toole PW, Moore S a. 2010. The 2.2-Å structure of the HP0958 protein from *Helicobacter pylori* reveals a kinked anti-parallel coiled-coil hairpin domain and a highly conserved ZN-ribbon domain. *J. Mol. Biol.* 403: 405–19.

Cambau E, Allerheiligen V, Coulon C, Corbel C, *et al.* 2009. Evaluation of a new test, genotype HelicoDR, for molecular detection of antibiotic resistance in *Helicobacter pylori*. *J. Clin. Microbiol.* 47: 3600–7.

Catrenich CE, Makin KM. 1991. Characterization of the morphologic conversion of *Helicobacter pylori* from bacillary to coccoid forms. *Scand. J. Gastroenterol. Suppl.* 181: 58–64.

Cattoli G, Bart A, Klaver PS, Robijn R, *et al.* 2000. *Helicobacter acinonychis* eradication leading to the resolution of gastric lesions in tigers. *Vet. Rec.* 147: 164–165.

Cavallo P, Cianciulli A, Mitolo V, Panaro MA. 2011. Lipopolysaccharide (LPS) of *Helicobacter* modulates cellular DNA repair systems in intestinal cells. *Clin. Exp. Med.* 11: 171–9.

Cave DR, Goddard PJ. 1999. Pathobiology of *Helicobacter pylori* infection. *Yale J. Biol. Med.* 71: 43–51.

Celli JP, Turner BS, Afdhal NH, Keates S, *et al.* 2009. *Helicobacter pylori* moves through mucus by reducing mucin viscoelasticity. *Proc. Natl. Acad. Sci. U. S. A.* 106: 14321–6.

Cellini L, Grande R, Artese L, Marzio L. 2010. Detection of *Helicobacter pylori* in saliva and esophagus. *New Microbiol.* 33: 351–7.

Censini S, Lange C, Xiang Z, Crabtree JE, *et al.* 1996. Cag, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc. Natl. Acad. Sci. U. S. A.* 93: 14648–53.

Chaddock AM, Mant A, Karnauchov I, Brink S, *et al.* 1995. A new type of signal peptide: central role of a twin-arginine motif in transfer signals for the ApH-dependent thylakoidal protein translocase. *EMBO J.* 14: 2715–2722.

Chan E-C, Chang C-C, Li Y-S, Chang CA, *et al.* 2000. Purification and Characterization of Neutral Sphingomyelinase from *Helicobacter pylori*. *Biochemistry* 39: 4838–4845.

Chen J, Lin M, Li N, Lin L, She F. 2012. Therapeutic vaccination with *Salmonella*-delivered codon-optimized outer inflammatory protein DNA vaccine enhances protection in *Helicobacter pylori* infected mice. *Vaccine* 30: 5310–5.

Chevance FF V, Hughes KT. 2008. Coordinating assembly of a bacterial macromolecular machine. *Nat. Rev. Microbiol.* 6: 455–65.

Chilcott GS, Hughes KT. 2000. Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica* serovar *typhimurium* and *Escherichia coli*. *Microbiol. Mol. Biol. Rev*. 64: 694–708.

Choi YJ, Kim N, Lim J, Jo SY, *et al.* 2012. Accuracy of diagnostic tests for *Helicobacter pylori* in patients with peptic ulcer bleeding. *Helicobacter* 17: 77–85.

Christensen JE, Pacheco S a, Konkel ME. 2009. Identification of a Campylobacter jejuni-secreted protein required for maximal invasion of host cells. *Mol. Microbiol.* 73: 650–62.

Chung C, Olivares A, Torres E, Yilmaz O, *et al.* 2010. Diversity of VacA intermediate region among *Helicobacter pylori* strains from several regions of the world. *J. Clin. Microbiol.* 48: 690–6.

Claret L, Calder SR, Higgins M, Hughes C. 2008. Oligomerization and activation of the FliI ATPase central to bacterial flagellum assembly. *Mol. Microbiol*. 48: 1349–1355.

Clemens J, Albert JM, Rao MM, Huda S, *et al.* 1996. Sociodemographic, hygienic and nutritional correlates of *Helicobacter pylori* infection of young Bangladeshi children. *Pediatr. Infect. Dis. J.* 15: 1113–1118.

Clyne M, Drumm B, Clyne M, Drumm B. 1996. The urease enzyme of *Helicobacter pylori* does not function as an adhesin. *Infect. Immun*. 64: 2817–2820.

Cornelis GR. 2006. The type III secretion injectisome. *Nat. Rev. Microbiol.* 4: 811–25.

Covacci A, Censini S, Bugnoli M, Petracca R, *et al.* 1993. Molecular characterization of the 128-kDa immunodominant antigen of *Helicobacter pylori* associated with cytotoxicity and duodenal ulcer. *Proc. Natl. Acad. Sci. U. S. A.* 90: 5791–5.

Crabtree JE, Farmery SM, Lindley IJ, Figura N, *et al.* 1994. CagA/cytotoxic strains of *Helicobacter pylori* and interleukin-8 in gastric epithelial cell lines. *J. Clin. Pathol.* 47: 945–950.

Croxen MA, Sisson G, Melano R, Hoffman PS. 2006. The *Helicobacter pylori* chemotaxis receptor TlpB (HP0103) is required for pH taxis and for colonization of the gastric mucosa. *J. Bacteriol.* 188: 2656–2665.

Cullen TW, Giles DK, Wolf LN, Ecobichon C, *et al.* 2011. *Helicobacter pylori* versus the host: remodeling of the bacterial outer membrane is required for survival in the gastric mucosa. *PLoS Pathog.* 7: e1002454.

Cullen TW, Trent MS. 2010. A link between the assembly of flagella and lipooligosaccharide of the Gram-negative bacterium *Campylobacter jejuni*. *Proc. Natl. Acad. Sci. U. S. A.* 107: 5160–5.

Cutler a F, Havstad S, Ma CK, Blaser MJ, *et al.* 1995. Accuracy of invasive and noninvasive tests to diagnose *Helicobacter pylori* infection. *Gastroenterology* 109: 136–41.

Daleke MH, Ummels R, Bawono P, Heringa J, *et al.* 2012. General secretion signal for the mycobacterial type VII secretion pathway. *Proc. Natl. Acad. Sci. U. S. A.* 109: 11342–7.

Danielli A, Amore G, Scarlato V. 2010. Built shallow to maintain homeostasis and persistent infection: insight into the transcriptional regulatory network of the gastric human pathogen *Helicobacter pylori*. *PLoS Pathog.* 6: e1000938.

Ding Z, Christie PJ. 2003. *Agrobacterium tumefaciens* Twin-Arginine-dependent translocation is important for virulence, flagellation, and chemotaxis but not type IV. *J. Bacteriol.* 185: 760-771.

Dorer MS, Talarico S, Salama NR. 2009. *Helicobacter pylori*'s unconventional role in health and disease. *PLoS Pathog.* 5: e1000544.

Dorji D, Dendup T, Malaty HM, Wangchuk K, *et al.* 2013. Epidemiology of *Helicobacter pylori* in Bhutan: The Role of Environment and Geographic Location. *Helicobacter* 19: 69-73.

Dorrell N, Martino MC, Stabler RA, Ward SJ, *et al.* 1999. Characterization of *Helicobacter pylori* PldA, a phospholipase with a role in colonization of the gastric mucosa. *Gastroenterology* 1098–1104.

Douillard FP, Ryan K a, Caly DL, Hinds J, *et al.* 2008. Posttranscriptional regulation of flagellin synthesis in *Helicobacter pylori* by the RpoN chaperone HP0958. *J. Bacteriol.* 190: 7975–84.

Douzi B, Filloux A, Voulhoux R. 2012. On the path to uncover the bacterial type II secretion system. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 367: 1059–72.

Duncan SS, Valk PL, McClain MS, Shaffer CL, *et al.* 2013. Comparative genomic analysis of East Asian and non-Asian *Helicobacter pylori* strains identifies rapidly evolving genes. *PLoS One* 8: e55120.

Dunn BE, Cohen H, Blaser MJ. 1997. *Helicobacter pylori*. *Clin. Microbiol. Rev.* 10: 720–41.

Duong F, Lazdunski A, Murgier M. 1996. Protein secretion by heterologous bacterial ABC-transporters: the C-terminus secretion signal of the secreted protein confers high recognition specificity. *Mol. Microbiol.* 21: 459–470.

Durand É, Bernadac A, Ball G, Sturgis JN, Filloux A. 2003. Type II Protein Secretion in *Pseudomonas aeruginosa*: the pseudopilus is a multifibrillar and adhesive structure. *J. Bacteriol.* 185: 2749-2758.

Eaton KA, Krakowka S. 1994. Effect of gastric pH on urease-dependent colonization of gnotobiotic piglets by *Helicobacter pylori. Infect. Immun.* 62: 3604–7.

Eaton KA, Morgan DR, Krakowka S. 1992. Motility as a factor in the colonisation of gnotobiotic piglets by *Helicobacter pylori*. *J. Med. Microbiol.* 37: 123–7.

Eaton KA, Dewhirst FE, Radin MJ, Fox JG, *et al.* 1993. *Helicobacter* acinonyx sp. nov., Isolated from Cheetahs with Gastritis. *Int. J. Syst. Bacteriol.* 43: 99–106.

Eidt S, Stolte M, Fischer R. 1994. *Helicobacter pylori* gastritis and primary gastric non-Hodgkin's lymphomas. *J. Clin. Pathol.* 47: 436–439.

Ekström AM, Held M, Hansson L-E, Engstrand L, Nyrén O. 2001. *Helicobacter pylori* in gastric cancer established by CagA immunoblot as a marker of past infection. *Gastroenterology* 121: 784–791.

Ellis JJ, Broom M, Jones S. 2007. Protein – RNA Interactions: structural analysis and functional classes. *Proteins Struct. Funct. Bioinforma.* 66: 903–911.

Erhardt M, Hirano T, Su Y, Paul K, *et al.* 2010. The role of the FliK molecular ruler in hook-length control in *Salmonella enterica*. *Mol. Microbiol.* 75: 1272–84.

Evans LDB, Poulter S, Terentjev EM, Hughes C, Fraser GM. 2013. A chain mechanism for flagellum growth. *Nature* 504: 287–90.

Evans LDB, Stafford GP, Ahmed S, Fraser GM, Hughes C. 2006. An escort mechanism for cycling of export chaperones during flagellum assembly. *Proc. Natl. Acad. Sci. U. S. A.* 103: 17474–9.

Fan F, Macnab RM. 1996. Enzymatic Characterization of FliI. *J. Biol. Chem.* 271: 31981–31988.

Fan F, Ohnishi K, Francis NR, Macnab RM. 1997. The FliP and FliR proteins of *Salmonella typhimurium*, putative components of the type III flagellar export apparatus, are located in the flagellar basal body. *Mol. Microbiol*. 26: 1035–46.

Farnbacher M, Jahns T, Willrodt D, Daniel R, *et al.* 2010. Sequencing, annotation, and comparative genome analysis of the gerbil-adapted *Helicobacter pylori* strain B8. *BMC Genomics* 11: 335.

Ferreira RM, Machado JC, Letley D, Atherton JC, *et al.* 2012. A novel method for genotyping the *Helicobacter pylori* vacA intermediate region directly in gastric biopsy specimens. *J. Clin. Microbiol.* 50: 3983–9.

Ferris HU, Furukawa Y, Minamino T, Kroetz MB, *et al.* 2005. FlhB regulates ordered export of flagellar components via autocleavage mechanism. *J. Biol. Chem.* 280: 41236–42.

Fialho AMN, Braga ABC, Braga Neto MB, Carneiro JG, *et al.* 2010. Younger siblings play a major role in *Helicobacter pylori* transmission among children from a low-income community in the Northeast of Brazil. *Helicobacter* 15: 491–6.

Fiedorik SC, Malaty HM, Evans DL, Pumphrey CL, *et al.* 1991. Factors influencing the epidemiology of *Helicobacter pylori* infection in children. Pediatrics 88: 578–582.

Fischbach W, Malfertheiner P, Hoffmann JC, Bolten W, *et al.* 2009. *Helicobacter pylori* and gastroduodenal ulcer disease. *Dtsch. Arztebl. Int.* 106: 801–8.

Fischer W, Windhager L, Rohrer S, Zeiller M, *et al.* 2010. Strain-specific genes of *Helicobacter pylori*: genome evolution driven by a novel type IV secretion system and genomic island transfer. *Nucleic Acids Res.* 38: 6089–101.

Fox JG. 1994. Gastric disease in ferrets: effects of *Helicobacter mustelae*, nitrosamines and reconstructive gastric surgery. *Eur. J. Gastroenterol. Hepatol*. 6: S57–65.

Fox JG. 1995. Non-human reservoirs of *H. pylori*. *Aliment. Pharmacol. Ther*. 9: 93–103.

Fox JG. 2002. The non-*H. pylori Helicobacter*s: their expanding role in gastrointestinal and systemic diseases. *Gut* 50: 273–283.

Fox JG, Boutin SR, Handt LK, Taylor NS, *et al.* 2007. Isolation and characterization of a novel *Helicobacter* species, "*Helicobacter macacae*," from rhesus monkeys with and without chronic idiopathic colitis. *J. Clin. Microbiol.* 45: 4061–3.

Fox JG, Dangler C a., Sager W, Borkowski R, Gliatto JM. 1997. *Helicobacter mustelae*-associated gastric adenocarcinoma in ferrets (*Mustela putorius furo*). *Vet. Pathol.* 34: 225–229.

Fox JG, Dewhirst FE, Tully JG, Paster BJ, *et al.* 1994. *Helicobacter hepaticus* sp. nov., a microaerophilic bacterium isolated from livers and intestinal mucosal scrapings from mice. *J. Clin. Microbiol*. 32: 1238–45.

Fox JG, Handt L, Xu S, Shen Z, *et al.* 2001. Novel *Helicobacter* species isolated from rhesus monkeys with chronic idiopathic colitis. *J. Med. Microbiol*. 50: 421–9.

Fox JG, Sheppard BJ, Dangler CA. 2002. Germ-line p53 -targeted disruption inhibits *Helicobacter*-induced premalignant lesions and invasive gastric carcinoma through down-regulation of Th1 proinflammatory responses. *Cancer Res*. 62: 696–702.

Fox JG, Taylor NS, Howe S, Tidd M, *et al.* 2006. *Helicobacter anseris* sp. nov. and *Helicobacter brantae* sp. nov., isolated from feces of resident Canada geese in the greater Boston area. *Appl. Environ. Microbiol.* 72: 4633–7.

Fox JG, Yan LL, Dewhirst FE, Paster BJ, *et al.* 1995. *Helicobacter bilis* sp . nov ., a novel *Helicobacter* species isolated from bile , livers , and intestines of aged , inbred mice. *J. Clin. Microbiol*. 33: 445.

Foynes S, Dorrell N, Ward SJ, Stabler RA, *et al.* 2000. *Helicobacter pylori* possesses two CheY response regulators and a histidine kinase sensor , CheA , which are essential for chemotaxis and colonization of the gastric mucosa. *Infect. Immun*. 68: 2016–2023.

Francis NR, Irikura VM, Yamaguchi S, DeRosier DJ, Macnab RM. 1992. Localization of the *Salmonella typhimurium* flagellar switch protein FliG to the cytoplasmic M-ring face of the basal body. *Proc. Natl. Acad. Sci. U. S. A.* 89: 6304–8.

Francis NR, Sosinsky GE, Thomas D, DeRosier DJ. 1994. Isolation, characterisation and structure of the bacterial flagellar motors containing the switch complex. *J. Mol. Biol.* 235: 1261–1270.

Fraser GM, Bennett JC, Hughes C. 1999. Substrate-specific binding of hook-associated proteins by FlgN and FliT, putative chaperones for flagellum assembly. *Mol. Microbiol*. 32: 569–80.

Frenck RW, Clemens J. 2003. *Helicobacter* in the developing world. *Microbes Infect.* 5: 705–713.

Fronzes R, Christie PJ, Waksman G. 2009. The structural biology of type IV secretion systems. *Nat. Rev. Microbiol.* 7: 703–14.

Furuta Y, Yahara K, Hatakeyama M, Kobayashi I. 2011. Evolution of *cagA* oncogene of *Helicobacter pylori* through recombination. *PLoS One* 6: e23499.

Galán JE, Collmer A. 1999. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* 284: 1322–1328.

Galmiche A, Rassow J, Doye A, Cagnol S, *et al.* 2000. The N-terminal 34 kDa fragment of *Helicobacter pylori* vacuolating cytotoxin targets mitochondria and induces cytochrome c release. *EMBO J.* 19: 6361–70.

Gamsjaeger R, Liew CK, Loughlin FE, Crossley M, Mackay JP. 2007. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem. Sci.* 32: 63–70.

Gebert B, Fischer W, Weiss E, Hoffmann R, Haas R. 2003. *Helicobacter pylori* vacuolating cytotoxin inhibits T lymphocyte activation. *Science* 301: 1099–102.

Geis G, Suerbaum S, Forsthoff B, Leying H, Opferkuch W. 1993. Ultrastructure and biochemical studies of the flagellar sheath of *Helicobacter pylori*. *J. Med. Microbiol.* 38: 371–7.

Gerhard M, Lehn N, Neumayer N, Borén T, *et al.* 1999. Clinical relevance of the *Helicobacter pylori* gene for blood-group antigen-binding adhesin. *Proc. Natl. Acad. Sci. U. S. A.* 96: 12778–83.

Ghosh S, Karin M, Haven N. 2002. Missing Pieces in the NF-κB Puzzle. *Cell* 109: S81–96.

Gillen KL, Hughes KT. 1991. Negative regulatory loci coupling flagellin synthesis to flagellar assembly in *Salmonella typhimurium*. *J. Bacteriol*. 173: 2301–2310.

Girardin SE, Boneca IG, Carneiro L a M, Antignac A, *et al.* 2003. Nod1 detects a unique muropeptide from gram-negative bacterial peptidoglycan. *Science* 300: 1584–7.

Gohlke U, Pullan L, McDevitt CA, Porcelli I, *et al.* 2005. The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter. *Proc. Natl. Acad. Sci. U. S. A.* 102: 10482–10486.

González-Pedrajo B, Fraser GM, Minamino T, Macnab RM. 2002. Molecular dissection of Salmonella FliH, a regulator of the ATPase FliI and the type III flagellar protein export pathway. *Mol. Microbiol.* 45: 967–82.

González-Pedrajo B, Minamino T, Kihara M, Namba K. 2006. Interactions between C ring proteins and export apparatus components: a possible mechanism for facilitating type III protein export. *Mol. Microbiol.* 60: 984–98.

Goodwin CS, Armstrong J a, Marshall BJ. 1986. Campylobacter *pyloridis*, gastritis, and peptic ulceration. *J. Clin. Pathol.* 39: 353–365.

Goodwin CS, Armstrong JA. 1990. Microbiological aspects of *Helicobacter pylori* (*Campylobacter pylori*). *Eur. J. Clin. Microbiol*. 9: 1–13.

Goodwin CS, Armstrong JA, Chilvers T, Peters M, *et al.* 1989. Transfer of *Campylobacter pylori* and *Campylobacter mustelae* to *Helicobacter* gen. nov. as *Helicobacter pylori* comb. nov. and *Helicobacter mustelae* comb. nov. respectively. *Int. J. Syst. Bacteriol.* 39: 397-405.

Goodwin CS, Mcculloch RK, Armstrong JA, Wee SH. 1985. Unusual cellular fatty acids and distinctive ultrastructure in a new spiral bacterium (*Campylobacter pyloridis*) from the human gastric mucosa. *J. Med. Microbiol*. 19: 257–267.

Gophna U, Ron EZ, Graur D. 2003. Bacterial type III secretion systems are ancient and evolved by multiple horizontal-transfer events. *Gene* 312: 151–163.

Graham DY, Kato M, Asaka M. 2008. Gastric endoscopy in the 21st century: appropriate use of an invasive procedure in the era of non-invasive testing. *Dig. Liver Dis.* 40: 497–503.

Grande R, Campli E Di, Bartolomeo S Di, Verginelli F, *et al.* 2012. *Helicobacter pylori* biofilm: a protective environment for bacterial recombination. *J. Appl. Microbiol.* 113: 669–76.

Grande R, Giulio M Di, Bessa LJ, Campli E Di, *et al.* 2011. Extracellular DNA in *Helicobacter pylori* biofilm: a backstairs rumour. *J. Appl. Microbiol.* 110: 490–8.

Gromiha MM, Parry DA. 2004. Characteristic features of amino acid residues in coiled-coil protein structures. *Biophys. Chem.* 111: 95–103.

Groote D De, Doorn LJ van, Ducatelle R, Verschuuren A, *et al.* 1999. "*Candidatus Helicobacter suis*", a gastric *Helicobacter* from pigs, and its phylogenetic relatedness to other gastrospirilla. *Int. J. Syst. Bacteriol*. 49: 1769–77.

Guo L, Liu K, Xu G, Li X, *et al.* 2012. Prophylactic and therapeutic efficacy of the epitope vaccine CTB-UA against *Helicobacter pylori* infection in a BALB/c mice model. *Appl. Microbiol. Biotechnol.* 95: 1437–44.

Haesebrouck F, Pasmans F, Flahou B, Chiers K, *et al.* 2009. Gastric *Helicobacter*s in domestic animals and nonhuman primates and their significance for human health. *Clin. Microbiol. Rev.* 22: 202–23.

Handa O, Naito Y, Yoshikawa T. 2010. *Helicobacter pylori*: a ROS-inducing bacterial species in the stomach. *Inflamm. Res.* 59: 997–1003.

Handt LK, Fox JG, Dewhirst FE, Fraser GJ, *et al.* 1994. *Helicobacter pylori* isolated from the domestic cat: public health implications. *Infect. Immun*. 62: 2367–74.

Harvey RF, Lane JA, Nair P, Egger M, *et al.* 2010. Clinical trial: prolonged beneficial effect of *Helicobacter pylori* eradication on dyspepsia consultations-the Bristol *Helicobacter* Project. *Aliment. Pharmacol. Ther.* 32: 394–400.

Hauser AR. 2009. The type III secretion system of *Pseudomonas aeruginosa*: infection by injection. *Nat. Rev. Microbiol.* 7: 654–65.

Hazell SL, Lee A, Brady L, Hennessy W. 1986. Campylobacter *pyloridis* and gastritis: association with intercellular spaces and adaptation to an environment of mucus as important factors in colonization of the gastric epithelium. *J. Infect. Dis*. 153: 658–63.

Heep M, Kist M, Strobel S, Beck D, Lehn N. 2000. Secondary resistance among 554 isolates of *Helicobacter pylori* after failure of therapy. *Eur. J. Clin. Microbiol. Infect. Dis*. 19: 538–41.

Henderson IR, Navarro-Garcia F, Desvaux M, Fernandez RC, Ala'Aldeen D. 2004. Type V protein secretion pathway: the autotransporter story. *Microbiol. Mol. Biol. Rev*. 68: 692–744.

Hendricks JK, Mobley HL. 1997. *Helicobacter pylori* ABC transporter: effect of allelic exchange mutagenesis on urease activity. *J. Bacteriol*. 179: 5892–902.

Heuermann D, Haas R. 1998. A stable shuttle vector system for efficient genetic complementation of *Helicobacter pylori* strains by transformation and conjugation. *Mol. Gen. Genet*. 257: 519–28.

Higashi H, Tsutsumi R, Muto S, Sugiyama T, *et al.* 2002. SHP-2 tyrosine phosphatase as an intracellular target of *Helicobacter pylori* CagA protein. *Science* 295: 683–6.

Hirano T, Minamino T, Macnab RM. 2001. The role in flagellar rod assembly of the N-terminal domain of *Salmonella* FlgJ, a flagellum-specific muramidase. *J. Mol. Biol.* 312: 359–69.

Hirano T, Yamaguchi S, Oosawa K, Aizawal S. 1994. Roles of FliK and FlhB in determination of flagellar hook length in *Salmonella typhimurium. J. Bacteriol*. 176: 5439–5449.

Hofreuter D, Odenbreit S, Haas R. 2001. Natural transformation competence in *Helicobacter pylori* is mediated by the basic components of a type IV secretion system. *Mol. Microbiol*. 41: 379–91.

Hoiczyk E, Roggenkamp A, Reichenbecher M, Lupas A, Heesemann J. 2000. Structure and sequence analysis of *Yersinia* YadA and *Moraxella* UspAs reveal a novel class of adhesins. *EMBO J*. 19: 5989–5999.

Homma M, Kutsukake K, Hasebe M, Iino T, Macnab RM. 1990. FlgB, FlgC, FlgF and FlgG: A family of structurally related proteins in the flagellar basal body of *Salmonella typhimurium. J. Mol. Biol*. 211: 465–477.

Hughes KT, Gillen KL, Semon MJ, Karlinsey JE. 1993. Sensing structural intermediates in bacterial flagellar assembly by export of a negative regulator. *Science* 262: 1277–1280.

Humbert O, Dorer MS, Salama NR. 2011. Characterization of *Helicobacter pylori* factors that control transformation frequency and integration length during inter-strain DNA recombination. *Mol. Microbiol.* 79: 387–401.

Humbert O, Salama NR. 2008. The *Helicobacter pylori* HpyAXII restriction-modification system limits exogenous DNA uptake by targeting GTAC sites but shows asymmetric conservation of the DNA methyltransferase and restriction endonuclease components. *Nucleic Acids Res.* 36: 6893–906.

Hunt RH, Xiao SD, Megraud F, Leon-Barua R, *et al.* 2011. *Helicobacter pylori* in developing countries. World Gastroenterology Organisation Global Guideline. *J. Gastrointest. Liver Dis.* 20: 299–304.

Hussein NR, Argent RH, Marx CK, Patel SR, *et al.* 2010. *Helicobacter pylori* dupA is polymorphic, and its active form induces proinflammatory cytokine secretion by mononuclear cells. *J. Infect. Dis.* 202: 261–9.

Ibuki T, Uchida Y, Hironaka Y, Namba K, *et al.* 2013. Interaction between FliJ and FlhA, components of the bacterial flagellar type III export apparatus. *J. Bacteriol.* 195: 466–73.

Ikeda T, Homma M, Iino T, Asakura S, Kamiya R. 1987. Localization and stoichiometry of hook-associated proteins within *Salmonella typhimurium* flagella. *J. Bacteriol*. 169: 1168–1173.

Imada K, Minamino T, Kinoshita M, Furukawa Y, Namba K. 2010. Structural insight into the regulatory mechanisms of interactions of the flagellar type III chaperone FliT with its binding partners. *Proc. Natl. Acad. Sci.* 107: 8812–8817.

Inoue M, Tsugane S. 2005. Epidemiology of gastric cancer in Japan. *Postgrad. Med. J.* 81: 419–24.

International Agency for Research on Cancer. 1994. WORLD HEALTH ORGANIZATION IARC Monographs on the evaluation of carcinogenic risks to humans: Schistosomes , Liver Flukes and *Helicobacter pylori*. 61: 177–240.

Jacob-Dubuisson F, Locht C, Antoine R. 2001. Two-partner secretion in Gram-negative bacteria: a thrifty, specific pathway for large virulence proteins. *Mol. Microbiol*. 40: 306–313.

Jagannathan A, Constantinidou C, Penn CW. 2001. Roles of *rpoN* , *fliA* , and *flgR* in expression of flagella in *Campylobacter jejuni*. *J. Bacteriol.* 183: 2937–2942.

Jemal A, Bray F, Ferlay J. 2011. Global Cancer Statistics. *Ca. Cancer J. Clin.* 61: 69–90.

Jenks PJ, Foynes S, Ward SJ, Constantinidou C, *et al.* 1997. A fagellar-specific ATPase (FliI) is necessary for fagellar export in *Helicobacter pylori*. *FEMS Microbiol. Lett*. 152: 205–211.

Jiang X, Doyle MP. 2000. Growth supplements for *Helicobacter pylori*. *J. Clin. Microbiol*. 38: 1984–7.

Jin F, Conrad JC, Gibiansky ML, Wong GCL. 2011. Bacteria use type-IV pili to slingshot on surfaces. *Proc. Natl. Acad. Sci. U. S. A.* 108: 12617–22.

Jones CJ, Homma M, Macnab RM. 1989. L- , P- , and M-ring proteins of the flagellar basal body of *Salmonella typhimurium*: gene sequences and deduced protein sequences. *J. Bacteriol*. 171: 3890–3900.

Jones DM, Curry A. 1989. The ultrastructure of *Campylobacter pylori*. In B.J. Rathbone and R. V Heatley (eds). *Campylobacter pylori* and gastroduodenal disease., pp. 48–59, Blackwell Scientific Publications, Oxford, United Kingdom.

Jones S, Daley DT, Luscombe NM, Berman HM, Thornton JM. 2001. Protein-RNA interactions: a structural analysis. *Nucleic Acids Res*. 29: 943–54.

Joo J-S, Park K-C, Song J-Y, Kim D-H, *et al.* 2010. A thin-layer liquid culture technique for the growth of *Helicobacter pylori*. *Helicobacter* 15: 295–302.

Joosten M, Flahou B, Meyns T, Smet A, *et al.* 2013. Case report: *Helicobacter suis* infection in a pig veterinarian. *Helicobacter* 18: 392–6.

Josenhans C, Beier D, Linz B, Meyer TF, Suerbaum S. 2007. Pathogenomics of *Helicobacter*. *Int. J. Med. Microbiol.* 297: 589–600.

Josenhans C, Eaton K a, Thevenot T, Suerbaum S. 2000. Switching of flagellar motility in *Helicobacter pylori* by reversible length variation of a short homopolymeric sequence repeat in *fliP*, a gene encoding a basal body protein. *Infect. Immun*. 68: 4598–603.

Josenhans C, Niehus E, Amersbach S, Hörster A, *et al.* 2002. Functional characterization of the antagonistic flagellar late regulators FliA and FlgM of *Helicobacter pylori* and their effects on the *H. pylori* transcriptome. *Mol. Microbiol.* 43: 307–22.

Jyot J, Balloy V, Jouvion G, Verma A, *et al.* 2011. Type II secretion system of *Pseudomonas aeruginosa*: *in vivo* evidence of a significant role in death due to lung infection. *J. Infect. Dis.* 203: 1369–77.

Kaparakis M, Turnbull L, Carneiro L, Firth S, *et al.* 2010. Bacterial membrane vesicles deliver peptidoglycan to NOD1 in epithelial cells. *Cell. Microbiol.* 12: 372–85.

Karim QN, Logan RPH, Puels J, Karnholz A, Worku ML. 1998. Measurement of motility of *Helicobacter* pylor , *Campylobacter jejuni* , and *Escherichia coli* by real time computer tracking using the Hobson BacTracker. *J. Clin. Pathol.* 623–628.

Kauser F, Khan AA, Hussain MA, Carroll IM, *et al.* 2004. The *cag* pathogenicity island of *Helicobacter pylori* is disrupted in the majority of patient isolates from different human populations. *J. Clin. Microbiol.* 42: 5302–5308.

Kazetani K-I, Minamino T, Miyata T, Kato T, Namba K. 2009. ATP-induced FliI hexamerization facilitates bacterial flagellar protein export. *Biochem. Biophys. Res. Commun.* 388: 323–7.

Kennemann L, Didelot X, Aebischer T, Kuhn S, *et al.* 2011. *Helicobacter pylori* genome evolution during human infection. *Proc. Natl. Acad. Sci. U. S. A.* 108: 5033–8.

Kerr JR. 2000. An association between sudden infant death syndrome (SIDS) and *Helicobacter pylori* infection. *Arch. Dis. Child.* 83: 429–434.

Kersulyte D, Lee W, Subramaniam D, Anant S, *et al.* 2009. *Helicobacter pylori*'s plasticity zones are novel transposable elements. *PLoS One* 4: e6859.

Khalifa MM, Sharaf RR, Aziz RK. 2010. *Helicobacter pylori*: a poor man's gut pathogen? *Gut Pathog.* 2: 1-12.

Kihara M, Minamino T, Yamaguchi S, Macnab RM. 2001. Intergenic suppression between the flagellar MS Ring protein FliF of *Salmonella* and FlhA, a membrane component of its export apparatus. *J. Bacteriol.* 183: 1655–1662.

Kikuchi H, Asako K, Tansho S, Ueda T, Koshio O. 2012. Recurrent *Helicobacter cinaedi* cellulitis and bacteremia in a patient with systemic lupus erythematosus. *Intern. Med*. 51: 3185–3188.

Kim DJ, Park K, Kim J-H, Yang S-H, *et al.* 2010. *Helicobacter pylori* proinflammatory protein up- regulates NF- κB as a cell-translocating Ser/Thr kinase. *Proc. Natl. Acad. Sci. U. S. A.*107: 21418–21423.

Kim E-H, Hong K-S, Hong H, Hahm KB. 2011. Detouring the undesired route of *Helicobacter pylori*-induced gastric carcinogenesis. *Cancers* 3: 3018–3028.

Kim JS, Chang JH, Chung S Il, Yum JSUN. 1999. Molecular Cloning and Characterization of the *Helicobacter pylori fliD* Gene , an essential factor in flagellar structure and motility. *J. Bacteriol*. 181: 6969–6976.

Kinoshita M, Hara N, Imada K, Namba K, Minamino T. 2013. Interactions of bacterial flagellar chaperone-substrate complexes with FlhA contribute to co-ordinating assembly of the flagellar filament. *Mol. Microbiol.* 90: 1249–61.

Koch M, Meyer TF, Moss SF. 2013. Inflammation, immunity, vaccines for *Helicobacter pylori* infection. *Helicobacter* 18: 18–23.

Kojima S, Furukawa Y, Matsunami H, Minamino T, Namba K. 2008. Characterization of the periplasmic domain of MotB and implications for its role in the stator assembly of the bacterial flagellar motor. *J. Bacteriol.* 190: 3314–22.

Koletzko S, Koletzko B, Haisch M, Hering P, *et al.* 1995. Isotope-selective non-dispersive infrared spectrometry for detection of *Helicobacter pylori* infection with 13C-urea breath test. *Lancet* 345: 961–962.

Korotkov KV, Sandkvist M, Hol WGJ. 2012. The type II secretion system: biogenesis, molecular architecture and mechanism. *Nat. Rev. Microbiol.* 10: 336–51.

Kostrzynska M, Betts JD, Austin JW, Trust TJ. 1991. Identification, characterization, and spatial localization of two flagellin species in *Helicobacter pylori* flagella. *J. Bacteriol*. 173: 937–46.

Kubori T, Yamaguchi S, Aizawa S. 1997. Assembly of the switch complex onto the MS ring complex of *Salmonella typhimurium* does not require any other flagellar proteins. *J. Bacteriol*. 179: 813–7.

Kuipers EJ, Israel DA, Kusters JG, Blaser MJ. 1998. Evidence for a conjugation-like mechanism of DNA transfer in *Helicobacter pylori*. *J. Bacteriol*. 180: 2901–5.

Kuipers EJ, Uyterlinde AM, Pena AS, Hazenberg HJ, *et al.* 1995. Increase of *Helicobacter pylori*-associated corpus gastritis during acid suppressive therapy: implications for long-term safety. *Am. J. Gastroenterol.* 90: 1401–1406.

Kusters JG, Gerrits MM, Strijp JA Van. 1997. Coccoid forms of *Helicobacter pylori* are the morphologic manifestation of cell death. *Infect. Immun.* 65: 3672-79.

Kusters JG, Vliet AHM van, Kuipers EJ. 2006. Pathogenesis of *Helicobacter pylori* infection. *Clin. Microbiol. Rev.* 19: 449–90.

Kutsukake K. 1994. Excretion of the anti-sigma factor through a flagellar substructure couples flagellar gene expression with flagellar assembly in *Salmonella typhimurium. Mol. Gen. Gentics* 243: 605–612.

Kutsukake K, Ohya Y, Iino T. 1990. Transcriptional analysis of the flagellar regulon of Salmonella typhimurium. *J. Bacteriol*. 172: 741–7.

Kutter S, Buhrdorf R, Haas J, Schneider-Brachert W, *et al.* 2008. Protein subassemblies of the *Helicobacter pylori* Cag type IV secretion system revealed by localization and interaction studies. *J. Bacteriol.* 190: 2161–71.

Kuwahara H, Miyamoto Y, Akaike T, Kubota T, *et al.* 2000. *Helicobacter pylori* urease suppresses bactericidal activity of peroxynitrite via carbon dioxide production. *Infect. Immun.* 68: 4378–83.

Laine L, Lewin D, Naritoku W, Estrada R, Cohen H. 1996. Protective comparison of commercially available rapid urease tests for the diagnosis of *Helicobacter pylori*. *Gastrointest. Endosc*. 44: 523–526.

Laity JH, Lee BM, Wright PE. 2001. Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* 11: 39–46.

Lam WWL, Woo EJ, Kotaka M, Tam WK, *et al.* 2010. Molecular interaction of flagellar export chaperone FliS and cochaperone HP1076 in *Helicobacter pylori*. *FASEB J.* 24: 4020–32.

Lamb A, Yang X-D, Tsang Y-HN, Li J-D, *et al.* 2009. *Helicobacter pylori* CagA activates NF-kappaB by targeting TAK1 for TRAF6-mediated Lys 63 ubiquitination. *EMBO Rep.* 10: 1242–9.

Lane MC, O'Toole PW, Moore SA. 2006. Molecular basis of the interaction between the flagellar export proteins FliI and FliH from *Helicobacter pylori*. *J. Biol. Chem.* 281: 508–17.

Lara-Ramírez EE, Segura-Cabrera A, Guo X, Yu G, *et al.* 2011. New implications on genomic adaptation derived from the *Helicobacter pylori* genome comparison. *PLoS One* 6: e17300.

Lata K, Paul K, Chattopadhyay. 2014 Functional characterization of *Helicobacter pylori* TlyA: Pore-forming hemolytic activity and cytotoxic property of the protein. *Biochem. Biophys. Res. Commun.* 444: 153-157.

Leal YA, Cedillo-Rivera R, Simón JA, Velázquez JR, *et al.* 2011. Utility of stool sample-based tests for the diagnosis of *Helicobacter pylori* infection in children. *J. Pediatr. Gastroenterol. Nutr.* 52: 718–28.

Lee A, Hazell SL, O'Rourke J, Kouprach S. 1988. Isolation of a spiral-shaped bacterium from the cat stomach. *Infect. Immun*. 56: 2843–50.

Lertsethtakarn P, Ottemann KM, Hendrixson DR. 2011. Motility and chemotaxis in *Campylobacter* and *Helicobacter* . *Annu. Rev. Microbiol.* 65: 389–410.

Leunk RD, Johnson PT, David BC, Kraft WG, Mrgan DR. 1988. Cytotoxic activity in broth-culture filtrates of Campylobacter *pylori*. *J. Med. Microbiol*. 26: 93–99.

Leying H, Suerbaum S, Geis G, Haas R. 1992. Cloning and genetic characterization of a *Helicobacter pylori* flagellin gene. *Mol. Microbiol*. 6: 2863–74.

Linz B, Balloux F, Moodley Y, Manica A, *et al.* 2007. An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* 445: 915–8.

Liu R, Ochman H. 2007. Stepwise formation of the bacterial flagellar system. *Proc. Natl. Acad. Sci. U. S. A.* 104: 7116–7121.

Liu X, Fujita N, Ishihama A, Matsumura P. 1995. The C-terminal region of the alpha subunit of Escherichia coli RNA polymerase is required for transcriptional activation of the flagellar level II operons by the FlhD/FlhC complex. *J. Bacteriol.* 177:5186-5188.

Lloyd SA, Tang HUA, Wang XUN, Billings S, Blair DF. 1996. Torque generation in the flagellar motor of *Escherichia coli*: evidence of a direct role for FliG but not for FliM or FliN. *J. Bacteriol.* 178: 223–231.

Loh JT, Forsyth MH, Cover TL. 2004. Growth phase regulation of *flaA* expression in *Helicobacter pylori* is luxS dependent. *Infect. Immun.* 72: 5506–5510.

Lowenthal AC, Hill M, Sycuro LK, Mehmood K, *et al.* 2009. Functional analysis of the *Helicobacter pylori* flagellar switch proteins. *J. Bacteriol.* 191: 7147–56.

Lu H, Hsu P-I, Graham DY, Yamaoka Y. 2005. Duodenal ulcer promoting gene of *Helicobacter pylori*. *Gastroenterology* 128: 833–848.

Lubelski J, Konings WN, Driessen AJM. 2007. Distribution and physiology of ABC-type transporters contributing to multidrug resistance in bacteria. *Microbiol. Mol. Biol. Rev.* 71: 463–76.

Luket CJ, Penn CW. 1995. Identification of a 29 kDa flagellar sheath protein in *Helicobacter pylori* using a murine monoclonal antibody. *Microbiology* 141:597–604.

Luo C-H, Chiou P-Y, Yang C-Y, Lin N-T. 2012. Genome, integration, and transduction of a novel temperate phage of *Helicobacter pylori*. *J. Virol.* 86: 8781–92.

Lupas AN, Gruber M. 2005. The structure of alpha-helical coiled coils. *Adv. Protein Chem.* 70: 37–78.

Lupetti P, Heuser JE, Manetti R, Massari P, *et al.* 1996. Oligomeric and subunit structure of the *Helicobacter Pylori* vacuolating cytotoxin. *J. Cell Biol.* 133: 801–807.

MacIntyre DL, Miyata ST, Kitaoka M, Pukatzki S. 2010. The *Vibrio cholerae* type VI secretion system displays antimicrobial properties. *Proc. Natl. Acad. Sci. U. S. A.* 107: 19520–4.

Mackman N, Nicaud J-M, Gray L, Holland IB. 1986. Secretion of haemolysin by *Escherichia coli*. *Protein Secret. Export Bact.* 125: 159–181.

Macnab RM. 1977. Bacterial flagella rotating in bundles: A study in helical geometry. *Proc. Natl. Acad. Sci. U. S. A.* 74: 221–225.

Macnab RM. 2003. How bacteria assemble flagella. *Annu. Rev. Microbiol.* 57: 77–100.

Macnab RM. 2004. Type III flagellar protein export and flagellar assembly. *Biochim. Biophys. Acta* 1694: 207–17.

Magdalena J, Hachani A, Chamekh M, Jouihri N, *et al.* 2002. Spa32 regulates a switch in substrate specificity of the type III secreton of *Shigella flexneri* from needle components to Ipa proteins. *J. Bacteriol.* 184: 3343–3441.

Mahdavi J, Sondén B, Hurtig M, Olfat FO, *et al.* 2002. *Helicobacter pylori* SabA adhesin in persistent infection and chronic inflammation. *Science* 297: 573–8.

Makishima S, Komoriya K, Yamaguchi S, Aizawa SI. 2001. Length of the flagellar hook and the capacity of the type III export apparatus. *Science* 291: 2411–3.

Malaty HM, El-Kasabany A, Graham DY, Miller CC, *et al.* 2002. Age at acquisition of *Helicobacter pylori* infection: a follow-up study from infancy to adulthood. *Lancet* 359: 931–5.

Malfertheiner P, Megraud F, O'Morain C a, Atherton J, *et al.* 2012. Management of *Helicobacter pylori* infection--the Maastricht IV/Florence Consensus Report. *Gut* 61: 646–64.

Marshall BJ, Armstrong JA, McGechie DB, Glancy RJ. 1985. Attempt to fulfill Koch's postulates for *pyloric Campylobacter*. *Medical J. Aust.* 142: 436–439.

Marshall BJ, Warren JR. 1984. Unidentified curved *bacilli* in the stomach of patients with gastritis and peptic ulceration. *Lancet* 1: 1311–5.

Martino MC, Stabler RA, Zhang ZW, Farthing MJG, *et al.* 2001. *Helicobacter pylori* pore-forming cytolysin orthologue TlyA possesses *in vitro* hemolytic activity and has a role in colonization of the gastric mucosa. *Infect. Immun.* 69: 1697.

McCarter LL. 2006. Regulation of flagella. *Curr. Opin. Microbiol.* 9: 180–6.

McMurry JL, Murphy JW, González-Pedrajo B. 2006. The FliN-FliH interaction mediates localization of flagellar export ATPase FliI to the C ring complex. *Biochemistry* 45: 11790–8.

McNulty CAM, Watson DM. 1984. Spiral bacteria of the gastric antrum. *Lancet* 323: 1068–1069.

Mégraud F. 2004. H *pylori* antibiotic resistance: prevalence, importance, and advances in testing. *Gut* 53: 1374–84.

Mégraud F. 2012. The challenge of *Helicobacter pylori* resistance to antibiotics: the comeback of bismuth-based quadruple therapy. *Therap. Adv. Gastroenterol.* 5: 103–9.

Miehlke S, Kirsch C, Agha-Amiri K, Günther T, *et al.* 2000. The *Helicobacter pylori* vacA s1, m1 genotype and cagA is associated with gastric carcinoma in Germany. *Int. J. Cancer* 87: 322–7.

Minamino T, Gonzalez-Pedrajo B, Kihara M, Namba K, Macnab RM. 2003. The ATPase FliI Can Interact with the Type III Flagellar Protein Export Apparatus in the Absence of Its Regulator, FliH. *J. Bacteriol.* 185: 3983–3988.

Minamino T, Imada K, Namba K. 2008. Molecular motors of the bacterial flagella. *Curr. Opin. Struct. Biol.* 18: 693–701.

Minamino T, Macnab RM. 1999. Components of the *Salmonella* flagellar export apparatus and classification of export substrates. *J. Bacteriol*. 181: 1388–94.

Minamino T, MacNab RM. 2000. FliH, a soluble component of the type III flagellar export apparatus of *Salmonella*, forms a complex with FliI and inhibits its ATPase activity. *Mol. Microbiol*. 37: 1494–503.

Minamino T, Namba K. 2008. Distinct roles of the FliI ATPase and proton motive force in bacterial flagellar protein export. *Nature* 451: 485–8.

Minamino T, Shimada M, Okabe M, Saijo-Hamano Y, *et al.* 2010. Role of the C-terminal cytoplasmic domain of FlhA in bacterial flagellar type III protein export. *J. Bacteriol.* 192: 1929–36.

Minamino T, Tame JR, Namba K, Macnab RM. 2001. Proteolytic analysis of the FliH/FliI complex, the ATPase component of the type III flagellar export apparatus of *Salmonella*. *J. Mol. Biol.* 312: 1027–36.

Minamino T, Yoshimura SDJ, Morimoto Y V, González-Pedrajo B, *et al.* 2009. Roles of the extreme N-terminal region of FliH for efficient localization of the FliH-FliI complex to the bacterial flagellar type III export apparatus. *Mol. Microbiol.* 74: 1471–83.

Minoli G, Prada A, Schuman R, Murnick D, Rigas B. 1998. A simplified urea breath test for the diagnosis of *Helicobacter pylori* infection using the LARA system. *J. Clin. Gastroenterol*. 26: 264–266.

Mitchell DJ, McClure BG, Tubman TR. 2001. Simultaneous monitoring of gastric and oesophageal pH reveals limitations of conventional oesophageal pH monitoring in milk fed infants. *Arch. Dis. Child*. 84: 273–6.

Mizushima T, Sugiyama T, Komatsu Y. 2001. Clinical relevance of the *babA2* genotype of *Helicobacter pylori* in Japanese clinical isolates. 39:2463-2465.

Mogensen JE, Otzen DE. 2005. Interactions between folding factors and bacterial outer membrane proteins. *Mol. Microbiol.* 57: 326–46.

Molinari M, Salio M, Galli C, Norais N, *et al.* 1998. Selective inhibition of Ii-dependent antigen presentation by *Helicobacter pylori* toxin VacA. *J. Exp. Med.* 187: 135–40.

Momtaz H, Souod N, Dabiri H, Sarshar M. 2012. Study of *Helicobacter pylori* genotype status in saliva, dental plaques, stool and gastric biopsy samples. *World J. Gastroenterol.* 18: 2105–11.

Moodley Y, Linz B, Bond RP, Nieuwoudt M, *et al.* 2012. Age of the association between *Helicobacter pylori* and man. *PLoS Pathog.* 8: e1002693.

Moodley Y, Linz B, Yamaoka Y, Windsor HM, *et al.* 2009. The peopling of the Pacific from a bacterial perspective. Science 323: 527–530.

Moraes TF, Spreter T, Strynadka NC. 2008. Piecing together the type III injectisome of bacterial pathogens. *Curr. Opin. Struct. Biol.* 18: 258–66.

Morgan DR, Freedman R, Depew CE, Kraft WG. 1987. Growth of *Campylobacter pylori* in liquid media. *J. Clin. Microbiol*. 25: 2123–5.

Moriya N, Minamino T, Hughes KT, Macnab RM, Namba K. 2006. The type III flagellar export specificity switch is dependent on FliK ruler and a molecular clock. *J. Mol. Biol.* 359: 466–77.

Moriya N, Minamino T, Imada K, Namba K. 2011. Genetic analysis of the bacterial hook-capping protein FlgD responsible for hook assembly. *Microbiology* 157: 1354–62.

Mougous JD, Gifford CA, Ramsdell TL, Mekalanos JJ. 2007. Threonine phosphorylation post-translationally regulates protein secretion in *Pseudomonas aeruginosa*. *Nat. Cell Biol.* 9: 797–803.

Mould RM, Robinson C. 1991. A proton gradient is required for the transport of two lumenal oxygen-evolving proteins across the thylakoid membrane. *J. Biol. Chem.* 266: 12189–12193.

Mueller C a, Broz P, Cornelis GR. 2008. The type III secretion system tip complex and translocon. *Mol. Microbiol.* 68: 1085–95.

Muotiala A, Helander IM, Pyhala L, Kosunen TU, Morant AP. 1992. Low biological activity of *Helicobacter pylori* lipopolysaccharide. *Infect. Immun*. 60: 1714–1716.

Murata-Kamiya N. 2011. Pathophysiological functions of the CagA oncoprotein during infection by *Helicobacter pylori*. *Microbes Infect.* 13: 799–807.

Murata-Kamiya N, Kikuchi K, Hayashi T, Higashi H, Hatakeyama M. 2010. *Helicobacter pylori* exploits host membrane phosphatidylserine for delivery, localization, and pathophysiological action of the CagA oncoprotein. *Cell Host Microbe* 7: 399–411.

Murata-Kamiya N, Kurashima Y, Teishikata Y, Yamahashi Y, *et al.* 2007. *Helicobacter pylori* CagA interacts with E-cadherin and deregulates the beta-catenin signal that promotes intestinal transdifferentiation in gastric epithelial cells. *Oncogene* 26: 4617–26.

Nagini S. 2012. Carcinoma of the stomach: A review of epidemiology, pathogenesis, molecular genetics and chemoprevention. *World J. Gastrointest. Oncol.* 4: 156–69.

Navabi N, Johansson ME V, Raghavan S, Lindén SK. 2013. *Helicobacter pylori* infection impairs the mucin production rate and turnover in the murine gastric mucosa. *Infect. Immun.* 81: 829–37.

NCBI. 2013. National Center for Biotechnology Information (NCBI) Genome. Available: http: //www.ncbi.nlm.nih.gov/genome/genomes/169.

Neal JT, Peterson TS, Kent ML, Guillemin K. 2013. *H. pylori* virulence factor CagA increases intestinal cell proliferation by Wnt pathway activation in a transgenic zebrafish model. *Dis. Model. Mech.* 6: 802–10.

Nedenskov P. 1994. Nutritional requirements for growth of *Helicobacter pylori*. *Appl. Environ. Microbiol*. 60: 3450–3.

Nedenskov-Sorensen P, Bukholm G, Bovre K. 1990. Natural competence for genetic transformation in *Campylobacter pylori. J. Infect. Dis*. 161: 365–366.

Niehus E, Gressmann H, Ye F, Schlapbach R, *et al.* 2004. Genome-wide analysis of transcriptional hierarchy and feedback regulation in the flagellar system of *Helicobacter pylori*. *Mol. Microbiol.* 52: 947–61.

O'Toole PW, Austin JW, Trust TJ. 1994. Identification and molecular characterization of a major ring-forming surface protein from the gastric pathogen *Helicobacter mustelae*. *Mol. Microbiol*. 11: 349–61.

Oakley MG, Hollenbeck JJ. 2001. The design of antiparallel coiled coils. *Curr. Opin. Struct. Biol.* 11: 450–7.

Oberhettinger P, Schütz M, Leo JC, Heinz N, *et al.* 2012. Intimin and invasin export their C-terminus to the bacterial cell surface using an inverse mechanism compared to classical autotransport. *PLoS One* 7: e47069.

Ohnishi K, Kutsukake K, Suzuki H, Lino T. 1992. A novel transcriptional regulation mechanism in the flagellar regulon of *Salmonella typhimurium*: an anti-sigma factor inhibits the activity of the flagellum-specific Sigma factor, σF. *Mol. Microbiol*. 6: 3149–3157.

Ohnishi K, Ohto Y, Aizawa S, Macnab RM. 1994. FlgD is a scaffolding protein needed for flagellar hook assembly in *Salmonella typhimurium*. *J. Bacteriol.* 176: 2272-2281.

Ohnishi N, Yuasa H, Tanaka S, Sawa H, *et al.* 2008. Transgenic expression of *Helicobacter pylori* CagA induces gastrointestinal and hematopoietic neoplasms in mouse. *Proc. Natl. Acad. Sci. U. S. A.* 105: 1003–8.

Omori K, Idei A. 2003. Gram-negative bacterial ATP-binding cassette protein exporter family and diverse secretory proteins. *J. Biosci. Bioeng.* 95: 1–12.

Ottemann KM, Lowenthal AC. 2002. *Helicobacter pylori* uses motility for initial colonization and to attain robust infection. *Infect. Immun.* 70: 1984–1990.

Ou Z, Xiong L, Li D-Y, Geng L, *et al.* 2013. Evaluation of a new fluorescence quantitative PCR test for diagnosing *Helicobacter pylori* infection in children. *BMC Gastroenterol.* 13: 1-7.

Paetzel M, Karla A, Strynadka NCJ, Dalbey RE. 2002. Signal peptidases. *Chem. Rev.* 102: 4549–4579.

Pallen MJ, Matzke NJ, Derosier D, Woese C, *et al.* 2006. From The Origin of Species to the origin of bacterial flagella. Nat. Rev. Microbiol. 4: 784–790.

Palmer T, Berks BC. 2012. The twin-arginine translocation (Tat) protein export pathway. *Nat. Rev. Microbiol.* 10: 483–96.

Papanikou E, Karamanou S, Economou A. 2007. Bacterial protein secretion through the translocase nanomachine. *Nat. Rev. Microbiol.* 5: 839–51.

Parkin DM. 2006. The global health burden of infection-associated cancers in the year 2002. *Int. J. Cancer* 118: 3030–44.

Parry D a D, Fraser RDB, Squire JM. 2008. Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure. *J. Struct. Biol.* 163: 258–69.

Parsonnet J, Shmuely H, Haggerty T. 1999. Fecal and oral shedding of *Helicobacter pylori* from healthy infected adults. *JAMA* 282: 2240–5.

Paulsen IT. 2003. Multidrug efflux pumps and resistance: regulation and evolution. *Curr. Opin. Microbiol.* 6: 446–451.

Peabody CR. 2003. Type II protein secretion and its relationship to bacterial type IV pili and archaeal flagella. *Microbiology* 149: 3051–3072.

Penfold S, Lastovica A, Elisha B. 1988. Demonstration of plasmids in *Campylobacter pylori*. *J. Infect. Dis*. 157: 850–851.

Pereira L, Hoover TR. 2005. Stable accumulation of sigma 54 in *Helicobacter pylori* requires the novel protein HP0958. *J. Bacteriol.* 187: 4463–4469.

Perez-perez GI, Rothenbacher D, Brenner H. 2004. Epidemiology of *Helicobacter pylori* infection. *Helicobacter* 9: 1–6.

Perrais M, Rousseaux C, Ducourouble M-P, Courcol R, *et al.* 2013. *Helicobacter pylori* urease and flagellin alter mucin gene expression in human gastric cancer cells. *Gastric cancer* DOI 10.1007/s10120-013-0267-5.

Peterson JD, Umayam LA, Dickinson TM, Hickey EK, White O. 2001. The comprehensive microbial resource. *Nucleic Acids Res*. 123–125. Available: http: //cmr.jcvi.org/cgi-bin/CMR/CMR/CmrHomePage.cgi.

Petersson C, Forsberg M, Aspholm M, Olfat FO, *et al.* 2006. *Helicobacter pylori* SabA adhesin evokes a strong inflammatory response in human neutrophils which is down-regulated by the neutrophil-activating protein. *Med. Microbiol. Immunol.* 195: 195–206.

Pettersson J, Nordfelth R, Dubinina E, Bergman T, *et al.* 1996. Modulation of virulence factor expression by pathogen target cell contact. *Science*. 273: 1231–1233.

Phadnis SH, Parlow MH, Levy M, Ilver D, *et al.* 1996. Surface localization of *Helicobacter pylori* urease and a heat shock protein homolog requires bacterial autolysis. *Infect. Immun.* 64: 905–912.

Pijkeren J-P van, Canchaya C, Ryan KA, Li Y, *et al.* 2006. Comparative and functional analysis of sortase-dependent proteins in the predicted secretome of *Lactobacillus salivarius* UCC118. *Appl. Environ. Microbiol.* 72: 4143–53.

PIMRider® H. https: //pimr.hybrigenics.com/htmlFiles/interactionData/*pylori*.html.

Pohlner J, Halter R, Beyreuther K, Meyer TF. 1987. Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. *Nature* 325: 458–462.

Poppe M, Feller SM, Römer G, Wessler S. 2007. Phosphorylation of *Helicobacter pylori* CagA by c-Abl leads to cell motility. *Oncogene* 26: 3462–72.

Pounder RE, Ng D. 1995. The prevalence of *Helicobacter pylori* infection in different countries. *Aliment. Pharmacol. Ther.* 9: 33–9.

Poursina F, Faghri J, Moghim S, Zarkesh-Esfahani H, *et al.* 2013. Assessment of *cagE* and *babA* mRNA expression during morphological conversion of *Helicobacter pylori* from spiral to coccoid. *Curr. Microbiol.* 66: 406–13.

Prinz C, Schwendy S, Voland P. 2006. H *pylori* and gastric cancer: shifting the global burden. *World J. Gastroenterol*. 12: 5458–64.

Prüss BM, Matsumura P. 1997. Cell cycle regulation of flagellar genes. *J. Bacteriol*. 179: 5602–4.

Pugsley AP. 1993. The complete general secretory pathway in Gram-negative bacteria. *Microbiol. Rev.* 57: 50–108.

Pukatzki S, Ma AT, Revel AT, Sturtevant D, Mekalanos JJ. 2007. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc. Natl. Acad. Sci. U. S. A.* 104:15508-15513.

Pukatzki S, Ma AT, Sturtevant D, Krastins B, *et al.* 2006. Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proc. Natl. Acad. Sci. U. S. A.* 103: 1528–33.

Pukatzki S, McAuley SB, Miyata ST. 2009. The type VI secretion system: translocation of effectors and effector-domains. *Curr. Opin. Microbiol.* 12: 11–7.

Radin JN, Gaddy JA, González-Rivera C, Loh JT, *et al.* 2013. Flagellar localization of a *Helicobacter pylori* autotransporter protein. *MBio* 4: e00613-12.

Rain JC, Selig L, Reuse H De, Battaglia V, *et al.* 2001. The protein-protein interaction map of *Helicobacter pylori*. *Nature* 409: 211–5.

Raymond J, Thiberge J-M, Kalach N, Bergeret M, *et al.* 2008. Using macro-arrays to study routes of infection of *Helicobacter pylori* in three families. *PLoS One* 3: e2259.

Renshaw PS, Lightbody KL, Veverka V, Muskett FW, *et al.* 2005. Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. *EMBO J.* 24: 2491–8.

Rhead JL, Letley DP, Mohammadi M, Hussein N, *et al.* 2007. A new *Helicobacter pylori* vacuolating cytotoxin determinant, the intermediate region, is associated with gastric cancer. *Gastroenterology* 133: 926–36.

Rohrer S, Holsten L, Weiss E, Benghezal M, *et al.* 2012. Multiple pathways of plasmid DNA transfer in *Helicobacter pylori*. *PLoS One* 7: e45623.

Rokita E, Makristathis A, Presterl E, Rotter ML, Hirschl AM. 1998. *Helicobacter pylori* urease significantly reduces opsonization by human complement. *J. Infect. Dis.* 178:1521–1525.

Rolig AS, Carter JE, Ottemann KM. 2011. Bacterial chemotaxis modulates host cell apoptosis to establish a T-helper cell, type 17 (Th17)-dominant immune response in *Helicobacter pylori* infection. *Proc. Natl. Acad. Sci. U. S. A.* 108: 19749–54.

Rollauer SE, Tarry MJ, Graham JE, Jääskeläinen M, *et al.* 2012. Structure of the TatC core of the twin-arginine protein transport system. *Nature* 492: 210–4.

Rosqvist R, Magnusson KE, Wolf-Watz H. 1994. Target cell contact triggers expression and polarized transfer of *Yersinia* YopE cytotoxin into mammalian cells. *EMBO J.* 13: 964–72.

Rossi M, Hänninen M-L. 2012. *Helicobacter* spp. other than *H. pylori*. *Helicobacter* 17: 56–61.

Rossier O, Starkenburg SR, Cianciotto NP. 2004. *Legionella pneumophila* type II protein secretion promotes virulence in the A/J mouse model of Legionnaires' disease pneumonia. *Infect. Immun.* 72: 310–321.

Rust M, Borchert S, Niehus E, Kuehne S a, *et al.* 2009. The *Helicobacter pylori* anti-sigma factor FlgM is predominantly cytoplasmic and cooperates with the flagellar basal body protein FlhA. *J. Bacteriol.* 191: 4824–34.

Ryan KA, Karim N, Worku M, Moore SA, *et al.* 2005a. HP0958 is an essential motility gene in *Helicobacter pylori*. *FEMS Microbiol. Lett.* 248: 47–55.

Ryan KA, Karim N, Worku M, Penn CW, O'Toole PW. 2005b. Hook-filament transition is controlled by a FliK functional homolog encoded by the gene HP0906. *J. Bacteriol.* 187: 5742–5750.

Saadat I, Higashi H, Obuse C, Umeda M, *et al.* 2007. *Helicobacter pylori* CagA targets PAR1/MARK kinase to disrupt epithelial cell polarity. *Nature* 447: 330–3.

Saez J, Belda S, Santibáñez M, Rodríguez JC, *et al.* 2012. Real-time PCR for diagnosing *Helicobacter pylori* infection in patients with upper gastrointestinal bleeding: comparison with other classical diagnostic methods. *J. Clin. Microbiol.* 50: 3233–7.

Saier MH. 2004. Evolution of bacterial type III protein secretion systems. *Trends Microbiol.* 12: 113–5.

Saitoa N, Konishia K, Satoa F, Katob M, *et al.* 2003. Plural transformation-processes from spiral to coccoid *Helicobacter pylori* and its viability. *J. Infect.* 46: 49–55.

Salacha R, Kovacić F, Brochier-Armanet C, Wilhelm S, *et al.* 2010. The *Pseudomonas aeruginosa* patatin-like protein PlpD is the archetype of a novel Type V secretion system. *Environ. Microbiol.* 12: 1498–512.

Salama NR, Hartung ML, Müller A. 2013. Life in the human stomach: persistence strategies of the bacterial pathogen *Helicobacter pylori*. *Nat. Rev. Microbiol.* 11: 385–99.

Salama NR, Otto G, Tompkins L, Falkow S. 2001. Vacuolating cytotoxin of *Helicobacter pylori* plays a role during colonization in a mouse model of infection. *Infect. Immun.* 69: 730–736.

Samatey FA, Imada K, Nagashima S, Vonderviszt F, *et al.* 2001. Structure of the bacterial flagellar protofilament and implications for a switch for supercoiling. *Nature* 410: 331–337.

Satin B, Norais N, Telford J, Rappuoli R, *et al.* 1997. Effect of *Helicobacter pylori* vacuolating toxin on maturation and extracellular release of procathepsin D and on epidermal growth factor degradation. *J. Biol. Chem.* 272: 25022–25028.

Sauvonnet N, Vignon G, Pugsley AP, Gounon P. 2000. Pilus formation and protein secretion by the same machinery in *Escherichia coli. EMBO J.* 19: 2221–8.

Schreiber S, Bucker R, Groll C, Azevedo-vethacke M, *et al.* 2005. Rapid Loss of Motility of *Helicobacter pylori* in the gastric lumen *in vivo*. *Infect. Immun.* 73: 1584–1589.

Schreiber S, Konradt M, Groll C, Scheid P, *et al.* 2004. The spatial orientation of *Helicobacter pylori* in the gastric mucus. *Proc. Natl. Acad. Sci. U. S. A.* 101: 5024–9.

Schwartz JT, Allen LH. 2006. Role of urease in megasome formation and *Helicobacter pylori* survival in macrophages. *J. Leuk. Biol.* 79: 1214-1255.

Segal ED, Cha J, Lo J, Falkow S, Tompkins LS. 1999. Altered states: involvement of phosphorylated CagA in the induction of host cellular growth changes by *Helicobacter pylori*. *Proc. Natl. Acad. Sci. U. S. A.* 96: 14559–64.

Selbach M, Moese S, Hauck CR, Meyer TF, Backert S. 2002. Src is the kinase of the *Helicobacter pylori* CagA protein *in vitro* and *in vivo*. *J. Biol. Chem.* 277: 6775–8.

Seyler RWJ, Olson JW, Maier RJ. 2001. Superoxide dismutase-deficient mutants of *Helicobacter pylori* are hypersensitive to oxidative stress and defective in host colonization. *Infect. Immun.* 69: 4034–4040.

Sheu S-M, Sheu B-S, Yang H-B, Lei H-Y, Wu J-J. 2007. Anti-Lewis X antibody promotes *Helicobacter pylori* adhesion to gastric epithelial cells. *Infect. Immun.* 75: 2661–7.

Shimada M, Ando T, Peek RM, Watanabe O, *et al.* 2008. *Helicobacter pylori* infection upregulates interleukin-18 production from gastric epithelial cells. *Eur. J. Gastroenterol. Hepatol.* 20: 1144–1150.

Shin S, Park C. 1995. Modulation of flagellar expression in *Escherichia coli* by acetyl phosphate and the osmoregulator OmpR. *J. Bacteriol*. 177: 4696–4702.

Silva DG, Tinoco EMB, Rocha GA, Rocha AMC, *et al.* 2010. *Helicobacter pylori* transiently in the mouth may participate in the transmission of infection. *Mem. Inst. Oswaldo Cruz* 105: 657–60.

Sitaraman R, Israel DA, Romero-Gallo J, Peek RM. 2012. Cell-associated hemolysis induced by *Helicobacter pylori* is mediated by phospholipases with mitogen-activated protein kinase-activating properties. *J. Clin. Microbiol.* 50: 1014–8.

Smet A, Flahou B, Mukhopadhya I, Ducatelle R, *et al.* 2011. The other *Helicobacter*s. *Helicobacter* 16: 70–5.

Smith TG, Pereira L, Hoover TR. 2009. *Helicobacter pylori* FlhB processing-deficient variants affect flagellar assembly but not flagellar gene expression. *Microbiology* 155: 1170–80.

Society AC. 2011. Global Cancer Facts & Figures 2nd Edition.

Soutourina O, Kolb A, Krin E, Danchin A, *et al.* 1999. Multiple control of flagellum biosynthesis in *Escherichia coli*: role of H-NS protein and the cyclic AMP-catabolite activator protein complex in transcription of the *flhDC* master operon. *J. Bacteriol.* 181: 7500-7508.

Spohn G, Scarlato V. 1999. Motility of *Helicobacter pylori* is coordinately regulated by the transcriptional activator FlgR , an NtrC homolog. *J. Bacteriol*. 181: 593–599.

Stafford GP, Ogi T, Hughes C. 2005. Binding and transcriptional activation of non-flagellar genes by the *Escherichia coli* flagellar master regulator $FlhD_2C_2$. *Microbiology* 151: 1779–1788.

Stanley NR, Findlay K, Berks BC, Palmer T. 2001. *Escherichia coli* strains blocked in Tat-dependent protein export exhibit pleiotropic defects in the cell envelope. *J. Bacteriol.* 183: 139–144.

Stanley SA, Raghavan S, Hwang WW, Cox JS. 2003. Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. *Proc. Natl. Acad. Sci. U. S. A.* 100: 13001–6.

Stevenson LG, Strisovsky K, Clemmer KM, Bhatt S, *et al.* 2007. Rhomboid protease AarA mediates quorum-sensing in *Providencia stuartii* by activating TatA of the twin-arginine translocase. *Proc. Natl. Acad. Sci. U. S. A.* 104: 1003-1008.

Stingl K, Müller S, Scheidgen-Kleyboldt G, Clausen M, Maier B. 2010. Composite system mediates two-step DNA uptake into *Helicobacter pylori*. *Proc. Natl. Acad. Sci. U. S. A.* 107: 1184–9.

Stingl K, Uhlemann E, Deckers-Hebestreit G, Schmid R, *et al.* 2001. Prolonged survival and cytoplasmic pH homeostasis of *Helicobacter pylori* at pH 1. *Infect. Immun.* 69: 1178–1181.

Stolte M, Kroher G, Meining A, Morgner A, *et al.* 997. A comparison of *Helicobacter pylori* and *H. heilmannii* gastritis. A matched control study involving 404 patients. *Scand. J. Gastroenterol*. 32: 28–33.

Suerbaum S, Josenhans C, Labigne A. 1993. Cloning and genetic characterization of the *Helicobacter pylori* and *Helicobacter mustelae flaB* flagellin genes and construction of *H. pylori flaA*- and *flaB*-negative mutants by electroporation-mediated allelic exchange. *J. Bacteriol.* 175: 3278–3288.

Suerbaum S, Michetti P. 2002. *Helicobacter pylori* infection. *N. Engl. J. Med.* 347: 1175–1186.

Sundrud MS, Torres VJ, Unutmaz D, Cover TL. 2004. Inhibition of primary human T cell proliferation by *Helicobacter pylori* vacuolating toxin (VacA) is independent of VacA effects on IL-2 secretion. *Proc. Natl. Acad. Sci. U. S. A.* 101: 7727–32.

Sycuro LK, Pincus Z, Gutierrez KD, Biboy J, *et al.* 2010. Peptidoglycan crosslinking relaxation promotes *Helicobacter pylori*'s helical shape and stomach colonization. *Cell* 141: 822–33.

Sycuro LK, Wyckoff TJ, Biboy J, Born P, *et al.* 2012. Multiple peptidoglycan modification networks modulate *Helicobacter pylori*'s cell shape, motility, and colonization potential. *PLoS Pathog.* 8: e1002603.

Tabassam FH, Graham DY, Yamaoka Y. 2008. OipA plays a role in *Helicobacter pylori*-induced focal adhesion kinase activation and cytoskeletal re-organization. *Cell. Microbiol.* 10: 1008–20.

Tarry MJ, Schafer E, Chen S, Buchanan G, *et al.* 2009. Structural analysis of substrate binding by the TatBC component of the twin-arginine protein transport system. *Proc. Natl. Acad. Sci. U. S. A* 106: 13284–13289.

Telford BJL, Ghiara P, Orco MD, Comanducci M, *et al.* 1994. Gene structure of the *Helicobacter pylori* cytotoxin and evidence of its key role in gastric disease. *J. Exp. Med.* 179: 1653–1658.

Terry K, Williams SM, Connolly L, Ottemann KM. 2005. Chemotaxis plays multiple roles during *Helicobacter pylori* animal infection. *Infect. Immun.* 73: 803–811.

The Eurogast Study Group. 1993. Epidemiology of, and risk factors for, *Helicobacter pylori* infection among 3194 asymptomatic subjects in 17 populations. *Gut* 34: 1672–6.

Thomas D, Morgan DG, DeRosier DJ. 1999. Rotational symmetry of the C ring and a mechanism for the flagellar rotary motor. *Proc. Natl. Acad. Sci. U. S. A.* 96: 10134–10139.

Thomas J, Stafford GP, Hughes C. 2004. Docking of cytosolic chaperone-substrate complexes at the membrane ATPase during flagellar type III protein export. *Proc. Natl. Acad. Sci. U. S. A.* 101: 3945–50.

Tian X-Y, Zhu H, Zhao J, She Q, Zhang G-X. 2012. Diagnostic Performance of urea breath test, rapid urea test, and histology for *Helicobacter pylori* infection in patients with partial gastrectomy: a meta-analysis. *J. Clin. Gastroenterol.* 46: 285–292.

Toller IM, Neelsen KJ, Steger M, Hartung ML, *et al.* 2011. Carcinogenic bacterial pathogen *Helicobacter pylori* triggers DNA double-strand breaks and a DNA damage response in its host cells. *Proc. Natl. Acad. Sci. U. S. A.* 108: 14944–9.

Tomb JF, White O, Kerlavage AR, Clayton RA, *et al.* 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388: 539–47.

Tombola F, Carlesso C, Szabò I, Bernard M de, *et al.* 1999. *Helicobacter pylori* vacuolating toxin forms anion-selective channels in planar lipid bilayers: possible implications for the mechanism of cellular vacuolation. *Biophys. J.* 76: 1401–9.

Tomita T, Jackson a M, Hida N, Hayat M, *et al.* 2001. Expression of Interleukin-18, a Th1 cytokine, in human gastric mucosa is increased in *Helicobacter pylori* infection. *J. Infect. Dis.* 183: 620–7.

Torres VJ, Ivie SE, McClain MS, Cover TL. 2005. Functional properties of the p33 and p55 domains of the *Helicobacter pylori* vacuolating cytotoxin. *J. Biol. Chem.* 280: 21107–14.

Trebesius K, Adler K, Vieth M, Stolte M, Haas R. 2001. Specific detection and prevalence of *Helicobacter heilmannii*-like organisms in the human gastric mucosa by fluorescent *in situ* hybridization and partial 16S ribosomal DNA sequencing. *J. Clin. Microbiol.* 39: 1510–1516.

Troisfontaines P, Cornelis GR. 2005. Type III secretion: more systems than you think. *Physiology* 20: 326–339.

Tseng H-J, Chan C-C, Chan E-C. 2004. Sphingomyelinase of *Helicobacter pylori*-induced cytotoxicity in AGS gastric epithelial cells via activation of JNK kinase. *Biochem. Biophys. Res. Commun.* 314: 513–518.

Tseng T-T, Tyler BM, Setubal JC. 2009. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol.* 9: S2.

Turvey SE, Leo SH, Boos A, Deans GD, *et al.* 2012. Successful approach to treatment of *Helicobacter* bilis infection in X-linked agammaglobulinemia. *J. Clin. Immunol.* 32: 1404–8.

Twing KI, Kirchman DL, Campbell BJ. 2011. Temporal study of *Helicobacter pylori* presence in coastal freshwater, estuary and marine waters. *Water Res.* 45: 1897–905.

Uchiyama J, Takeuchi H, Kato S, Takemura-Uchiyama I, *et al.* 2012. Complete Genome Sequences of Two *Helicobacter pylori* Bacteriophages Isolated from Japanese Patients. *J. Virol.* 86: 11400–11401.

Uemura N, Okamoto S, Yamamoto S, Matsumura N, *et al.* 2001. *Helicobacter pylori* infection and the development of gastric cancer. *N. Engl. J. Med.* 345: 784–789.

Ueno T, Oosawa K, Aizawa S-I. 1994. Domain structures of the MS ring component protein (FliF) of the flagellar basal body of *Salmonella typhimurium*. *J. Mol. Biol*. 236: 548–555.

Unemo M, Aspholm-Hurtig M, Ilver D, Bergström J, *et al.* 2005. The sialic acid binding SabA adhesin of *Helicobacter pylori* is essential for nonopsonic activation of human neutrophils. *J. Biol. Chem.* 280: 15390–7.

Viala J, Chaput C, Boneca IG, Cardona A, *et al.* 2004. Nod1 responds to peptidoglycan delivered by the *Helicobacter pylori* cag pathogenicity island. *Nat. Immunol.* 5: 1166–74.

Voulhoux Â, Ball Á, Lazdunski Â, Wu L, *et al.* 2001. Involvement of the twin-arginine translocation system in protein secretion via the type II pathway. *Eur. Mol. Biol. Organ.* 20: 6735–6741.

Wallden K, Rivera-Calzada A, Waksman G. 2010. Type IV secretion systems: versatility and diversity in function. *Cell. Microbiol.* 12: 1203–12.

Wang G, Hong Y, Olczak A, Maier SE, Maier RJ. 2006. Dual roles of *Helicobacter pylori* NapA in inducing and combating oxidative stress. *Infect. Immun.* 74: 6839–46.

Ward JM, Anver MR, Haines DC, Benveniste RE. 1994. Chronic active hepatitis in mice caused by *Helicobacter* hepaticus. *Am. J. Pathol.* 145: 959–68.

Warren JR, Marshall BJ. 1983. Unidentified curved *bacilli* on gastric epithelium in active chronic gastritis. *Lancet* 321: 1273–1275.

Weeks DL. 2000. A H+-gated urea channel: the link between *Helicobacter pylori* urease and gastric colonization. *Science* 287: 482–485.

Wei BL, Brun-Zinkernagel a M, Simecka JW, Prüss BM, *et al.* 2001. Positive regulation of motility and *flhDC* expression by the RNA-binding protein CsrA of *Escherichia coli*. *Mol. Microbiol*. 40: 245–56.

WGO E. 2010. World Gastroenterology Organisation Global Guidelines: *Helicobacter pylori* in developing countries. 1–15.

Willén R, Carlén B, Wang X, Papadogiannakis N, *et al.* 2000. Morphologic conversion of *Helicobacter pylori* from spiral to coccoid form. Scanning (SEM) and transmission electron microscopy (TEM) suggest viability. *Ups. J. Med. Sci.* 105: 31–40.

Willhite DC, Blanke SR. 2004. *Helicobacter pylori* vacuolating cytotoxin enters cells, localizes to the mitochondria, and induces mitochondrial membrane permeability changes correlated to toxin channel activity. *Cell. Microbiol.* 6: 143–154.

Williams AW, Yamaguchi S, Togashi F, Aizawa SI, *et al.* 1996. Mutations in *fliK* and *flhB* affecting flagellar hook and filament assembly in *Salmonella typhimurium*. *J. Bacteriol*. 178: 2960–70.

Wirth H-P, Yang M, Peek RM, Tham KT, Blaser MJ. 1997. *Helicobacter pylori* Lewis Expression Is Related to the Host Lewis Phenotype. *Gastroenterology* 113: 1091–1098.

Woestyn S, Allaoui A, Wattiau P, Cornelis GR. 1994. YscN, the putative energizer of the Yersinia Yop secretion machinery. *J. Bacteriol.* 176: 1561–9.

Wroblewski LE, Asim M, Chaturvedi R, Schumacher M, *et al.* 2013. Use of a novel ex-vivo three-dimensional system to define host-microbial interactions with carcinogenic potential. 17th *Int. Work. Campylobacter, Helicobacter Relat. Org.* 8.

Wroblewski LE, Peek RM. 2011. "Targeted disruption of the epithelial-barrier by *Helicobacter pylori*". *Cell Commun. Signal.* 9: 1-11.

Wroblewski LE, Peek RM, Wilson KT. 2010. *Helicobacter pylori* and gastric cancer: factors that modulate disease risk. *Clin. Microbiol. Rev.* 23: 713–39.

Wroblewski LE, Shen L, Ogden S, Romero-Gallo J, *et al.* 2009. *Helicobacter pylori* dysregulation of gastric epithelial tight junctions by urease-mediated myosin II activation. *Gastroenterology* 136: 236–246.

Xiang Z, Censini S, Bayeli PF, Telford JL, *et al.* 1995. Analysis of expression of CagA and VacA virulence factors in 43 strains of *Helicobacter pylori* reveals that clinical isolates can be divided into two major types and that CagA is not necessary for expression of the vacuolating cytotoxin. *Infect. Immun*. 63: 94–8.

Xu Q, Morgan RD, Roberts RJ, Blaser MJ. 2000. Identification of type II restriction and modification systems in *Helicobacter pylori* reveals their substantial diversity among strains. *Proc. Natl. Acad. Sci. U. S. A.* 97: 9671–6.

Yahara K, Kawai M, Furuta Y, Takahashi N, *et al.* 2012. Genome-wide survey of mutual homologous recombination in a highly sexual bacterial species. *Genome Biol. Evol.* 4: 628–640.

Yamaguchi S, Aizawa S, Kihara M, Isomura M, Jones CJ. 1986. Genetic evidence for a switching and energy-transducing complex in the flagellar motor of *Salmonella typhimurium*. *J. Bacteriol.* 168: 1172-1179.

Yamamoto S, Kutsukake K. 2006. FliT acts as an anti-FlhD2C2 factor in the transcriptional control of the flagellar regulon in *Salmonella enterica* serovar *typhimurium*. *J. Bacteriol.* 188: 6703–8.

Yamaoka Y, Kwon DH, Graham DY. 2000. A M(r) 34,000 proinflammatory outer membrane protein (oipA) of *Helicobacter pylori*. *Proc. Natl. Acad. Sci. U. S. A.* 97: 7533–8.

Ye F, Brauer T, Niehus E, Drlica K, *et al.* 2007. Flagellar and global gene regulation in *Helicobacter pylori* modulated by changes in DNA supercoiling. *Int. J. Med. Microbiol.* 297: 65–81.

Yokoseki T, Iino T, Kutsukake K. 1996. Negative regulation by *fliD*, *fliS*, and *fliT* of the export of the flagellum-specific anti-sigma factor, FlgM , in *Salmonella typhimurium*. *J. Bacteriol*. 178: 899–901.

Yonekura K. 2000. The bacterial flagellar cap as the rotary promoter of flagellin self-assembly. *Science* 290: 2148–2152.

Yoshiyama H, Nakazawa T. 2000. Unique mechanism of *Helicobacter pylori* for colonizing the gastric mucus. *Microbes Infect.* 2: 55–60.

Young GM, Schmiel DH, Miller VL. 1999. A new pathway for the secretion of virulence factors by bacteria: the flagellar export apparatus functions as a protein-secretion system. *Proc. Natl. Acad. Sci. U. S. A.* 96: 6456–61.

Zarivach R, Vuckovic M, Deng W, Finlay BB, Strynadka NCJ. 2007. Structural analysis of a prototypical ATPase from the type III secretion system. *Nat. Struct. Mol. Biol.* 14: 131–7.

Zhang Z-W, Dorrell N, Wren BW, Farthingt MJG. 2002. *Helicobacter pylori* adherence to gastric epithelial cells: a role for non-adhesin virulence genes. *J. Med. Microbiol.* 51: 495–502.

Zhu K, González-Pedrajo B, Macnab RM. 2002. Interactions among membrane and soluble components of the flagellar export apparatus of *Salmonella*. *Biochemistry* 41: 9516–24.

Zou Q-H, Li R-Q. 2011. *Helicobacter pylori* in the oral cavity and gastric mucosa: a meta-analysis. *J. Oral Pathol. Med.* 40: 317–24.

Chapter 2
Genome Sequences and Comparative Genomics of Two Motile Helicobacter pylori Strains

Note:

- Some of the following sections contain work that has been published as

Ceara D. Clancy, Brian M. Forde, Stanley A. Moore and Paul W. O'Toole: Draft Genome Sequences of *Helicobacter pylori* Strains 17874 and P79. *J Bacteriol.* 2012; 194:2402.

- Dr. Brian Forde performed genome assembly described in Section 2 of the genome announcement and Section 3.3 of the expanded thesis analysis.

Published Genome Announcement:

Draft Genome Sequences of Helicobacter pylori Strains 17874 and P79

1 Abstract

Helicobacter pylori is a human pathogen which colonises the human gastric mucosa, causing gastritis, duodenal and gastric ulcers, and gastric carcinoma. Here we announce the draft genomes of *H. pylori* strain 17874, commonly used for studying motility, and P79, a strain for which plasmid vectors have been developed.

2 Genome Announcement

H. pylori genomes sequenced to date exhibit significant variation. *H. pylori* CCUG 17874 was originally isolated from the gastric antrum of a patient in Perth, Australia and is the type strain for the species (4) that is often used for flagellum biogenesis studies. P79 is a derivative of strain P1, transformed with 17874 chromosomal DNA to generate a streptomycin resistant mutant (3). This readily-transformable strain facilitates *in vivo* studies on *H. pylori*. The genomes of these strains were sequenced to provide a clearer genomic platform for *H. pylori* motility investigation.

The *H. pylori* 17874 and P79 genomes were sequenced at the Beijing Genomics Institute (BGI) on the Illumina HiSeq platform, generating a paired-end library containing 20,154,284 and 13,298,804 reads of 90 bp, respectively. In a reference-guided assembly strategy using MIRA (version 3.2.1), reads for both genomes were mapped to the genomes of *H. pylori* 26695 (GenBank acc. NC_000915) (5) and J99 (NC_000921.1) (1). A *de novo* assembly using Velvet was also performed and aligned to the MIRA assembly to close gaps. 17874 and P79 contigs were assembled into 80 and 48 scaffolds, respectively. Protein coding regions were identified using the NCBI Prokaryotic Genome Automated Annotation Pipeline (PGAAP) and manually curated, with particular interest in flagellum-related genes. Predicted coding regions were identified with a minimum cut-off size of 30 amino acids.

H. pylori 17874 and P79 have genome sizes of 1,615,763 bp and of 1,641,495 bp, respectively and GC content of 38.97% and 38.86%, respectively. Both strains are *cagA*+ and *vacA*+, well described virulence factors (2). Strain-unique genes were identified using a pairwise bi-directional BLASTP comparison, where the query sequence has no detectable homologues. The 17874 genome contains 1,639 open reading frames, with 35, 45 and 24 unique genes that are absent in 26695, J99 and

P79, respectively. Sixteen genes from 26695 and 6 genes from J99 are absent in 17874. *H. pylori* P79 contains 1,699 open reading frames, with 40, 52 and 36 unique genes that are absent in 26695, J99 and 17874, respectively. Twelve genes from 26695 and 6 genes from J99 are absent in P79. Twenty one genes are unique to the 17874 and P79 genomes compared across these four strains.

The majority of strain-unique genes identified encode hypothetical protein products. Of note, 17874 possesses a unique type II restriction enzyme, and P79 possesses a unique hypothetical membrane protein that is absent in 26695/J99. 17874 and P79 lack metal-binding proteins present in both 26695 and J99, but possess Cag island protein B. All major flagellar and outer membrane proteins are present and intact in both 17874 and P79 compared to 26695 and J99. A hypothetical protein with predicted involvement in ATPase activity during flagellum biogenesis is absent in P79.

3 Nucleotide Sequence Accession Numbers

The draft genome sequence of *H. pylori* 17874 has been deposited in GenBank, available through the BioProject accession number PRJNA76569 and project ID 76569. Similarly, the draft sequence of P79 is available in GenBank through the BioProject accession number PRJNA76567 and project ID 76567.

4 Acknowledgements

This project was supported by a Research Frontiers Programme award (09_RFP_GEN2443) from Science Foundation Ireland to PWOT, and by an Embark scholarship from IRCSET to CDC.

5 References

1. **Alm RA, Ling L-SL, Moir DT, King BL, Brown ED, Doig PC, Smith DR, Noonan B, Guild BC, deJonge BL, Carmel G, Tummino PJ, Caruso A, Uria-Nickelsen M, Mills DM, Ives C, Gibson R, Merberg D, Mills SD, Jiang Q, *et al.*** 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. Nature 397:176-180.
2. **Basso D, Plebani M, Kusters JG.** 2010. Pathogenesis of *Helicobacter pylori* Infection. Helicobacter **15**:14-20.
3. **Heuermann D, Haas R.** 1998. A stable shuttle vector system for efficient genetic complementation of *Helicobacter pylori* strains by transformation and conjugation. Mol. Gen. Genet. **257**:519-528.
4. **Marshall BJ, Royce H, Annear D.** 1984. Original isolation of *Campylobacter pyloris* from human gastric mucosa. Microbios. Lett. **25**:83-88.
5. **Tomb J-F, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA, Nelson K, Quackenbush J, Zhou L, Kirkness EF, Peterson S, Loftus B, Richardson D, Dodson R, Khalak HG, Glodek A, *et al.*** 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature 388:539-547.

Expanded Thesis Analysis:

Comparative Genomics of Two Motile Helicobacter pylori Strains and Core Genome Analysis of 43 Strains

1 Abstract

H. pylori was the first species for which two genomes were sequenced (Alm *et al.*, 1999; Tomb *et al.*, 1997). This revealed a large degree of synteny in overall genome organisation between the strains 26695 and J99. However, *H. pylori* genomes possess regions of hypervariability *e.g.* cytotoxin-associated gene pathogenicity island (*cag*PAI) and plasticity zones (Alm and Trust, 1999).

The draft genome sequence of *Helicobacter pylori* CCUG 17874, the type-strain for the species, comprises 80 scaffolds containing 1.61 Mbp of sequence with a GC content of 38.97%. *In silico* analysis identified 1,639 coding genes, including *vacA*, *cagA*, 3 IS elements and 28 pseudogenes. This motile strain is widely used in flagellum biogenesis studies. P79 is a readily-transformable derivative of strain P1, which facilitates genetic manipulation of *H. pylori* cells. A draft assembly of P79, comprising 48 scaffolds, contains 1.64 Mbp of sequence with a GC content of 38.86%. Similarly, *in silico* analysis identified 1,699 coding genes including *vacA*, *cagA*, 5 IS elements and 33 pseudogenes.

Comparative analysis of these two strains revealed that both possess the full complement of flagellar genes. *H. pylori* CCUG 17874 possesses 35, 45 and 24 unique genes that are absent in 26695, J99 and P79, respectively. *H. pylori* P79 possesses 40, 52 and 36 unique genes that are absent in 26695, J99 and 17874, respectively. The core genome of *H. pylori* comprises 898 genes, based on analysis of 43 sequenced strains, including 17874 and P79. Core genomes were also identified for the following disease subtypes: gastritis, duodenal ulcer, gastric cancer and MALT lymphoma.

This analysis provides sequence information for these useful lab strains, and insights into the genetic organisation of *H. pylori*. As a result, a more conservative core genome for the species has now been determined. The genomes of these strains provide a clearer genomic platform for *H. pylori* motility investigation.

2 Introduction

Helicobacter pylori is a pathogen which colonises the human gastric mucosa, causing gastritis, duodenal and gastric ulcers, and gastric carcinoma (Blaser, 1997; Goodwin *et al.*, 1986; Uemura *et al.*, 2001). Motility is an essential feature for colonisation as it enables *H. pylori* to move from the lumen of the stomach, through the mucosal lining where it can interact with host epithelial cells (Algood and Cover, 2006; Eaton *et al.*, 1992). *H. pylori* motility requires the presence of 2 - 6 polar, sheathed flagella (Eaton *et al.*, 1992; Yoshiyama and Nakazawa, 2000). Flagellum biogenesis in *H. pylori* is a complex, hierarchical process which differs from other model organisms such as *Salmonella* and *E. coli* (Lertsethtakarn *et al.*, 2011; Macnab, 2003). The genome of *H. pylori* strain 26695 was the first to be sequenced for the species. The genome features of this strain have been well described and used in subsequent comparative studies (Alm *et al.*, 1999; Tomb *et al.*, 1997). However, this non-motile strain is not appropriate for motility studies because of a frameshift in the *flip* gene (Josenhans *et al.*, 2000). The type-strain, CCUG 17874, has been extensively used to investigate motility, but the genome had not been sequenced.

H. pylori was the first species for which the genome of more than one strain was sequenced (Alm *et al.*, 1999; Tomb *et al.*, 1997). Comparative analysis revealed that *H. pylori* genomes exhibit significant variation in defined regions of hypervariability, while retaining synteny in the overall genome organisation. The striking genome plasticity of this gastric pathogen coupled with the variety of clinical outcomes which can arise from infection have served as the impetus for genome mining studies to identify strain-specific virulence factors and genetic markers for disease. At the time of writing, the genomes of 52 *H. pylori* strains had been fully sequenced, annotated and are publically available through the NCBI web resource (NCBI, 2013). Additionally, the draft genome sequences of a further 228 strains are available, though some lack annotation.

The first live organisms to have their genomes sequenced were *Haemophilus influenzae* (Fleischmann *et al.*, 1995) and *Mycobacterium genitalium* (Fraser *et al.*, 1995) at the J. Craig Venter Institute (formerly the Institute of Genomic Research (TIGR)). These were shortly followed by the genome sequence of other organisms including that of *E. coli* (Blattner, 1997), and later *Salmonella enterica* serovar Typhimurium (McClelland *et al.*, 2001), the model

organisms for bacterial motility studies. Sequencing methods have developed considerably from the initial chain termination sequencing to next generation, and now third generation sequencing technologies. Next generation sequencing (NGS) can be subdivided into single nucleotide addition (pyrosequencing), cyclic reversible termination (Illumina) and sequencing by ligation (Applied Biosystems SOLiD). Each of these methods involves random fragmentation of genomic DNA and hybridisation to adapters followed by sequencing by either non-fluorescent/fluorescent means (Metzker, 2010). These methodologies have limitations including slow processing time due to the large number of sequencing cycles per run, amplification of errors in the PCR-based sequencing and short read lengths. Nevertheless, NGS allows the generation of a large volume of sequence data cheaply. Third generation sequencing methods including single real time (Pacific Bioscience) (Eid *et al.*, 2009) and nanopore (Oxford) sequencing have recently been developed. These eliminate amplification bias problems, and generate longer read lengths, thus improving the quality of genome sequencing for the future (Koren *et al.*, 2013).

As genome sequencing has become increasingly affordable and rapid in recent years, this has lead to an increase in the volume of genome information available for mining (Horner *et al.*, 2010; Loman *et al.*, 2012). A whole-genome shotgun sequencing approach is a powerful strategy to sequence and assemble whole genome data both rapidly and cheaply. Illumina sequencing followed by mapping to a reference genome is a reliable means for genome analysis and comparative studies. Furthermore, comparison with a *de novo* assembly of the same reads improves resolution of genome assembly.

H. pylori CCUG 17874 was originally isolated from the gastric antrum of a patient in Perth, Australia and is the type-strain for the species (Marshall *et al.*, 1984). P79 is a derivative of strain P1 (isolated from a patient with non-ulcer dyspepsia), transformed with 17874 derivative chromosomal DNA to generate a streptomycin-resistant mutant (Heuermann and Haas, 1998). This readily-transformable strain facilitates genetic studies of *H. pylori*. The genomes of these strains were sequenced to enhance *H. pylori* motility investigation and contribute to our understanding of the genomic organisation of this pathogen. Comparative analyses with other sequenced strains of *H. pylori* provided an updated core genome for the species and disease-associated subtypes.

3 Methods

3.1 Bacterial Strains and Culture Conditions

H. pylori strains CCUG 17874 and P79 were grown on Columbia Base Agar (CBA) solid medium (Oxoid, UK), supplemented with 5% v/v heat-inactivated, defibrinated horse blood (Cruinn, Ireland). Plates were incubated at 37°C, 5% CO_2 and sub-cultured every two days.

3.2 Genomic DNA Extraction

Cells were harvested from 2 day old full plates in sterile phosphate buffered saline (PBS). Genomic DNA was extracted using DNeasy Blood and Tissue DNA Extraction Kit (Qiagen, Germany) according to manufacturer's instructions. Briefly, cells were lysed and treated with proteinase K and RNase A. DNA was then bound to a silica column and washed before elution. DNA concentration and quality was estimated using Nanodrop 2000 (Thermo Scientific). Genomic DNA was run on a 1% agarose gel in TAE buffer at 90 V for 30 min to confirm quality.

3.3 Genome Sequencing and Annotation

The genomes of *H. pylori* CCUG 17874 and *H. pylori* P79 were sequenced on the Illumina HiSeq platform (Beijing Genomics Institute, China). A paired-end library was generated and sequenced containing 20,154,284 and 13,298,804 reads of 90 bp for the genomes of CCUG 17874 and P79, respectively. In a reference-guided assembly strategy using MIRA (version 3.2.1) (Chevreux *et al.*, 1999), reads for both genomes were mapped to the genomes of *H. pylori* 26695 (GenBank acc. NC_000915) and J99 (NC_000921.1) (Alm *et al.*, 1999; Tomb *et al.*, 1997). A *de novo* assembly using Velvet was also performed for each genome and aligned to the MIRA assembly to close gaps. *H. pylori* CCUG 17874 and P79 contigs were assembled into 80 and 48 scaffolds, respectively.

Automated gene calling and annotation were performed by the NCBI Prokaryotic Genome Automated Annotation Pipeline (PGAAP). Open reading frames (ORFs) were predicted by Genemark searches within the manually curated Protein Clusters database. Reverse PSI-BLAST (RPS-BLAST) was performed against the Clusters of Orthologous Groups (COGs) database to assign COG fuctional categories to the predicted ORFs. Additionally, InterProScan was used to identify protein domains

and signatures (Quevillon *et al.*, 2005). Frameshifts and partial gene fragments indicating potential pseudogenes were identified by alignment of proteins from the target set to the genome with ProSplign (a global alignment algorithm) and then checked with GeneMarkS. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRS) were identified by searching the CRISPR database (Grissa *et al.*, 2007). tRNAs were identified using tRNA-scan (Lowe and Eddy, 1997) and ribosomal binding sites using RBSfinder (Suzek *et al.*, 2001).

Identified protein coding regions were manually curated in Artemis (Rutherford *et al.*, 2000), with particular interest in flagellum-related genes. Predicted ORFs had a minimum cut-off size of 30 amino acids. Each of the draft genomes were submitted to the KAAS (KEGG Automatic Annotation Server) online server which automatically assigns K numbers to genes, enabling reconstruction of KEGG pathways and BRITE hierarchies using a bi-directional best hit BLAST approach (Moriya *et al.*, 2007).

Accession numbers: The draft genome of *H. pylori* CCUG 17874 is available under the BioProject accession number PRJNA76569. Similarly, the draft genome sequence of P79 is available under accession number PRJNA76567.

3.4 Genome Comparisons

Whole genome alignments were generated using Big Blast software (available from the Welcome Trust Institute (http://www.sanger.ac.uk)) to compare the genomes. Artemis Comparison Tool (ACT) was used to visualise the alignments (Carver *et al.*, 2005). Nucmer, part of the MUMmer software package, was used to generate further alignments (Kurtz *et al.*, 2004). METAPHORE is a custom in-house software which was used to identify orthologues, unique genes and core genes by performing bi-directional BLASTP comparisons on two or more genomes (Van der Veen *et al.*, 2014). Proteins with minimum 30% identity over 80% of their sequence length were classed as orthologues. Core genes were defined as those present in all possible pairwise genome combinations tested. Unique genes were defined as genes with no detectable homologues in bidirectional BLASTP comparisons. Phylogenetic trees were built based on concatenated MLST (multi locus sequence typing) analysis of 7 housekeeping genes that are distributed throughout the genome (Jolley and Maiden, 2010).

4 Results and Discussion

4.1 General Genome Features

The draft genome sequences of two *H. pylori* strains, CCUG 17874 and P79, are described in the following sections. Due to the draft nature of these sequences, information generated regarding pseudogenes, unique genes and specific gene numbers are estimates limited by the coverage of the genome sequences in this project. In order to ascertain absolute numbers, gap closure of the draft sequences would be required. For the purposes of this thesis, the sections ahead will refer to details generated from the draft genome sequences.

H. pylori CCUG 17874 and P79 have assembled genome sizes of 1,615,763 bp and of 1,641,495 bp, respectively and GC content of 38.97% and 38.86%, respectively (Table 2). Bioinformatic analysis of the *H. pylori* CCUG 17874 genome identified 1,639 coding regions with a coding density of 86.5% and an average gene length of 853 bp (Figure 15). In *H. pylori* P79, 1,699 coding regions were identified, representing a coding density of 85.1% and an average gene length of 812 bp (Figure 16). Gene synteny was largely conserved between 17874/P79 and reference strain 26695 (Figure 17). Biological functions could be assigned to 1,114 (67.9%) of the predicted proteins of *H. pylori* CCUG 17874. Of the remaining 525 predicted hypothetical proteins, 182 had COG functional categories assigned. Similarly, biological functions could be assigned to 1,079 (63.5%) of the predicted proteins of *H. pylori* P79. Of the remaining 620 predicted hypothetical proteins, 218 had COG functional categories assigned. The rest were either homologous to hypothetical proteins in other species or had no match to any known proteins, and hence were classified as unique proteins. Phylogenetic analysis based on MLST core genes predicted that 17874 clusters with the European strains, where it's predicted most closely related, sequenced strain is G27 (Figure 18). P79 also clusters with the European strains, closest to 26695 and P12.

Thirty six tRNA genes were identified in the genome of *H. pylori* CCUG 17874, while 35 were identified in the genome of P79. In both cases, the genes represent all 20 amino acids (redundant genes were present for 8 tRNAs in 17874, and 9 tRNAs in P79). In both genomes, 22 of these tRNAs were located on the lagging strand, most of which cluster near the 23S rRNA gene.

The genome of *H. pylori* CCUG 17874 contains 25 predicted pseudogenes (1.5% of coding sequences) (Table 2), generally as a result of homopolynucleotide mutations which cause in-sequence frame shifts. The *H pylori* P79 genome contains 29 predicted pseudogenes (1.7% of coding sequences). These predicted pseudogenes include recombinase A, genes in the plasticity regions of the strains and components of the restriction modification systems. Three transposases were indentified in the genome of *H. pylori* CCUG 17874, all part of the IS605 family. Four transposases were identified in the genome of P79, from IS605, IS606 and PS3IS, as well as an IS200 from *H. pylori* SARA17 (Table 3). The genomes of both strains were also found to harbour phage-associated genes (Table 4).

4.2 Plasticity Zones

There are 5 and 3 regions with deviating GC content in *H. pylori* CCUG 17874 and P79, respectively. The genes encoding *cagA* and *vacA* are both located in these low GC regions ("plasticity zones"), as well as many of the strain-specific genes of *H. pylori* (Alm and Trust, 1999; Boneca *et al.*, 2003). The plasticity zones of the 26695 and J99 genomes are flanked by the *ftsZ* gene and the rRNA 5S/23S subunit genes (Alm *et al.*, 1999; Tomb *et al.*, 1997). Other strains have been identified with three plasticity regions *e.g.* P12 which have since been identified as transposable elements (Kersulyte *et al.*, 2009). Plasticity regions include large genomic islands containing genes acquired by horizontal gene transfer, whereas PAIs are plasticity regions which encode virulence factors which contribute to the pathogenicity of the strain. Recent analysis of previously sequenced strains including 26695 have identified novel PAIs such as the *tfs3*PAI (Wang *et al.*, 2013). The availability of sequence data for a large number of *H. pylori* strains will enable a better understanding of these hypervariable regions and their potential for uncovering novel disease markers.

The fifth concentric circle in the genome atlas of 17874 shows 3 regions where the GC content is below the whole-genome average (Figure 15). The first region, which occurs near the origin, contains genes encoding transposases A and B, hypothetical proteins, RM system components and replicase A. The second region contains the *cag*PAI which includes genes encoding a T4SS apparatus which assemble to facilitate secretion of the CagA effector protein. The third region of low % GC includes genes encoding hypothetical proteins flanking a competence-like

protein, RM genes and an integrase. Other genes with low GC content which occur outside of these regions include ABC-type multidrug resistance genes, hypothetical genes and an inactivated helicase.

In the P79 genome, there are two plasticity zones: left and right. Plasticity zone left has a GC content of 33.3%, while plasticity zone right has a GC content of 32.66%, both lower than the rest of the genome (Figure 16). Genes present in plasticity zone left include those encoding hypothetical proteins, phage/colicin/tellurite resistance cluster terY, transposases A and B, topoisomerase and helicase. The *ftsZ* gene flanks plasticity zone right of P79, containing genesof the *cag*PAI. Similar to 17874, low GC content genes which occur outside of these two plasticity zones encode hypothetical proteins, multi-drug resistance proteins and RM system components.

4.3 Motility Genes, Virulence Factors and OMPs

All of the major regulatory and structural components required for flagellum biogenesis are present in the genomes of *H. pylori* CCGUG 17874 and P79 with reference to 26695 and J99 (Table 5). While the flagellar genes are not organised into discrete operons as is the case for *Salmonella*, the gene order is largely preserved across strains (Figure 19). The gene for a hypothetical protein with predicted involvement in ATPase activity during archaeal flagellum biogenesis is absent in P79. Both 17874 and P79 are motile strains of *H. pylori* with fully functional flagella, which is supported by the presence of the flagellar gene complement essential for motility (Figure 20).

Cytotoxin-associated gene A (CagA) and vacuolating cytotoxin (VacA) are two key virulence factors involved in *H. pylori* pathogenesis (Basso *et al.*, 2010). CCUG 17874 and P79 are both *cagA* and *vacA* positive and encode homologues of the virulence factor mviN protein (HP17_03394 and HP79_02579) (Table 6). In addition to these, P79 contains two virulence genes not present in 17874, both of which are annotated as encoding virulence associated protein D (VapD): HP79_08912 and HP79_08333. Both strains are urease and catalase positive, key virulence factors that enhance colonisation and infection of the host. HP17_07827 encodes a labile enterotoxin product which is absent in the genome of P79. Both strains also express a number of multidrug-resistance proteins enabling bacterial

survival in the presence of compounds such as methicillin and tetracycline in the case of P79.

CagA is a potent oncoprotein which induces inflammation and is associated with the more severe clinical outcomes of *H. pylori* infection including gastric cancer and mucosa-associated lymphoid tissue (MALT) lymphoma (Murata-Kamiya, 2011; Ohnishi *et al.*, 2008; Wroblewski and Peek, 2011). The *cag*PAI of 17874 contains 30 genes with an average GC content which is lower than is typical for the *H. pylori* genome (Figure 15). The gene encoding Cag15 is absent from the genome of 17874, while it contains two genes encoding hypothetical proteins which are absent in P79 (Table 6). The function of Cag15 is unknown. However, Cag15 contains putative transmembrane domains (Joyce *et al.*, 2001). The *cag7* gene is identified as a pseudogene in the genome of P79 and *cag3* is potentially inactive due to a frameshift mutation. Cag3 is hypothesised to be a novel secreted effector protein or an interaction partner of CagA (Olbermann *et al.*, 2010). Both strains have prematurely truncated *cag epsilon* genes which are thus, also likely to be pseudogenes.

VacA is a virulence factor which is associated with pathological outcomes of infection, depending on the allele present in a given strain of *H. pylori* (Atherton *et al.*, 1995; Leunk *et al.*, 1988; Rhead *et al.*, 2007). In addition to the *vacA* gene, 17874 contains 3 vacuolating cytotoxin paralogues. P79 contains 4 genes annotated as toxin-like outer membrane proteins (*vacA* paralogues) and 1 copy of *vacA*. One of these *vacA* paralogues may be inactive due to frameshift mutation. Both strains possess the i1 allele which is strongly associated with *vacA* and *cagA* positive *H. pylori* strains and has been associated with an increased risk of gastric atrophy and gastric carcinoma (Ferreira *et al.*, 2012).

Duodenal ulcer-promoting protein (DupA) has been identified as a marker of virulence when accompanied by an intact T4SS cluster (Jung *et al.*, 2012; Lu *et al.*, 2005). 17874 does not contain the *dupA* gene cluster which is present as a complete unit in G27, and incomplete in reference strains J99 and 26695 (Jung *et al.*, 2012). P79 is similar to 26695 as it contains a *tfs3b* partial cluster which contains a *virB4* gene with ~60% homology to *dupA*, and the secretion system genes *virB8*, *virD4* and *virD2*. Outer membrane proteins and adhesins can contribute to virulence by mediating bacterial-host interactions. The annotated genes encoding OMPs of 17874 (53 genes) and P79 (43 genes) are listed in Table 7 and Table 8. HopL is potentially non-functional in 17874 due to a frameshift mutation while an iron-regulated OMP

and a VacA paralogue are potentially frameshifted in the genome of P79. 17874 contains the gene encoding sialic acid-binding adhesin, SabA, which is absent from the genome of P79. The blood group antigen binding protein (BabA) is also absent from these strains. Outer inflammatory protein (OipA/HopH) and HopQ homologues as well as Hom and Hof families are present in both 17874 and P79. Therefore, the diverse selection of OMPs available for presentation on the cell surface of *H. pylori* contributes to the inter-strain variation of this pathogen. Table 9 and Table 10 list the ABC transporter proteins annotated in the genomes of 17874 and P79, respectively.

4.4 Homopolymeric Repeats

The genomes of *H. pylori* strains contain a characteristic high number of repeat sequences of varying lengths composed of cytosine or guanine nucleotides (Saunders *et al.*, 1998; Tomb *et al.*, 1997). Variation in the number of repeats incorporated during DNA replication by slipped-strand mispairing can alter the strength of expression of genes where the homopolymeric tract determines the promoter distance from its transcriptional start site (Moxon *et al.*, 2006). Alternatively, changes in GC tract length located within a gene can alter the reading frame and introduce premature stop codons. In this way, *H. pylori* can switch its expression profile in a process termed phase variation (Lertsethtakarn *et al.*, 2011). FliP is an example of this within the motility genes of *H. pylori*; a homopolymeric repeat of nine cytosines is responsible for the altered reading frame during expression of *fliP* in the non-motile 26695 strain (Figure 21). A motile pseudorevertant was isolated in which the C9 tract had reverted to a C8 repeat, enabling expression of the functional form of FliP. Analysis of this tract in the genome sequence of P79 confirmed the presence of a C8 repeat. In 17874, the repeat is disrupted by the presence of an adenine at position 5, a silent mutation, as is the case in J99 (Figure 21).

All genes containing homopolymeric GC repeats (≥ 8) either within the gene itself or upstream in the promoter region in the genomes of 17874 and P79 are listed in Table 11 and Table 12. In 17874, 39 genes have GC tracts of ≥ 8 bp, while there are 21 such genes in P79. There are hundreds of instances of 6 - 8 bp homopolymeric repeats. Broadly, the genes potentially subject to antigenic variation are involved in LPS and outer membrane synthesis, replication and cell division, motility, virulence and restriction modification systems. Additionally, some ABC transporter genes and genes encoding tRNA synthetases for amino acid synthesis and multidrug efflux

pumps were identified containing poly GC tracts. A number of predicted pseudogenes are potentially silenced by phase variation, including an endonuclease and a type III restriction enzyme (Table 13 and Table 14).

Therefore, *H. pylori* is complex in its regulation of gene expression. Phase variation enables this pathogen to conserve energy by restricting the genes of central cellular processes which can be expressed while retaining them for future use when they are required *e.g.* motility and virulence genes which are expressed during infection.

4.5 Natural Competence

H. pylori strains are generally naturally competent for the uptake of exogenous DNA (Yeh *et al.*, 2002). Natural competence of *H. pylori* involves core components of the T4SS, also known as the Com apparatus (Hofreuter *et al.*, 2001; Karnholz *et al.*, 2006). While competence is not essential during colonisation, it does promote colonisation during chronic infection as shown in a murine model (Dorer *et al.*, 2010, 2013). In *H. pylori*, most of the competence-related genes are found in two operons (*comB2 - 4* and *comB6 - 10*) (Hofreuter *et al.*, 2001).

All of the known components involved in natural competence are present in 17874 and P79 (Table 15). Comparing the Com apparatus components at a protein level highlighted some differences between strains 17874 and P79. ComB2 and ComB3 homologues have identical amino acid sequences while those of ComB4 vary in 22 positions and ComB6 varies in 14 positions. ComB4 is an ATPase which is thought to power DNA translocation; ComB6 is an inner membrane component (Hofreuter *et al.*, 2001; Karnholz *et al.*, 2006). In P79, ComB7 is prematurely truncated due to the presence of an additional thymidine which changes the reading frame, resulting in a protein which is 7 aa shorter than its homologue in 17874. However, ComB7 is not an essential competence gene and as P79 is more readily transformable compared to 17874, this truncation does not have a negative impact on the natural competence of the strain (Hofreuter *et al.*, 2001). Similarly, ComB8 is disrupted by a frame shift which changes the reading frame, resulting in a prematurely-truncated gene which encodes a 101 aa protein. This is due to the absence of one thymidine nucleotide in P79: 5'-TTGATG-3' in P79 where the sequence is 5'-TTTGATG-3' in 17874. This is likely a sequencing error as this truncated form of *comB8* lacks the transmembrane domain which is critical for its

function, yet P79 is more efficient at DNA uptake through the Com apparatus than 17874, which has the full length form of *comB*. The ComB9 homologue in 17874 (HP17_04796) encodes 6 additional amino acids which are absent in that of P79. HP17_04796 also differs in protein sequence from HP79_04966 at 6 amino acid positions, though these are likely not to impact on the function of this protein *e.g.* L/I substitution.

ComEC is an essential gene for *H. pylori* competence which is homologous to ComE3 in *Bacillus subtilis*, a channel-forming protein (Yeh *et al.*, 2003). A proposed model for the mechanism of DNA uptake by natural competence is a two-step process, whereby double-stranded DNA is transported across the outer-membrane to the periplasm. ComEC then produces a pore through which the DNA can traverse the inner-membrane and enter the cytoplasm (Stingl *et al.*, 2010). ComEC protein sequence is highly conserved among sequenced strains of *H. pylori* which harbour plasmids. While the sequences of 17874 and P79 ComEC homologues are not identical, these differences are confined to the variable regions, and hence are likely not to be responsible for the difference in competence between these strains.

HP17_03604 is a putative periplasmic competence protein which is absent in P79. ComH is an essential component of natural transformation and is conserved across strains with varying transformation efficiencies (Smeets *et al.*, 2000). It contains an N-terminal leader sequence, though the role of this protein in natural transformation is currently unknown. At a protein level, ComH of 17874 and P79 share 96% sequence similarity. DNA processing A (DprA) protein contributes to natural competence in *H. pylori*, where disruption of the gene causes reduced transformation efficiency of both plasmid and chromosomal DNA (Ando *et al.*, 1999). DprB is co-transcribed with DprA; it promotes DprA activity and may also function as a resolvase (Humbert *et al.*, 2011; Sharma *et al.*, 2010). The homologues of DprA and DprB in 17874 and P79 share 96% and 93% amino acid sequence similarity. The variation in DprB sequence is higher than is typical for *H. pylori* homologous proteins (4 - 5%), which possibly impacts on the function of this competence protein.

Restriction modification (RM) poses a barrier to transformation and recombination of exogenous DNA into *H. pylori* (Ando *et al.*, 2000). There are four types of RM sytems; the type II ststem is the most common and well-described of these. Type II RM requires the action of two enzyme types: methylases and endonucleases (Xu *et al.*, 2000). Methylases methylate DNA at specific recognition

sites to label DNA as "self". Endonucleases cleave DNA at specific recognition sites if these are not methylated. In this way, *H. pylori* can restrict the level of genetic exchange between strains. In 2012, a derivative strain of 26695 lacking type II restriction enzymes was found to have enhanced natural competence (Zhang and Blaser, 2012).

H. pylori strains possess a large number of RM systems and many of these are strain-specific (Huimin *et al.*, 2000). Comparative genomics of the RM systems of strains 26695 and J99 revealed that although *H. pylori* strains possess genes for multiple type II RM systems, many of these may not be biologically active (< 30% in J99/26695) (Lin *et al.*, 2001). Table 16 and Table 17 list the type II restriction modification genes present in the genomes of 17874 and P79, respectively. There are 21 such genes in the genome of 17874 while 9 were identified in that of P79. This is likely to be a key factor in the differing capacity of these strains for natural transformation due to altered restriction profiles. However, it must be considered that many of these systems may not be functionally active.

Interestingly, several restriction enzymes contain homopolymeric tracts (Table 12) and others are annotated as potential pseudogenes (Table 13) indicating that their expression can be modulated through phase variation. In P79, the gene encoding a type II methyltransferase (HP79_01260) is prematurely truncated due to a homopolymeric G tract containing 12 bp where the full length gene has 14 bp, as in reference strain 26695. RM and hypothetical genes are often strain-specific, as is the case for a number of gastric cancer-associated strains of *H. pylori* (McClain *et al.*, 2009). HP17_01508 and HP79_04682 encode homologues of the type II RM enzyme HsdR which is absent from the genome of J99. There is also a striking difference between the total number of methylases and endonucleases in 17874 and P79. 17874 possesses genes encoding 32 methylases and 21 endonucleases, while P79 possesses 37 methylases and just 13 endonucleases (Appendix 1 - Appendix 4), highlighting the inter-strain variation of RM components in *H. pylori*.

4.6 The Core Genome of *H. pylori*

The most recently determined core genome identified 1,063 genes common to 39 strains of *H. pylori* (Lu *et al.*, 2013). The revised core genome of *H. pylori* was determined by bi-directional BLASTP analysis of 43 sequenced strains including 17874 and P79 (Appendix 5). Here we present a more refined core genome

containing 898 genes (Appendix 6). Twenty six percent of the core genes are involved in metabolic pathways, while 10% are involved in the biosynthesis of secondary metabolites. Twenty two flagellum-related genes, 9 chemotaxis and 15 LPS biosynthesis genes are conserved across all 43 strains of *H. pylori*. Thirteen ABC transport genes are conserved along with components of the ComB natural transformation system. There are 172 genes encoding hypothetical proteins in the core genome of *H. pylori*, emphasising its capacity for encoding novel biological functions.

Of the 43 strains used to determine the core genome, 26 strains were isolated from individuals suffering from 1 of 4 *H. pylori*-related diseases: gastritis (10), duodenal ulcer (7), gastric cancer (7) and MALT lymphoma (2) (Appendix 5). The "disease core" genome was determined for these 26 strains, resulting in 977 genes which are common to all sequenced strains isolated from individuals suffering from *H. pylori*-related disease. Approximately 40% of the proteome of *H. pylori* consists of hypothetical proteins with no known function (Alm and Trust, 1999; Boneca *et al.*, 2003). Hypothetical proteins account for 330 products of the "disease core" genes according to the annotation of strain 26695. Seventy nine genes are unique to the disease-type strains which may include disease markers and virulence genes (Appendix 7). Functional analysis of the 28 hypothetical proteins may contribute to our understanding of the mechanisms behind the induction of disease by *H. pylori*. Of note, *vacA* is a core gene of the disease-inducing strains which is not conserved by all 43 sequenced strains. Gene content comparisons of *H. pylori* isolates from patients suffering from gastroduodenal diseases have also been employed to probe for biomarkers of disease (Romo-González *et al.*, 2009). Recently, Blanchard *et al.* reported the sequences of 65 *H. pylori* strains isolated from patients suffering from 4 disease states as well as asymptomatic adults (Blanchard *et al.*, 2013). In addition to the currently sequenced "disease" strains, this provides a valuable resource for further research into the pathogenesis of this gastric pathogen.

Core genome analysis for strains isolated from patients suffering from gastritis, duodenal ulcer, gastric cancer and MALT lymphoma was performed to serve as a platform for the identification of potential disease-specific genetic markers in *H. pylori*. The "gastritis core" genome consists of 1,186 core genes, 288 of which are conserved by this group in addition to the "total core" *H. pylori* genome (Appendix 8). These include the *cag*PAI, and genes encoding DNA translocase FtsK and

methicillin resistance protein. Additionally, 86 genes encoding hypothetical proteins form part of the "gastritis core" (26695 annotation). The "duodenal ulcer core" genome of *H. pylori* contains 1,236 conserved genes, 338 of which are duodenal ulcer-specific when compared to the "total core" genes (Appendix 9). In addition to 127 genes encoding hypothetical proteins, genes encoding urease accessory proteins, spore coat polysaccharide biosynthesis protein C and RM system components are also conserved (J99 annotation). The "gastric cancer core" genome consists of 1,114 genes, 216 of which are "gastric cancer core"-specific (Appendix 10). The additional genes encode products including 56 hypothetical proteins, OMPs, the *cag*PAI, virulence factor MviN, recombinase A and topoisomerase (F32 annotation). The "MALT core" genome contains 1,311 conserved genes and 413 "MALT core"-specific genes with reference to the "total core" genome (Appendix 11). Among these 413 genes are those encoding OMPs, chemotaxis proteins, cobalt-zinc-cadmium resistance protein, superozide dismutase and 128 hypothetical proteins (HELPY annotation).

4.7 Unique Genes

Strain-unique genes were identified using a pairwise bi-directional BLASTP comparison of 17874, P79, 26695 and J99, where the query sequence has no detectable homologues. The 17874 genome contains 41, 45 and 25 unique genes that are absent in 26695, J99 and P79, respectively, including 20 which are absent in all three other strains (Appendix 12). All of these genes encode hypothetical proteins except for the type II restriction enzyme R which is present in 26695 and P79 but absent in the genome of J99. Twenty four genes from 26695 are absent from the genome of 17874, including genes encoding a metal-binding polypeptide and hypothetical proteins (Appendix 13). Seven genes all encoding hypothetical proteins from the genome of J99 are absent in 17874 (Appendix 14). P79 contains 41, 54 and 38 unique genes that are absent in 26695, J99 and 17874, respectively (Appendix 15). Twenty two genes from 26695 are absent from P79, many of which are also absent in 17874. Six genes from J99 encoding hypothetical proteins are absent in P79 (Appendix 14). Thirty six genes are shared by 17874 and P79 but absent in 26695 and J99.

H. pylori has been associated with its anatomically modern human host for ~60,000 years (Linz *et al.*, 2007). Many of the strain-specific genes of *H. pylori*

encoding hypothetical proteins have a % GC content which is lower than the whole-genome average, indicating that they have been acquired by horizontal gene transfer. This is refective of the evolutionary proccess by which *H. pylori* continuously adapts to its environment. Therefore, it is likely that many of these uncharacterised hypothetical proteins have biological functions which give competitive advantage to the strain from which they were isolated for survival in their specific host *e.g.* colonisation factors and stress response genes. *H. pylori* is one of the most genetically variable pathogens described and genetic recombination has a significant impact on population genetics in which allelic diversity can be associated with pathogenicity *e.g. vacA* s1/m1 allele is linked to higher risk of gastric cancer development (Miehlke *et al.*, 2000; Suerbaum and Josenhans, 2007).

5 Conclusions

The increased ease and reduced expense in bacterial genome sequencing has lead to the availability of large volumes of data for gene mining. The genome of *H. pylori* undergoes extensive genetic flux, as revealed by comparative genomics of many strains. The draft genome sequence of 17874, the type strain for the species, has been deposited in GenBank. In addition, the draft genome of P79, a motile and readily transformable strain has also been deposited. Availability of these sequences will contribute to future motility studies as well as studies requiring genetic modification.

Comparative genomic analysis of strains 17874 and P79 revealed 1,639 and 1,699 coding genes in genomes of 1.61 and 1.64 Mbp, respectively. *H. pylori* genome size is much smaller than that of enteric pathogens *Staphylococcus aureus* (~2.8 Mbp), *Salmonella enterica* (4.5 - 4.8 Mb) and *Yersinia sp.* (4.3 - 4.8 Mb) (Chen *et al.*, 2010; Deng *et al.*, 2003; Gill *et al.*, 2005; Holt *et al.*, 2009). Variations in the ComB complement and differences between the RM systems encoded for in the genomes of 17874 and P79 are likely to be responsible for the difference in natural transformation efficiency between these strains. Phase variation facilitated by the presence of homopolymeric nucleotide repeats may also contribute to this difference. Many of the strain-unique genes in 17874 and P79 (compared to reference strains 26695 and J99) are in regions of low % GC content and encode hypothetical proteins. Core genome analysis of 43 sequenced strains of *H. pylori* identified a more conservative 898 core genes for the species. The core genome of *H. pylori* is much smaller than those of other pathogens including *S. aureus* (2,245 genes based on the genomes of 13 strains) and *Salmonella enterica* (2,882 genes based on the genomes of 73 strains) (Boissy *et al.*, 2011). Analysis of core genes conserved by strains isolated from patients suffering from *H. pylori*-related diseases identified potential biomarkers of disease including the *cag*PAI and *vacA*.

6 Tables and Figures

Table 2 General genome features of two *H. pylori* strains compared with reference strain *H. pylori* 26695

Feature	*H. pylori* CCUG 17874*	*H. pylori* P79*	*H. pylori* 26695
Genome size (bp)	1,615,763	1,641,495	1,667,867
G+C content (%)	38.97	38.86	39.00
Coding genes	1,639	1,699	1,590
Coding density (%)	86.5	85.1	90.4
rRNA operons	2	2	7
tRNAs	36	35	36
Pseudogenes	25	29	3
IS elements	3	5	14

*Figures for CCUG 17874 and P79 are estimates based on the draft assembly automated annotation.

Table 3 IS elements identified in the genomes of *H. pylori* CCUG 17874 and P79

Locus tag	Contig	Product
HP17_01198	22	IS605 transposase (tnpB)
HP17_01203	22	IS605 transposase (tnpA)
HP17_08409	160	IS605 transposase (tnpB)
HP79_04127	112	IS200 insertion sequence from SARA17
HP79_04132	112	Transposase-like protein, PS3IS
HP79_04137	112	IS606 transposase
HP79_06476	159	IS605 transposase (tnpA)
HP79_06481	159	IS605 transposase (tnpB)

Table 4 Phage-associated genes identified in the genomes of *H. pylori* CCUG 17874 and P79

Locus tag	Contig	Product
HP17_04069	73	Uncharacterised phage-associated protein
HP17_06262	106	Phage integrase family site-specific recombinase
HP17_08434	162	Phage/colicin/tellurite resistance cluster Y protein
HP79_00375	14	Phage integrase family site-specific recombinase
HP79_02479	69	Phage/colicin/tellurite resistance cluster Y protein

Table 5 Flagellar genes identified in the genomes of *H. pylori* CCUG 17874 and P79 and their orthologues in *H. pylori* strains 26695 and J99

Strain:	17874		P79		26695	J99	
Gene name	Locus tag	Contig	Locus tag	Contig	Locus tag	Locus tag	Product function
putative secreted heat shock protein	HP17_03754	71	HP79_01330	35	HP_1462	jhp_1355	Secreted protein involved in motility
fliR	HP17_00295	6	HP79_07780	188	HP_0173	jhp_0159	Flagellar biosynthetic protein
flgE1	HP17_03309, HP17_0329, HP17_03314	60, 62, 63	HP79_02664, HP79_02669	74, 75	HP_0870	jhp_0804	Fagellar hook protein
flgK	HP17_03469	68	HP79_08992	214	HP_1119	jhp_1047	Hook-associated protein 1 (HAP 1)
flgA	HP17_03844	72	HP79_01230	31	HP_1477	jhp_1370	Flagellar basal body P-ring biosynthesis protein
flaB	HP17_00935	17	HP79_04542	123	HP_0115	jhp_0107	Flagellin B
pflA	HP17_01373	24	HP79_06751	164	HP_1274	jhp_1195	Paralysed flagella protein PflA
flgH	HP17_01838	31	HP79_00560	19	HP_0325	jhp_0308	Flagellar basal body L ring protein
flaG 1	HP17_01853	31	HP79_00575	19	HP_0327	jhp_0310	Flagellar associated protein-glycosylation
fliF	HP17_01975	33	HP79_00717	21	HP_0351	jhp_0325	Flagellar basal body M ring protein
fliG	HP17_01980	33	HP79_00722	21	HP_0352	jhp_0326	Flagellar motor switch protein
fliH	HP17_01985	33	HP79_00727	21	HP_0353	jhp_0327	Flagellar export protein
hypothetical protein	HP17_02010	33	Absent		HP_0206	jhp_0192	Predicted ATPase involved in biogenesis of archaeal flagella
hypothetical protein	HP17_04675		HP79_08575		HP_0256	jhp_0240	Involved in motility and cell envelope architecture
flgG	HP17_02025	33	HP79_06991	170	HP_1092	jhp_0333	Basal body rod protein
flhA	HP17_02297	36	HP79_07273	175	HP_1041	jhp_0383	Flagellar basal body protein involved in export
flhF	HP17_02327	36	HP79_07303	175	HP_1035	jhp_0389	Flagellar biosynthesis regulator/GTP-binding

							protein
flhG	HP17_02332	36	HP79_07308	175	HP_1034	jhp_0390	ATP-binding protein
fliA	HP17_02342	36	HP79_07318	175	HP_1032	jhp_0392	Sigma 28 subunit of DNA-dependent RNA polymerase
fliM	HP17_02347	36	HP79_07323	175	HP_1031	jhp_0393	Flagellar motor switch protein
fliY	HP17_02352	36	HP79_07328	175	HP_1030	jhp_0394	Flagellar motor switch protein
flgE2	HP17_02559	41	HP79_01645	43	HP_0908	jhp_0844	Flagellar hook protein homolog
flgD	HP17_02564	41	HP79_01650	43	HP_0907	jhp_0843	Flagellar hook capping protein
fliE	HP17_04401	81	HP79_05396	135	HP_1557	jhp_1465	Flagellar hook-basal body protein
flgC	HP17_04406	81	HP79_05391	135	HP_1558	jhp_1466	Flagellar basal body rod protein
flgB	HP17_04411	81	HP79_05386	135	HP_1559	jhp_1467	Flagellar basal body rod protein
flgG	HP17_04556	82	HP79_05246	133	HP_1585	jhp_1492	Distal rod protein
flgI	HP17_04911	86	HP79_08525	202	HP_0246	jhp_0231	Flagellar basal body P-ring protein
flgL	HP17_05185	90	HP79_08807	209	HP_0295	jhp_0280	Flagellar hook-associated protein
flaA	HP17_05595	99	HP79_02059	61	HP_0601	jhp_0548	Flagellin A
fliN	HP17_05695	99	HP79_02184	66	HP_0584	jhp_0531	Flagellar motor switch protein
fliP	HP17_06317	108	HP79_00335	14	HP_0684, HP_0685	jhp_0625	Flagellar biosynthesis protein
fliW 2	HP17_06902	130	HP79_05914	149	HP_1377	jhp_1291	Flagellar assembly protein
hpaA3	HP17_07457	145	HP79_04057	109	HP_0492	jhp_0444	Flagellar sheath adhesin
fliI	HP17_07767	148	HP79_00065	4	HP_1420	jhp_1315	Flagellum-specific ATP synthase
fliQ	HP17_07772	148	HP79_00060	4	HP_1419	jhp_1314	Flagellar biosynthesis protein
flaG 2	HP17_07932	149	HP79_03366	95	HP_0751	jhp_0688	Uncharacterised flagellar protein
fliD	HP17_07937	149	HP79_03361	95	HP_0752	jhp_0689	Flagellar hook associated protein 2 (capping protein)

fliS	HP17_07942	149	HP79_03356	95	HP_0753	jhp_0690	Flagellin specific chaperone
flhB1	HP17_08034	150	HP79_03271	93	HP_0770	jhp_0707	Flagellar basal body protein
motB	HP17_08666	165	HP79_03001	83	HP_0816	jhp_0752	Flagellar motor protein
motA	HP17_08671	165	HP79_03011, HP79_03006	84, 83	HP_0815	jhp_0751	Flagellar motor protein
fliL	HP17_08701	165	HP79_03046	86	HP_0809	jhp_0745	Flagellar basal body-associated protein
rpoN	HP17_06522	116	HP79_00175	8	HP_0714	jhp_0652	RNA polymerase factor sigma-54
flgM	HP17_03484	68	HP79_09007	214	HP_1122	jhp_1051	Anti-fliA
envA/lpxC	HP17_02230	35	HP79_07213	174	HP_1052	jhp_0373	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase
fliT	HP17_07947	149	HP79_03351	95	HP_0754	jhp_0691	Flagellar chaperone, hypothetical protein
fliK	HP17_02569	41	HP79_01655	43	HP_0906	jhp_0842	Hypothetical protein
flgJ	HP17_02688	44	HP79_03583	101	HP_1233	jhp_1154	Hypothetical protein
atoS	HP17_04901	86	HP79_08515	202	HP_0244	jhp_0229	Histidine kinase specific for flgR
neuA/flmD	HP17_01843	31	HP79_00565	19	HP_0326	absent	CMP-N-acetylneuraminic acid synthetase
flaA1	HP17_06000	101	HP79_02849	79	HP_0840	jhp_0778	UDP-GlcNAc C6 dehydratase
putative fliZ	HP17_01473	24	HP79_06681	163	HP_1286	jhp_1206	Uncharacterised conserved protein *potential frameshift in P79
flhB2	HP17_04491	81	HP79_05311	135	HP_1575	jhp_1483	Homologue of flhB protein cytoplasmic domain
hpaA2	HP17_07264	143	HP79_05691	139	HP_0410	jhp_0971	Flagellar sheath associated protein paralog
hpaA	HP17_08561	165	HP79_03111	87	HP_0797	jhp_0733	Flagellar sheath associated protein paralog
flgZ	HP17_08776	166	HP79_08283	199	HP_0958	jhp_0892	Hypothetical protein
fliB	HP17_05065	88	HP79_08677	204	HP_0274	jhp_0259	Flagellin N-methylase family protein

Table 6 List of genes in the *cag*PAIs of *H pylori* CCUG 17874 and P79

17874		**P79**		
Locus tag	Contig	Locus tag	Contig	Product
HP17_07617	147	HP79_03897	108	Cag pathogenicity island protein (Cag1)
HP17_07622*	147	HP79_03892*	108	Cag pathogenicity island protein epsilon
HP17_07627	147	HP79_03887/82*	108	Cag pathogenicity island protein (Cag3)
HP17_07632	147	HP97_03877	108	Cag pathogenicity island protein gamma
HP17_07637	147	HP79_03872/67	108/7	Cag pathogenicity island protein 5
HP17_07642	147	HP79_03862	107	Cag pathogenicity island protein alpha
HP17_07647	147	HP79_03857	107	Cag pathogenicity island protein (Cag6)
HP17_07652	147	HP79_03852	107	Hypothetical protein
HP17_07657/03179/84	147/58/57	Pseudogene	107	Cag pathogenicity island protein Y VirB10-like protein (Cag7)
HP17_03174	57	HP79_02474	68	Cag pathogenicity island protein X (Cag8)
HP17_03169	57	HP79_02469	68	Cag pathogenicity island protein W (Cag9)
HP17_03164	57	HP79_02464	68	Cag pathogenicity island protein V (Cag10)
HP17_03159	57	HP79_02459	68	Cag pathogenicity island protein U (Cag11)
HP17_03154	57	HP79_02454	68	Cag pathogenicity island protein T (Cag12)
HP17_03149	57	HP79_02449	68	Cag pathogenicity island protein S (Cag13)
HP17_03144	57	Absent		Hypothetical protein
HP17_05960	100	HP79_02444	68	Cag pathogenicity island protein Q (Cag14)
HP17_05955	100	Absent		Hypothetical protein
Absent		HP79_02443	68	Cag island protein (Cag15)
HP17_05950	100	HP79_02439	68	Cag pathogenicity island protein M (Cag16)
HP17_05945	100	HP79_02434	68	Cag pathogenicity island protein N (Cag17)

HP17_05940	100	HP79_02429	68	Cag pathogenicity island protein L (Cag18)
HP17_05935	100	HP79_02424	68	Cag pathogenicity island protein (Cag19)
HP17_05930	100	HP79_02419	68	Cag pathogenicity island protein H (Cag20)
HP17_05925	100	HP79_02414	68	Cag pathogenicity island protein G (Cag21)
HP17_05920	100	HP79_02409	68	Cag pathogenicity island protein F (Cag22)
HP17_05915	100	HP79_02404	68	Cag pathogenicity island protein E (Cag23)
HP17_05910	100	HP79_02399	68	Cag pathogenicity island protein D (Cag24)
HP17_05905	100	HP79_02394	68	Cag pathogenicity island protein C (Cag25)
HP17_05900	100	HP79_02389	68	Cag pathogenicity island protein B
HP17_05895	100	HP79_02384	68	Cytotoxin-associated protein A (Cag26)

*Potentially frameshifted.

Table 7 List of OMPs annotated in the genome of *H. pylori* CCUG 17874

Locus tag	Contig	Product
HP17_05565	97	Outer-membrane protein of the hefABC efflux system
HP17_06167	104	Protective surface antigen D15
HP17_03504	68	Peptidoglycan-associated lipoprotein precursor
HP17_03904	72	Lipase-like protein
HP17_00746	15	Outer membrane protein
HP17_00045	2	Outer membrane protein
HP17_06737	125	Outer membrane protein
HP17_02957	48	Outer membrane protein
HP17_04826	86	Outer membrane protein HopA; signal peptide
HP17_04953	87	Outer membrane protein HopF; putative signal peptide
HP17_04958	87	Outer membrane protein HopG
HP17_00320	7	Outer membrane protein
HP17_00355	8	Outer membrane protein; signal peptide
HP17_00415	10	Putative outer membrane protein
HP17_00501	12	Putative outer membrane protein
HP17_07134	140	Outer membrane protein HofC
HP17_07129	140	Outer membrane protein HofD; signal peptide
HP17_05785	99	Outer membrane protein, predicted permease
HP17_05385	93	Outer membrane protein, OipA
HP17_06247	106	Outer membrane protein
HP17_06387	112	Outer membrane protein
HP17_06472	115	Outer membrane protein HopE
HP17_06497	116	Putative outer membrane protein HomB
HP17_06502	116	Outer membrane protein
HP17_06567	121	Outer membrane protein SabA
HP17_08129	153	Outer membrane protein HofF
HP17_08556	165	Outer membrane protein HorG
HP17_06005	101	Outer membrane protein P1
HP17_03344	64	Iron-regulated outer membrane protein
HP17_02524	41	Outer membrane protein HopB
HP17_02519	41	Outer membrane protein
HP17_02509	41	Iron-regulated outer membrane protein
HP17_01010	19	Putative outer membrane protein
HP17_02160	35	Outer membrane protein HorD
HP17_02080	34	Outer membrane protein HofB; signal peptide
HP17_01833	31	Outer membrane protein HorC; signal peptide
HP17_01788	29	Outer membrane protein
HP17_08264	158	Outer membrane protein HopI
HP17_08259/54	158	Outer membrane protein HopL *

HP17_08199	157	Outer membrane protein HofH
HP17_03139	56	Outer membrane protein HopQ; signal peptide
HP17_03102	54	Outer membrane protein (omp27) (HopQ)
HP17_08154/49	156/5	Outer membrane protein BabA
HP17_03689	71	Outer membrane protein HomD; signal peptide
HP17_03784	71	Outer membrane protein; signal peptide
HP17_03799	72	Outer membrane protein HorJ; signal peptide
HP17_03979	73	Outer membrane protein (omp32)
HP17_04039	73	Iron-regulated outer membrane protein
HP17_04226	78	Outer membrane protein
HP17_04436	81	Outer membrane protein
HP17_04631	82	Outer membrane protein
HP17_04716	83	Outer membrane protein (omp2)
HP17_08806	168	Outer membrane protein HopK; signal peptide

*Potentially frameshifted.

Table 8 List of OMPs annotated in the genome of *H. pylori* P79

Locus tag	Contig	Product
HP79_02039	61	Hypothetical protein
HP79_00490/85	16/15	Protective surface antigen D15
HP79_09027	214	Peptidoglycan-associated lipoprotein precursor
HP79_01170	31	Lipase-like protein
HP79_05146/41	131/30	Outer membrane protein
HP79_05056	129	Outer membrane protein (omp2)
HP79_04747	126	Outer membrane protein (omp3)
HP79_08030	191	Outer membrane protein (omp4)
HP79_08433	201	Outer membrane protein (omp6)
HP79_08560	202	Outer membrane protein (omp7)
HP79_08565	202	Outer membrane protein
HP79_08762/67/72	206/7	Toxin-like outer membrane protein*
HP79_00515	18	Outer membrane protein (omp9)
HP79_00555	19	Outer membrane protein (omp10)
HP79_01819	13	Outer membrane protein (omp13) (OipA)
HP79_00395	14	Outer membrane protein (omp14)
HP79_00285	12	Outer membrane protein
HP79_00220/15	9/8	Outer membrane protein HopE; signal peptide
HP79_03156	88	Outer membrane protein
HP79_03116	87	Outer membrane protein (omp18)
HP79_02854	79	Outer membrane protein P1 (ompP1)
HP79_02634/29	73/72	Iron-regulated outer membrane protein
HP79_01725	45	Outer membrane protein (omp19)
HP79_01625	43	Outer membrane protein (omp20)
HP79_01620	43	Outer membrane protein (omp21)
HP79_01605/00	43	Iron-regulated outer membrane protein*
HP79_01570	42	Toxin-like outer membrane protein
HP79_08137	195	Putative outer membrane protein
HP79_01495	38	Outer membrane protein (omp23)
HP79_01460	37	Outer membrane protein (omp24)
HP79_06306/01	154/3	Outer membrane protein (omp25)
HP79_06296	153	Outer membrane protein (omp26)
HP79_06179	152	Outer membrane protein (omp27) (HopQ)
HP79_06029	149	Outer membrane protein (omp30)
HP79_01045	29	Iron-regulated outer membrane protein
HP79_01105	30	Outer membrane protein (omp32)
HP79_01275/80	32/33	Outer membrane protein (omp31)
HP79_01555	39	Outer membrane protein (omp12)
HP79_01560	40	Outer membrane protein

HP79_07460	179	Outer membrane protein (omp9)
HP79_08077	194	Outer membrane protein (omp29)
HP79_01754	48	Outer membrane protein
HP79_05366	135	Outer membrane protein

*Potentially frameshifted.

Table 9 List of ABC transport genes annotated in *H. pylori* CCUG 17874

Locus tag	Contig	Product
HP17_08389	159	ABC-type antimicrobial peptide transport system, ATPase component
HP17_04938	87	Oligopeptide permease ATPase protein
HP17_04943	87	Oligopeptide permease integral membrane protein
HP17_05200	90	Peptide ABC transporter substrate-binding protein
HP17_05205	90	Peptide ABC transporter permease
HP17_05210	90	ABC-type transport system, permease; dipeptide transporter protein 3; membrane protein
HP17_05215	90	ABC-type transport system, ATP-binding protein; dipeptide transporter protein 4
HP17_05220	90	Dipeptide ABC transporter
HP17_05605	99	ABC-type transport system, permease and ATP- binding protein; putative membrane protein
HP17_05600	99	Multidrug resistance protein SpaB
HP17_05530	95	Multidrug resistance protein SpaB
HP17_05525	95	ABC transporter, permease
HP17_05520	95	ABC transporter, ATP-binding protein
HP17_06527	116	ABC-type transport system, ATP binding protein
HP17_07837	149	Hypothetical protein, ABC-type multidrug transport system
HP17_07917	149	Cell division protein, ABC-type antimicrobial peptide transport system, ATPase component
HP17_08119	152	Hypothetical protein, ABC-type transport system, involved in lipoprotein release
HP17_08651	165	Osmoprotection protein (proV)
HP17_08656	165	ABC-type transport system, permease; betaine/proline/choline transporter; membrane protein
HP17_03409	66	Iron (III) dicitrate transport system ATP-binding protein
HP17_03414	66	Iron(III) dicitrate ABC transporter permease protein (fecD)
HP17_01025	19	Amino acid ABC transporter permease
HP17_01030	19	Putative polar amino acid transport system substrate-binding protein
HP17_02972	50	Molybdenum ABC transporter ATP-binding protein (modD)
HP17_02982	51	Molybdenum ABC transporter (modB)
HP17_02987	51	Molybdenum ABC transporter
HP17_02210	35	Hypothetical protein, ABC-type multidrug transport system
HP17_02085	34	ABC-type transport system, ATP binding protein; lipid A and glycerophospholipid transporter; membrane protein; signal peptide
HP17_02020	33	Hypothetical protein, ABC-type multidrug transport system, ATPase component
HP17_08189	157	Glutamine ABC transporter permease
HP17_08184	157	Glutamine ABC transporter permease

HP17_08179	157	Phosphate ABC transporter ATP-binding protein
HP17_08174	157	Glutamine ABC transporter periplasmic glutamine-binding protein
HP17_02818	44	ABC transporter ATP-binding protein
HP17_02763	44	ABC-2 type transport system ATP-binding protein
HP17_01258	24	Oligopeptide ABC transporter, permease protein
HP17_01263	24	ABC transporter substrate-binding protein
HP17_03764	71	ABC transport system substrate binding protein t
HP17_03769	71	ABC transporter ATP-binding protein
HP17_03774	71	ABC transporter permease protein
HP17_03889	72	Antibiotic transport system permease protein
HP17_03894	72	Hypothetical protein, ABC-type multidrug transport system
HP17_04421	81	Iron(III) ABC transporter periplasmic iron-binding protein (ceuE)
HP17_04426	81	Iron(III) ABC transporter periplasmic iron-binding protein
HP17_04436	81	Outer membrane protein, ABC-type metal ion transport system
HP17_04496	81	DL-methionine transporter ATP-binding subunit
HP17_04501	81	ABC-type transport system, permease; putative D- and L-methionine transport protein; putative membrane protein

Table 10 List of ABC transport genes annotated in *H. pylori* P79

Locus tag	Contig	Product
HP79_04682	126	Type II restriction enzyme R protein (hsdR), ABC-type sugar transport systems
HP79_07745	188	ABC transporter ATP-binding protein
HP79_08545	202	Oligopeptide ABC transporter ATP-binding protein (oppD)
HP79_08550	202	Oligopeptide permease integral membrane protein
HP79_08822	209	Dipeptide ABC transporter periplasmic dipeptide- binding protein (dppA)
HP79_08827	209	Dipeptide ABC transporter periplasmic dipeptide- binding protein (dppB)
HP79_08832	209	Dipeptide ABC transporter periplasmic dipeptide- binding protein (dppC)
HP79_08837	209	Dipeptide ABC transporter periplasmic dipeptide- binding protein (dppD)
HP79_08842	209	Dipeptide ABC transporter ATP- binding protein (dppF)
HP79_04212	114	Molybdenum ABC transporter periplasmic molybdate-binding protein (modA)
HP79_04217	114	Molybdenum ABC transporter ModB
HP79_04222	114	Molybdenum ABC transporter ATP-binding protein (modD)
HP79_02069	61	Multidrug resistance protein (spaB)
HP79_01984	57	Hypothetical protein, ABC-type multidrug transport system
HP79_01979	57	ABC transporter, permease
HP79_01974	57	ABC transporter ATP-binding protein
HP79_00170	8	ABC-type transport system, ATP binding protein
HP79_03381	95	Cell division protein (ftsE)
HP79_03171	90	Hypothetical protein, ABC-type transport system involved in lipoprotein release
HP79_02991	83	Osmoprotection protein (proWX)
HP79_02986	83	Osmoprotection protein (proV)
HP79_02559	70	Iron compounds ABC transporter ATP-binding protein
HP79_02554	70	Iron(III) dicitrate ABC transporter permease protein (fecD)
HP79_08157	195	Amino acid ABC transporter permease protein (yckJ)
HP79_08162/67	195/6	Amino acid ABC transporter periplasmic binding protein (yckK)
HP79_07051/46	172/1	Multidrug resistance protein (msbA)
HP79_06219	152	Glutamine ABC transporter permease protein (glnP)
HP79_06214	152	Glutamine ABC transporter, permease protein
HP79_06209	152	Phosphate ABC transporter ATP-binding protein
HP79_06204	152	Glutamine ABC transporter periplasmic glutamine-binding protein (glnH)
HP79_03708	102	Multidrug resistance protein (hetA)
HP79_03653	102	ABC transporter ATP-binding protein

HP79_06881	166	Oligopeptide ABC transporter permease protein (oppB)
HP79_06496	161	Hypothetical protein, ABC-type multidrug transport system
HP79_01180	31	Hypothetical protein, ABC-type multidrug transport system
HP79_01185	31	Antibiotic transport system permease protein
HP79_01305	35	Hypothetical protein, ABC-type transport system involved in resistance to organic solvents
HP79_01320	35	ABC transport system substrate binding protein
HP79_06491	160	Iron(III) ABC transporter periplasmic iron- binding protein (ceuE)
HP79_06486	160	Iron(III) ABC transporter periplasmic iron- binding protein (ceuE)
HP79_05376	135	Iron(III) ABC transporter periplasmic iron- binding protein
HP79_05366	135	Outer membrane protein, ABC-type metal ion transport system
HP79_05306	135	DL-methionine transporter ATP-binding subunit
HP79_05301	134	D-methionine transport system permease protein

Table 11 List of the homopolymer G/C tracts in the genomes of *H. pylori* CCUG 17874

Track length	Homopolymer	Coordinates	Within or upstream	Locus tag	Gene
14	C	1066317	Within	HP17_05490	Glycosyltransferase involved in LPS biosynthesis
	C	317042c	Within	HP17_07012	Type I restriction-modification system methyltransferase subunit
11	C	122576c	Within	HP17_00245	Histidine kinase sensor protein
	C	571724	Within	HP17_01893	Unique hypothetical protein
	G	358564c	Within	HP17_01553	Methionine aminopeptidase
	C	317904c	Within	HP17_07017	Hypothetical protein; possible helicase
	G	1332106	Upstream	HP17_06847	Biotin synthase
	G	160957c	Within	HP17_04104	Adenine specific DNA methylase Mod
10	G	286752c	Within	Pseudogene	Putative type III restriction enzyme M protein
9	C	17213	Upstream	HP17_00596	Hypothetical protein
	G	589898	Within	HP17_06182	Processing zinc-metalloprotease
	G	609849c	Upstream	HP17_06272	Hypothetical protein, predicted permease
	C	935726	Within	HP17_03032	Type I restriction enzyme R protein (HsdR)
	C	1061205	Within	HP17_03459	Unique hypothetical protein
	G	495211c	Upstream	HP17_08159	Hypothetical protein, predicted permease
	G	127478c	Within	HP17_04266	DNA polymerase III subunit epsilon
	G	1555297	Within	HP17_04656	DNA primase

8	G	1570131c	Within	HP17_00781	Hypothetical protein, predicted cell wall-associated hydrolase
	G	79680	Upstream	HP17_00950	Fe-S oxidoreductases
	G	261927	Within	HP17_05160	Diaminopimelate decarboxylase
	C	371781	Within	HP17_07239	Hypothetical protein
	C	1119744c	Within	HP17_05720	Hypothetical protein, predicted neuraminidase (sialidase)
	G	539301	Within	HP17_05535	Vacuolating cytotoxin VacA
	C	1060411c	Within	HP17_05460	Hypothetical protein, predicted aspartate/tyrosine/aromatic aminotransferase
	G	586755	Within	HP17_06167	Outer membrane protein, protective surface antigen D15
	G	994105c	Within	HP17_06332	Ferrous iron transport protein B
	G	915690c	Within	HP17_08017	Hypothetical protein
	G	769945	Within	HP17_06005	Outer membrane protein P1
	C	832222	Within	HP17_02559	FlgE
	G	745461c	Within	HP17_08791	GpsA, NAD(P)H-dependent glycerol-3-phosphate dehydrogenase
	G	607041c	Within	HP17_02085	ABC-type transport system, ATP binding protein; lipid A and glycerophospholipid transporter; membrane protein; signal peptide
	G	498344c	Upstream	HP17_08194	Carbon starvation protein
	G	1132966	Within	HP17_03072	Multidrug-efflux transporter
	G	448806c	Within	HP17_02808	Adenine-specific DNA methylase
	G	326181c	Within	HP17_07072	Pgk, phosphoglycerate kinase
	G	218249c	Within	HP17_03809	Type IIS R-M system restriction enzyme
	G	108811c	Within	Pseudogene	Preprotein translocase subunit SecD
	C	11186	Within	HP17_07112	RepA

Table 12 List of the homopolymer G/C tracts in the genomes of *H. pylori* P79

Track length	Homopolymer	Coordinate	Within or upstream	Locus tag	Gene
23	C	736422c	Within	HP79_04869	Hypothetical protein, predicted chromosome segregation ATPase
14	C	755064c	Upstream	HP79_04157	Type I restriction enzyme R protein
	C	846702	Within	HP79_04667	Hypothetical protein
13	C	1389708	Within	HP79_07820	Histidine kinase sensor protein
12	G	236345	Within	HP79_01260	Type IIS restriction enzyme R protein (BCGIB)
11	C	785102c	Within	HP79_04617	Hypothetical protein
	G	284324c	Within	HP79_07520	Hypothetical protein
10	G	775826c	Within	HP79_04657	2-hydroxyacid dehydrogenase
	G	1062663	Within	HP79_05864	Adenine-specific DNA methylase
9	G	116689	Within	HP79_00622	Hypothetical protein
	C	575481c	Within	HP79_05774	Hypothetical protein, predicted helicase
8	C	1556675c	Within	HP79_00335	FliP
	G	1524647c	Within	HP79_00490	Protective surface antigen D15
	G	1427680c	Within	HP79_01035	Selenocysteine synthase
	C	1305344c	Within	HP79_01645	FlgE

C	1205534	Within	HP79_02214	Hypothetical protein, predicted neuraminidase (sialidase)
G	893230c	Within	HP79_04022	Outer membrane phospholipase A1
G	981644	Within	HP79_05436	Preprotein translocase subunit SecD
G	531200c	Within	HP79_06044	Alanine dehydrogenase
G	509784c	Within	HP79_06149	Multidrug-efflux transporter
G	1311993	Within	HP79_07363	Hypothetical protein

Table 13 Predicted pseudogenes in the genome of *H. pylori* CCUG 17874

Locus tag	Contig	Product
HP17_00100	3	Iron-sulphur cluster binding protein
HP17_00656	13	ATP-binding protein
HP17_00691	13	Urease accessory protein UreE
HP17_00940	17	DNA topoisomerase I
HP17_01378	24	Phosphomannomutase
HP17_02000	33	Hypothetical protein
HP17_02907	46	Aldo-keto reductase
HP17_03969	73	Putative endonuclease
HP17_04089	74	Type III restriction enzyme
HP17_04134	75	Chromosomal replication initiation protein
HP17_04526	80	Undecaprenyl phosphate N-acetylglucosaminyltransferase
HP17_04576	82	Hypothetical protein
HP17_04721	84	Type II citrate synthase
HP17_05270	91	Type II restriction enzyme
HP17_05850	99	Sialidase A
HP17_05975	101	Thiamine biosynthesis protein
HP17_06322	108	Iron (III) dicitrate transport protein FecA; signal peptide
HP17_06397	112	N-methyl hydantoinase
HP17_06452	115	Hypothetical protein
HP17_06852	130	Type III restriction enzyme R protein (res 1)
HP17_06862	130	Putative type III restriction enzyme M protein
HP17_07144	140	Non-functional type II restriction endonuclease
HP17_07487	145	Sodium- and chloride-dependent transporter; membrane protein
HP17_08209	157	Tetracycline resistance protein tetA (P)
HP17_08349	159	MobC-like protein

Table 14 Predicted pseudogenes in the genome of *H. pylori* P79

Locus tag	Contig	Product
HP79_00005	1	Outer membrane protein (omp 29)
HP79_00320	14	Iron (II) transport protein (feoB)
HP79_00410	14	Type II R-M system protein
HP79_00807	21	Type II DNA modification enzyme (methyltransferase)
HP79_01005	29	Cytoplasmic protein
HP79_01310	35	ABC transporter ATP-binding protein
HP79_01670	43	Phosphate acetyltransferase
HP79_01994	57	Vacuolating cytotoxin VacA
HP79_02024	61	Acriflavine resistance protein (acrB)
HP79_03401	95	Rod shape determining protein RodA
HP79_03446	96	D-alanyl-alanine synthetase A3
HP79_03837	107	Cag pathogenicity island protein (cag7)
HP79_03842	107	Cag pathogenicity island protein Y VirB10-like protein
HP79_03967	109	Glycolate oxidase subunit (glcD)
HP79_04252	115	Nicotinate-nucleotide adenyltransferase
HP79_04789	127	Urease subunit beta
HP79_04901	128	Restriction endonuclease
HP79_04926	128	Transcriptional regulator (hypF)
HP79_05256	133	DNA-binding/iron metalloprotein/AP endonuclease
HP79_06004	149	DNA repair protein (recN)
HP79_06246	153	Glucose-6-phosphate isomerise
HP79_06446	158	Type II DNA modification methyltransferase
HP79_06681	163	Hypothetical protein
HP79_06831	166	NAD+-dependent deacetylase, Sir2 family
HP79_06871	166	Oligopeptide ABC transporter periplasmic oligopeptide-binding
HP79_07870	189	Recombinase A
HP79_07935	189	Sodium/sulphate symporter
HP79_07950	189	L-lactate permease (lctP)
HP79_08777	207	Diaminopimelate decarboxylase (dap decarboxylase) (lysA)

Table 15 List of competence-related genes in the genomes of *H. pylori* CCUG 17874 and P79

Locus tag	Contig	Product
HP17_03604	70	Periplasmic competence protein-like protein
HP17_04671	83	ComB2
HP17_04676	83	ComB3
HP17_04681	83	ComB4
HP17_04781	84	ComB6
HP17_04786	84	ComB7
HP17_04791	84	ComB8
HP17_04796	84	ComB9
HP17_04801	84	ComB10
HP17_06972	131	ComEC
HP17_01883	31	DNA processing chain A (DprA)
HP17_01888	31	DprB
HP17_04129	75	ComH
HP79_05106	129	ComB2
HP79_05101	129	ComB3
HP79_05096	129	ComB4
HP79_04981	128	ComB6
HP79_04976	128	ComB7
HP79_04971	128	ComB8
HP79_04966	128	ComB9
HP79_04961	128	ComB10
HP79_05824	147	ComEC
HP79_00612	20	DNA processing chain A (DprA)
HP79_00617	20	DprB
HP79_00963	28	ComH

Table 16 List of the type II restriction modification system components identified in the genome of *H. pylori* CCUG 17874

Locus tag	Contig	Product
HP17_00556	13	Type II adenine methyltransferase
HP17_00811	15	Type II adenine methyltransferase
HP17_04995	88	Type II DNA modification methyltransferase
HP17_05005	88	Type II R-M system restriction endonuclease
HP17_05010	88	Type II DNA modification enzyme (methyltransferase)
HP17_08484	162	Type II restriction endonuclease
HP17_07184	141	Type II adenine methyltransferase
HP17_07164	140	Type II DNA modification enzyme (methyltransferase)
HP17_07159	140	Type II restriction endonuclease
HP17_07154	140	Type II DNA modification enzyme
HP17_06342	110	Putative type II cytosine specific methyltransferase
HP17_06347	110	Putative type II restriction enzyme
HP17_03479	68	M. HpyAVIII, type II cytosine specific DNA methyltransferase
HP17_01443	24	Type II restriction endonuclease
HP17_07037	133	Type II R-M system restriction endonuclease
HP17_03809	72	Type IIS R-M system restriction enzyme
HP17_03814	72	Type IIS restriction enzyme M protein (Mod)
HP17_03959	73	Type II methylase
HP17_03964	73	Type II adenine methyltransferase
HP17_04079/84	73/74	Type IIS restriction-modification protein
HP17_04286	80	Putative type II methylase protein

Table 17 List of the type II restriction modification system components identified in the genome of *H. pylori* P79

Locus tag	Contig	Product
HP79_01020	29	Type IIS restriction enzyme R and M protein (ECO57IR)
HP79_04677	126	Type II restriction enzyme M protein (hsdM)
HP79_05511	137	Type II N-6 Adenine-specific DNA methylase
HP79_05849	147	Type IIS restriction enzyme R protein (MBOIIR)
HP79_05854	147	Type IIS restriction enzyme M1 protein (mod)
HP79_05859	147	Type IIS restriction enzyme M2 protein (mod)
HP79_01255	31	Type IIS restriction enzyme M protein (mod)
HP79_01260/65	31	Type IIS restriction enzyme R protein (BCGIB)
HP79_04682	126	Type II restriction enzyme R protein (hsdR)

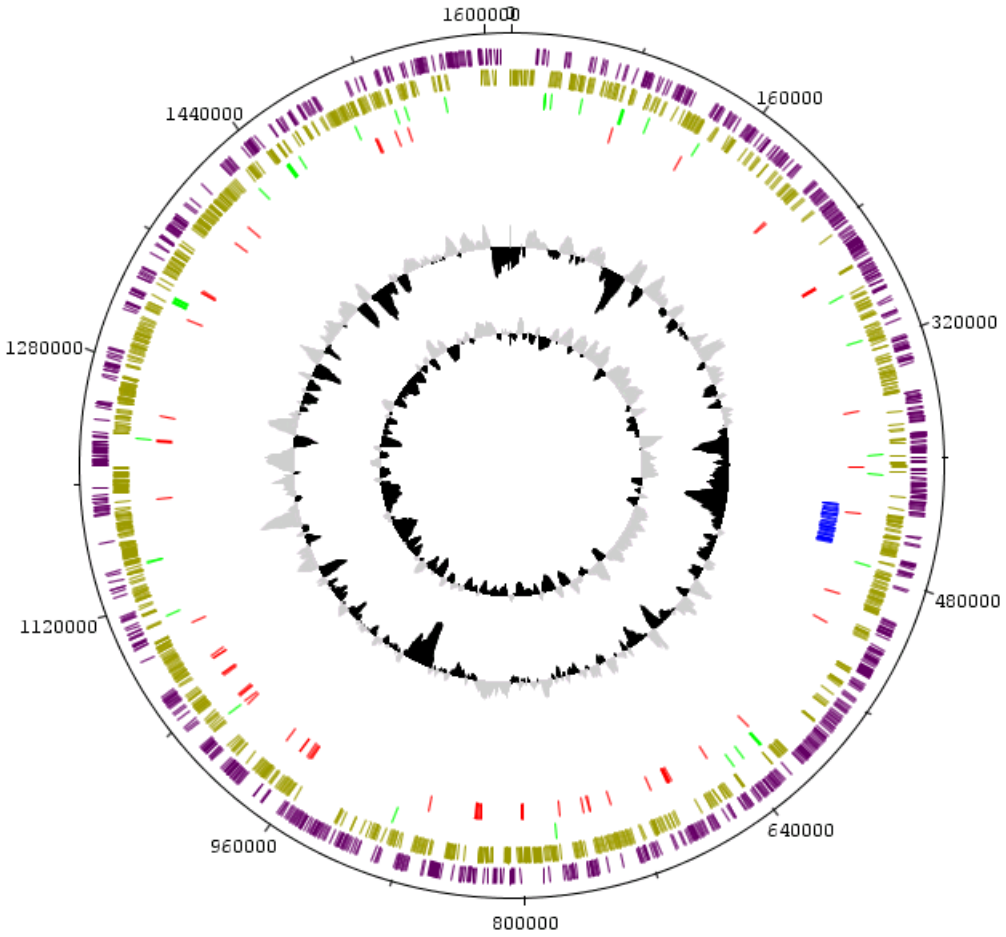


Figure 15 Genome atlas of *H. pylori* CCUG 17874.

Graphical representation of the genome was generated using Artemis. Numbers are nucleotide co-ordinates. From the outermost circle to the innermost: *H. pylori* genes on the forward strand (purple); *H. pylori* genes on the reverse strand (gold); pseudogenes (green); flagellar genes (red); cag PAI genes (blue); % GC (black= below the mean, grey= above the mean); and GC skew.

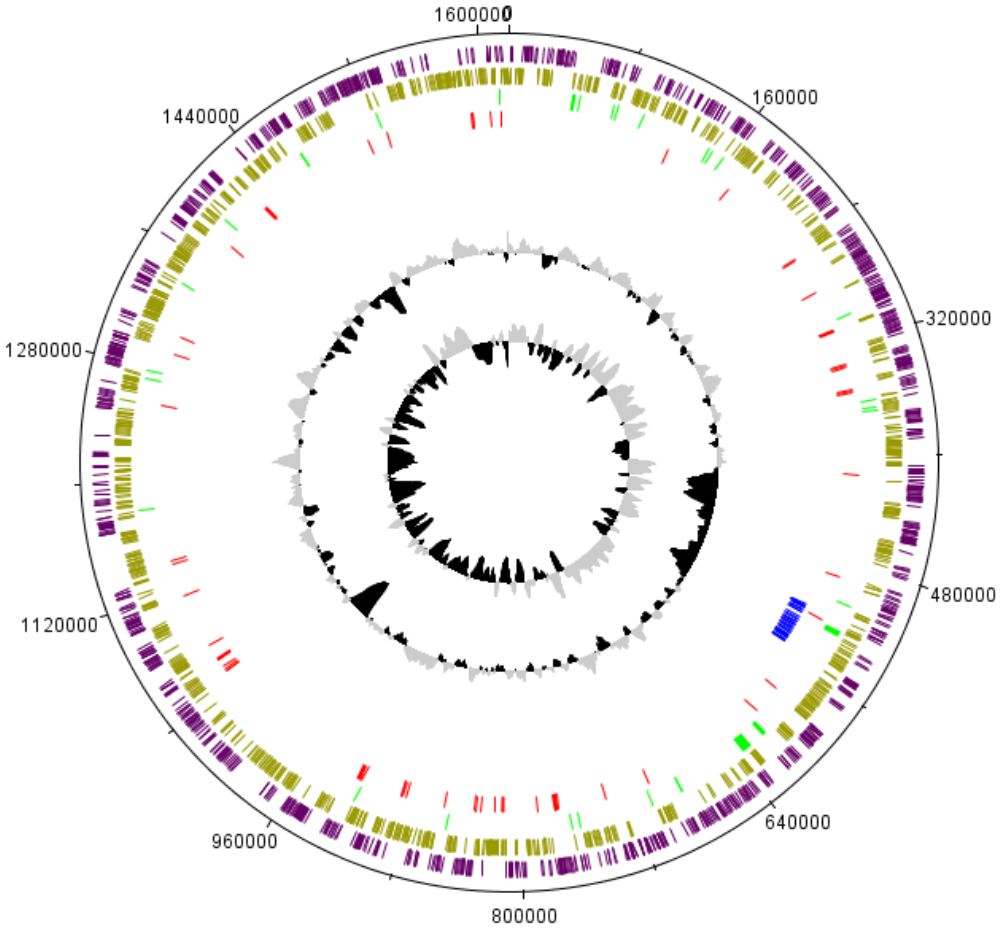
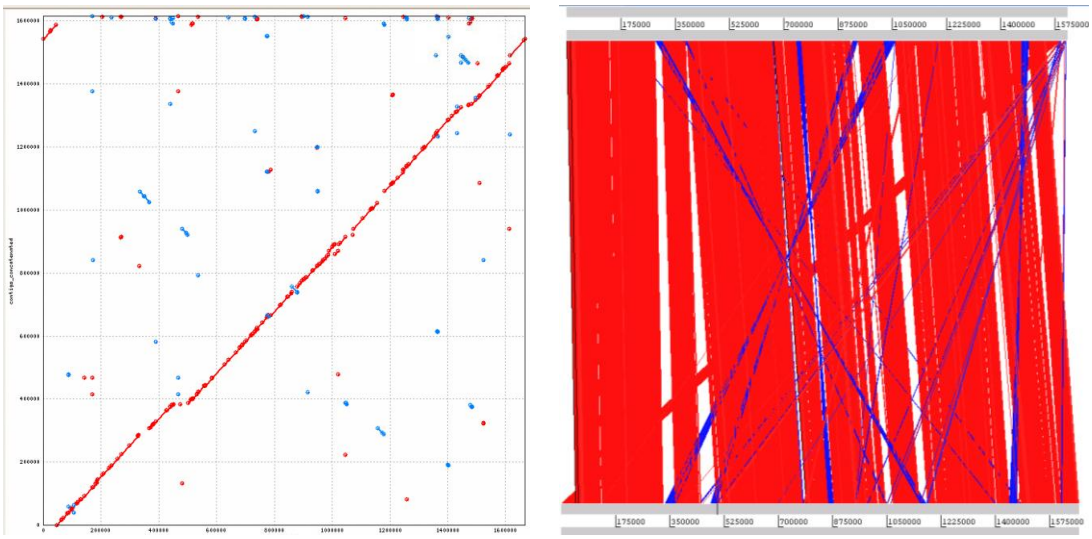


Figure 16 Genome atlas of *H. pylori* P79.

Graphical representation of the genome was generated using Artemis. Numbers are nucleotide co-ordinates. From the outermost circle to the innermost: *H. pylori* genes on the forward strand (purple); *H. pylori* genes on the reverse strand (gold); pseudogenes (green); flagellar genes (red); cag PAI genes (blue); %GC (black= below the mean, grey= above the mean); and GC skew.

A *H. pylori* CCUG 17874



B *H. pylori* P79

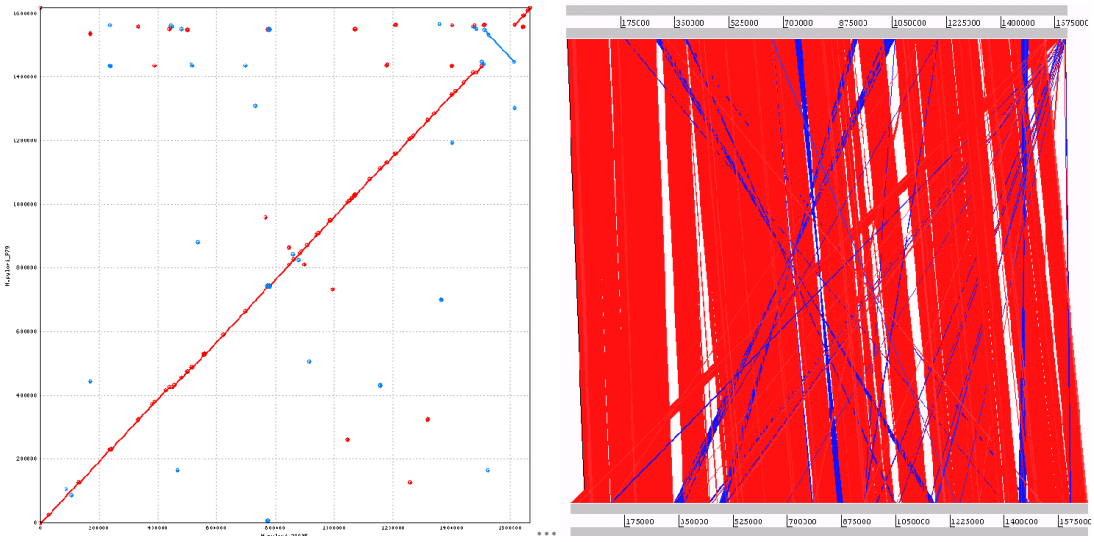


Figure 17 Genome synteny of *H. pylori* strains CCUG 17874/P79 and 26695.

Left panel: Mummerplot alignment of *H. pylori* CCUG 17874 (**A**) and P79 (**B**) (Y-axis) and reference strain *H. pylori* 26695. Red dots represent regions of homology between the genomes which are in the same orientation. Blue dots represent homology between the genomes which are in the opposite orientation. Right panel: ACT comparison (DNA *vs* DNA) of *H. pylori* CCUG 17874 (**A**) and P79 (**B**) (top) and reference strain *H. pylori* 26695 (bottom).

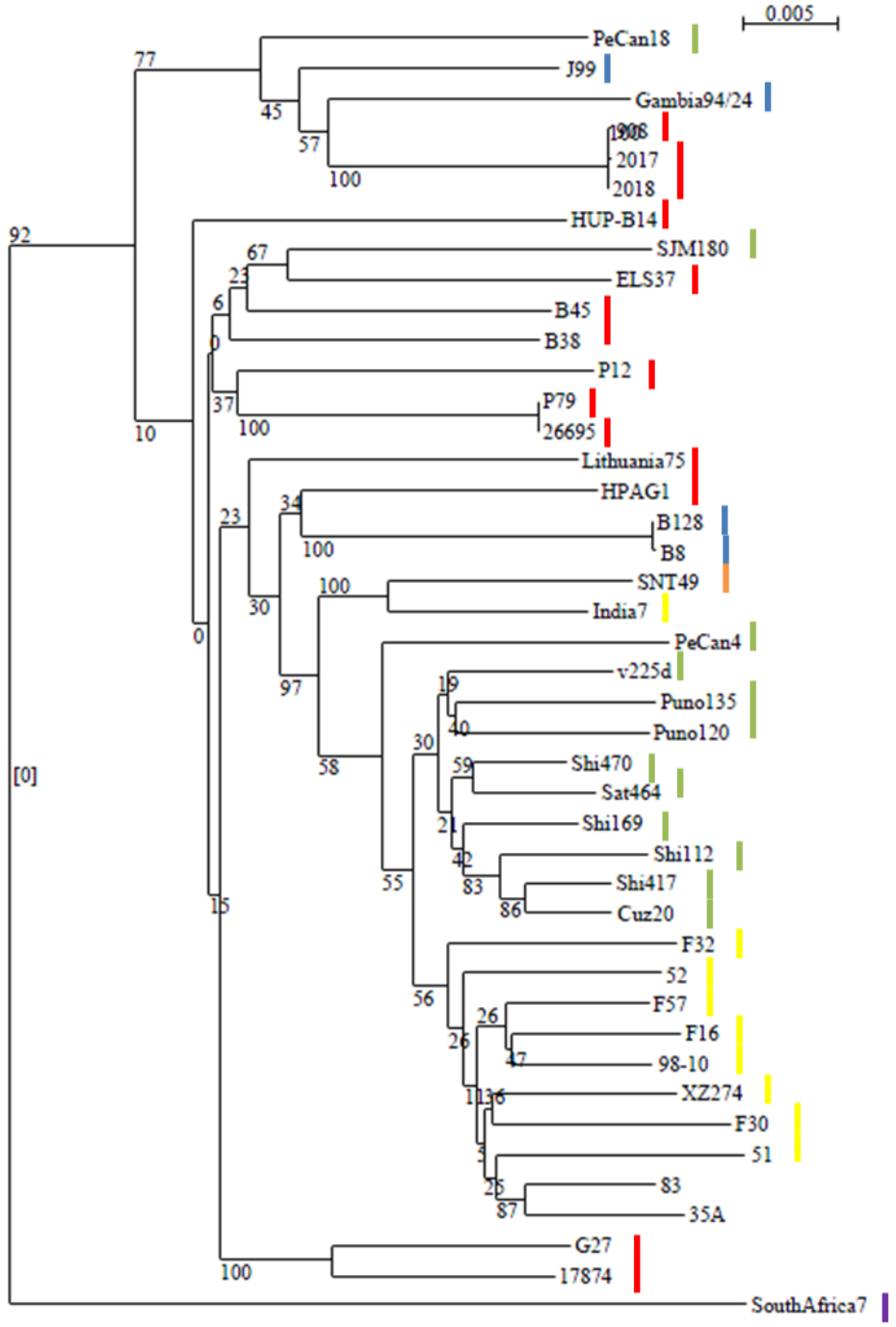


Figure 18 Phylogenetic structure based on MLST analysis of 43 *H. pylori* strains.

Neighbour-joining tree illustrating clustering of strains by geographical location where: red = Europe; blue = North America; green = South America; brown = Asia; yellow = East Asia and purple = Africa. Bootstrap values (100 replicates) are listed on each branch.

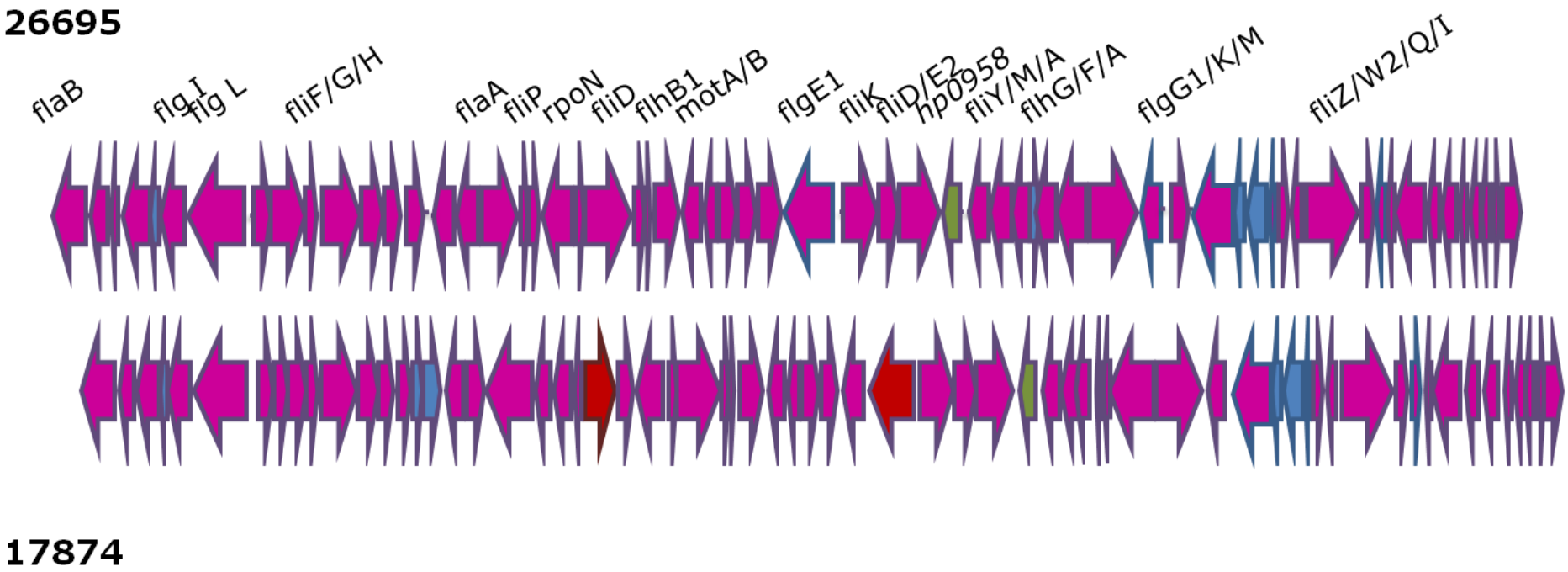


Figure 19 Flagellar gene organisation of *H. pylori* CCUG 17874 and P79 based on reference strain 26695.

Distribution of flagellar genes in both strains is across the genome, represented here as a single locus for illustrative purposes where: purple = flagellar genes; blue = non-flagellar genes; and red = genes whose size is estimated due to lack of sequencing coverage.

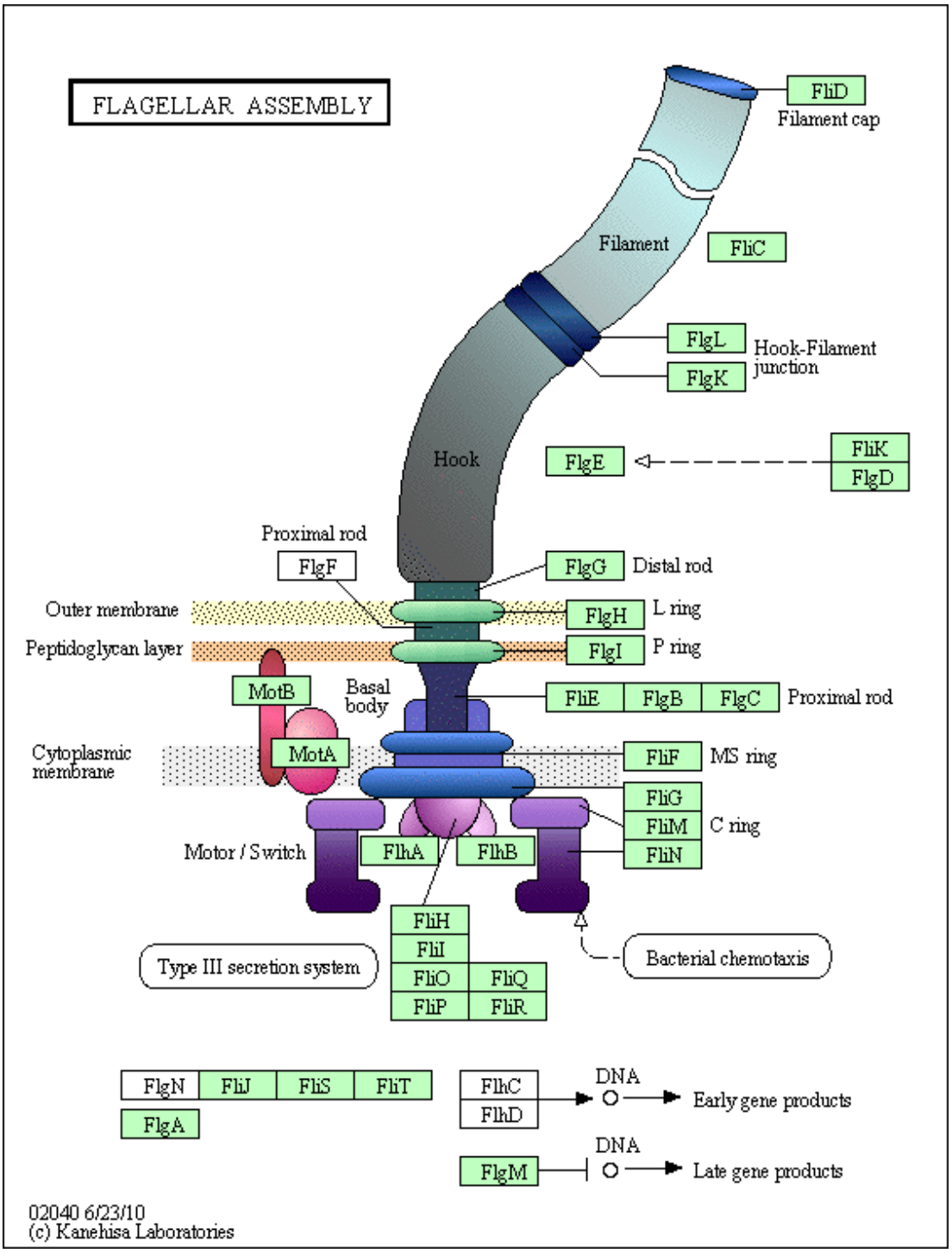


Figure 20 Flagellar genes present in the genomes of *H. pylori* CCUG 17874 and P79.

Image generated by KEGG Automatic Annotation Server bi-directional best hit BLAST against a database of publicly available *H. pylori* genomes. Map of *H. pylori* CCUG 17874 and P79 are identical. Image based on flagellum of *Salmonella enterica* where green = present and white = absent.

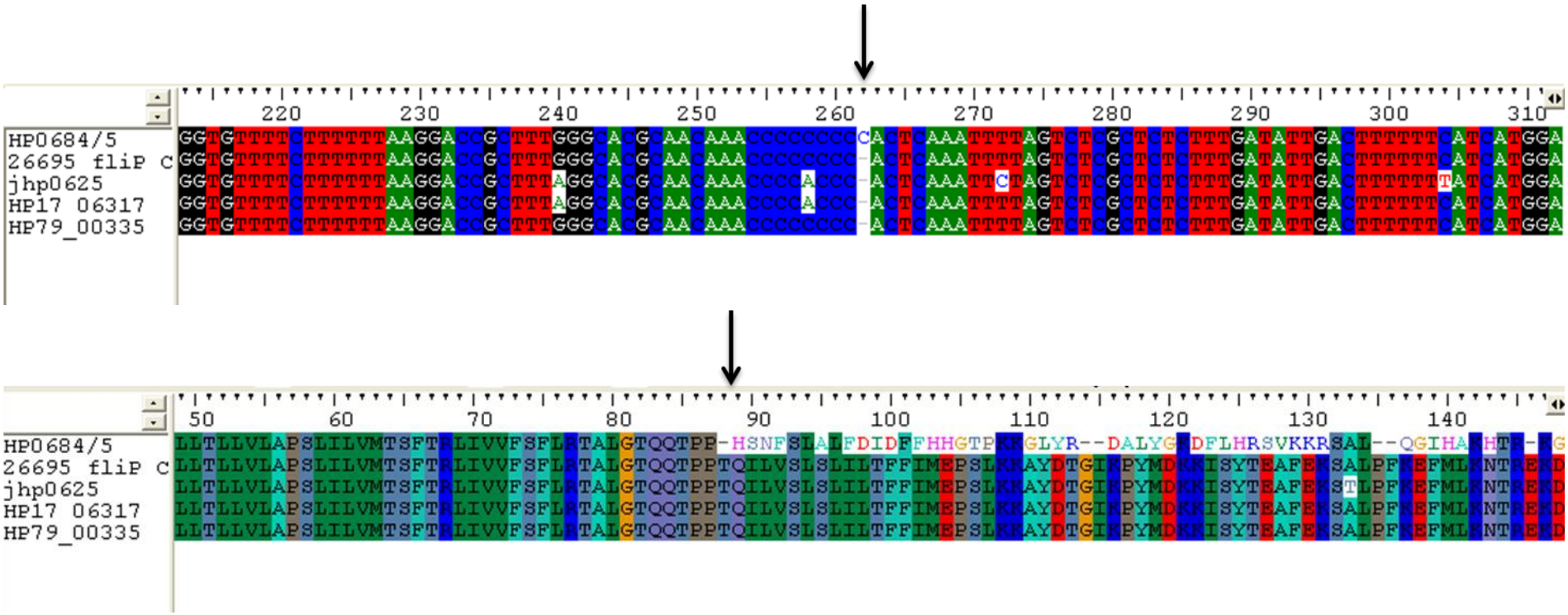


Figure 21 ClustalW multiple sequence alignment of *flip* nucleotide and translated amino acid sequences.

Sequences of the *fliP* gene of reference strains *H. pylori* 26695 and J99 compared to those of 17874 and P79. "26695 *fliP* C" is the altered sequence of *flip* from 26695 where one C has been deleted in the homopolymeric tract to illustrate the frameshift caused by this phase variable tract.

7 References

Algood HMS, Cover TL. 2006. *Helicobacter pylori* persistence: an overview of interactions between *H. pylori* and host immune defenses. *Clin. Microbiol. Rev.* 19: 597–613.

Alm RA, Ling LL, Moir DT, King BL, *et al.* 1999. Genomic sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 397: 186–190.

Alm RA, Trust TJ. 1999. Analysis of the genetic diversity of *Helicobacter pylori*: the tale of two genomes. *J. Mol. Med.* 77: 834–46.

Ando T, Israel DA, Kusugami K, Blaser MJ. 1999. HP0333, a member of the *dprA* family, is involved in natural transformation in *Helicobacter pylori*. *J. Bacteriol.* 181: 5572–80.

Ando T, Xu Q, Torres M, Kusugami K, *et al.* 2000. Restriction-modification system differences in *Helicobacter pylori* are a barrier to interstrain plasmid transfer. *Mol. Microbiol.* 37: 1052–65.

Atherton JC, Cao P, Peek RMJ, Tummuru MKR, *et al.* 1995. Mosaicism in Vacuolating Cytotoxin Alleles of *Helicobacter pylori*. *J. Biol. Chem.* 270: 17771–17777.

Basso D, Plebani M, Kusters JG. 2010. Pathogenesis of *Helicobacter pylori* infection. *Helicobacter* 15 Suppl 1: 14–20.

Blanchard TG, Czinn SJ, Correa P, Nakazawa T, *et al.* 2013. Genome sequences of 65 *Helicobacter pylori* strains isolated from asymptomatic individuals and patients with gastric cancer, peptic ulcer disease, or gastritis. *Pathog. Dis.* 68: 39–43.

Blaser MJ. 1997. Perspectives Series: Host/pathogen interactions; ecology of *Helicobacter pylori* in the human stomach. *J. Clin. Invest.* 100: 759–762.

Blattner FR. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277: 1453–1462.

Boissy R, Ahmed A, Janto B, Earl J, *et al.* 2011. Comparative supragenomic analyses among the pathogens *Staphylococcus aureus*, *Streptococcus pneumoniae*, and *Haemophilus influenzae* using a modification of the finite supragenome model. *BMC Genomics* 12: 187-201.

Boneca IG, de Reuse H, Epinat J-C, Pupin M, *et al.* 2003. A revised annotation and comparative analysis of *Helicobacter pylori* genomes. *Nucleic Acids Res.* 31: 1704–1714.

Caly DL, O'Toole PW, Moore SA. 2010. The 2.2-Å structure of the HP0958 protein from *Helicobacter pylori* reveals a kinked anti-parallel coiled-coil hairpin domain and a highly conserved Zn-ribbon domain. *J. Mol. Biol.* 403: 405–19.

Carver TJ, Rutherford KM, Berriman M, Rajandream M-A, *et al.* 2005. ACT: the Artemis Comparison Tool. *Bioinformatics* 21: 3422–3.

Chen PE, Cook C, Stewart AC, Nagarajan N, *et al.* 2010. Genomic characterization of the *Yersinia* genus. *Genome Biol.* 11: R1.

Chevreux B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information. GCB99 (1999), pp. 45-56. *Comput. Sci. Biol.: Proc. Ger. Coference Bioinforma.* 45–56.

Deng W, Liou S, Plunkett III G, George F, *et al.* 2003. Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18 (2003) *J. Bact.* 185: 2330-2337.

Dorer MS, Cohen IE, Sessler TH, Fero J, Salama NR. 2013. Natural competence promotes *Helicobacter pylori* chronic infection. *Infect. Immun.* 81: 209–15.

Dorer MS, Fero J, Salama NR. 2010. DNA damage triggers genetic exchange in *Helicobacter pylori*. *PLoS Pathog.* 6: e1001026.

Douillard FP, Ryan KA, Caly DL, Hinds J, *et al.* 2008. Posttranscriptional regulation of flagellin synthesis in *Helicobacter pylori* by the RpoN chaperone HP0958. *J. Bacteriol.* 190: 7975–84.

Eaton KA, Morgan DR, Krakowka S. 1992. Motility as a factor in the colonisation of gnotobiotic piglets by *Helicobacter pylori*. *J. Med. Microbiol.* 37: 123–7.

Eid J, Fehr A, Gray J, Luong K, *et al.* 2009. Real-Time DNA sequencing from single polymerase molecules. *Science* 323: 133–138.

Ferreira RM, Machado JC, Letley D, Atherton JC, *et al.* 2012. A novel method for genotyping the *Helicobacter pylori vacA* intermediate region directly in gastric biopsy specimens. *J. Clin. Microbiol.* 50: 3983–9.

Fleischmann RD, Adams MD, White O, Clayton RA, *et al.* 1995. Whole-genome random sequencing and assembly of the *Haemophilus influenzae* RD. *Science* 269: 496–512.

Fraser CM, Gocayne JD, White O, Adams MD, *et al.* 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science* 270: 397–403.

Gill SR, Fouts DE, Archer GL, Mongodin EF, *et al.* 2005. Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a *Staphylococcus epidermidis* strain. *J. Bacteriol.* 187: 2426–2438.

Goodwin CS, Armstrong JA, Marshall BJ. 1986. *Campylobacter pyloridis*, gastritis, and peptic ulceration. *J. Clin. Pathol.* 39: 353–365.

Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35: W52–7.

Heuermann D, Haas R. 1998. A stable shuttle vector system for efficient genetic complementation of *Helicobacter pylori* strains by transformation and conjugation. *Mol. Gen. Genet.* 257: 519–28.

Hofreuter D, Odenbreit S, Haas R. 2001. Natural transformation competence in *Helicobacter pylori* is mediated by the basic components of a type IV secretion system. *Mol. Microbiol.* 41: 379–91.

Holt KE, Thomson NR, Wain J, Langridge GC, *et al.* 2009. Pseudogene accumulation in the evolutionary histories of *Salmonella enterica* serovars Paratyphi A and Typhi. *BMC Genomics* 10: 36-47.

Horner DS, Pavesi G, Castrignanò T, De Meo PD, *et al.* 2010. Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. *Brief. Bioinform.* 11: 181–97.

Huimin K, Lin L-F, Porter N, Stickel S, *et al.* 2000. Functional analysis of putative restriction–modification system genes in the *Helicobacter pylori* J99 genome. *Nucleic Acids Res.* 28: 3216–3223.

Humbert O, Dorer MS, Salama NR. 2011. Characterization of *Helicobacter pylori* factors that control transformation frequency and integration length during inter-strain DNA recombination. *Mol. Microbiol.* 79: 387–401.

Jolley KA, Maiden MCJ. 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11:595.

Josenhans C, Eaton KA, Thevenot T, Suerbaum S. 2000. Switching of flagellar motility in *Helicobacter pylori* by reversible length variation of a short homopolymeric sequence repeat in fliP, a gene encoding a basal body protein. *Infect. Immun.* 68: 4598–603.

Joyce EA, Gilbert J V, Eaton KA, Plaut A, Wright A. 2001. Differential gene expression from two transcriptional units in the cag pathogenicity island of *Helicobacter pylori*. *Infect. Immun.* 69: 4202–4209.

Jung SW, Sugimoto M, Shiota S, Graham DY, Yamaoka Y. 2012. The intact *dupA* cluster is a more reliable *Helicobacter pylori* virulence marker than *dupA* alone. *Infect. Immun.* 80: 381–7.

Karnholz A, Hoefler C, Odenbreit S, Fischer W, *et al.* 2006. Functional and topological characterization of novel components of the *comB* DNA

transformation competence system in *Helicobacter pylori*. *J. Bacteriol.* 188: 882–893.

Kersulyte D, Lee W, Subramaniam D, Anant S, *et al.* 2009. *Helicobacter pylori*'s plasticity zones are novel transposable elements. *PLoS One* 4: e6859.

Koren S, Harhay GP, Smith TP, Bono JL, *et al.* 2013. Reducing assembly complexity of microbial genomes with single-molecule sequencing. *Genome Biol.* 14: R101.

Kurtz S, Phillippy A, Delcher AL, Smoot M, *et al.* 2004. Versatile and open software for comparing large genomes. *Genome Biol.* 5: R12.

Lertsethtakarn P, Ottemann KM, Hendrixson DR. 2011. Motility and chemotaxis in *Campylobacter* and *Helicobacter* . *Annu. Rev. Microbiol.* 65: 389–410.

Leunk RD, Johnson PT, David BC, Kraft WG, Mrgan DR. 1988. Cytotoxic activity in broth-culture filtrates of *Campylobacter pylori*. *J. Med. Microbiol.* 26: 93–99.

Lin L, Posfai J, Roberts RJ, Kong H. 2001. Comparative genomics of the restriction-modification systems in *Helicobacter pylori*. *Proc. Natl. Acad. Sci. U. S. A.* 98: 2740–2745.

Linz B, Balloux F, Moodley Y, Manica A, *et al.* 2007. An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* 445: 915–8.

Loman NJ, Constantinidou C, Chan JZM, Halachev M, *et al.* 2012. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat. Rev. Microbiol.* 10: 599–606.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25: 955–64.

Lu H, Hsu P-I, Graham DY, Yamaoka Y. 2005. Duodenal ulcer promoting gene of *Helicobacter pylori*. *Gastroenterology* 128: 833–848.

Lu W, Wise MJ, Tay CY, Windsor HM, *et al.* 2013. Comparative analysis of the full genome of the *Helicobacter pylori* isolate, Sahul64, identifies genes of high divergence. *J. Bacteriol.* 2: doi:10.1128/JB.01021-13.

Macnab RM. 2003. How bacteria assemble flagella. *Annu. Rev. Microbiol.* 57: 77–100.

Marshall BJ, Royce H, Annear D. 1984. Original isolation of *Campylobacter pylori*s from human gastric mucosa. Microbios Lett 25: 83–88.

McClain MS, Shaffer CL, Israel DA, Peek RM, Cover TL. 2009. Genome sequence analysis of *Helicobacter pylori* strains associated with gastric ulceration and gastric cancer. *BMC Genomics* 10: 3-16.

McClelland M, Sanderson KE, Spieth J, Clifton SW, *et al.* 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Lett. to Nat.* 413: 852–856.

Metzker ML. 2010. Sequencing technologies- the next generation. *Nat. Rev. Genet.* 11: 31–46.

Miehlke S, Kirsch C, Agha-Amiri K, Günther T, *et al.* 2000. The *Helicobacter pylori* vacA s1, m1 genotype and cagA is associated with gastric carcinoma in Germany. *Int. J. Cancer* 87: 322–7.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35: W182–5.

Moxon R, Bayliss C, Hood D. 2006. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu. Rev. Genet.* 40: 307–33.

Murata-Kamiya N. 2011. Pathophysiological functions of the CagA oncoprotein during infection by *Helicobacter pylori*. *Microbes Infect.* 13: 799–807.

NCBI. 2013. National Center for Biotechnology Information (NCBI) Genome. Available: http://www.ncbi.nlm.nih.gov/genome/genomes/169 [accessed 26 September 2013].

Ohnishi N, Yuasa H, Tanaka S, Sawa H, *et al.* 2008. Transgenic expression of *Helicobacter pylori* CagA induces gastrointestinal and hematopoietic neoplasms in mouse. *Proc. Natl. Acad. Sci. U. S. A.* 105: 1003–8.

Olbermann P, Josenhans C, Moodley Y, Uhr M, *et al.* 2010. A global overview of the genetic and functional diversity in the *Helicobacter pylori* cag pathogenicity island. *PLoS Genet.* 6: e1001069.

Quevillon E, Silventoinen V, Pillai S, Harte N, *et al.* 2005. InterProScan: protein domains identifier. *Nucleic Acids Res.* 33: W116–20.

Rhead JL, Letley DP, Mohammadi M, Hussein N, *et al.* 2007. A new *Helicobacter pylori* vacuolating cytotoxin determinant, the intermediate region, is associated with gastric cancer. *Gastroenterology* 133: 926–36.

Romo-González C, Salama NR, Burgeño-Ferreira J, Ponce-Castañeda V, *et al.* 2009. Differences in genome content among *Helicobacter pylori* isolates from patients with gastritis, duodenal ulcer, or gastric cancer reveal novel disease-associated genes. *Infect. Immun.* 77: 2201–11.

Rutherford K, Parkhill J, Crook J, Horsnell T, *et al.* 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16: 944–5.

Saunders NJ, Peden JF, Hood DW, Moxon ER. 1998. Simple sequence repeats in the *Helicobacter pylori* genome. *Mol. Micro.* 27: 1091–1098.

Sharma CM, Hoffmann S, Darfeuille F, Reignier J, *et al.* 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* 464: 250–5.

Smeets LC, Bijlsma JJE, Boomkens SY, Vandenbroucke-Grauls CMJE, *et al.* 2000. *comH* , a novel gene essential for natural transformation of *Helicobacter pylori*. *J. Bacteriol.* 182: 3948–3954.

Stingl K, Müller S, Scheidgen-Kleyboldt G, Clausen M, Maier B. 2010. Composite system mediates two-step DNA uptake into *Helicobacter pylori*. *Proc. Natl. Acad. Sci. U. S. A.* 107: 1184–9.

Suerbaum S, Josenhans C. 2007. *Helicobacter pylori* evolution and phenotypic diversification in a changing host. *Nat. Rev. Microbiol.* 5: 441–52.

Suzek BE, Ermolaeva MD, Schreiber M, Salzberg SL. 2001. A probabilistic method for identifying start codons in bacterial genomes. *Bioinformatics* 17: 1123–30.

Tomb JF, White O, Kerlavage AR, Clayton RA, *et al.* 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388: 539–47.

Uemura N, Okamoto S, Yamamoto S, Matsumura N, *et al.* 2001. *Helicobacter pylori* infection and the development of gastric cancer. *N. Engl. J. Med.* 345: 784–789.

van der Veen B, O'Toole PW, Claesson MJ. 2014. METAPHORE- Automated bidirectional best hit homology analyses. *submitted.*

Wang G-Q, Xu J-T, Xu G-Y, Zhang Y, *et al.* 2013. Predicting a novel pathogenicity island in *Helicobacter pylori* by genomic barcoding. *World J. Gastroenterol.* 19: 5006–10.

Wroblewski LE, Peek RM. 2011. Targeted disruption of the epithelial-barrier by *Helicobacter pylori*. *Cell Commun. Signal.* 9: 1–11.

Xu Q, Morgan RD, Roberts RJ, Blaser MJ. 2000. Identification of type II restriction and modification systems in *Helicobacter pylori* reveals their substantial diversity among strains. *Proc. Natl. Acad. Sci. U. S. A.* 97: 9671–6.

Yeh Y-C, Chang K-C, Yang J-C, Fang C-T, Wang J-T. 2002. Association of metronidazole resistance and natural competence in *Helicobacter pylori*. *Antimicrob. Agents Chemother.* 46: 1564–1567.

Yeh Y-C, Lin T-L, Chang K-C, Wang J-T. 2003. Characterization of a ComE3 homologue essential for DNA transformation in *Helicobacter pylori*. *Infect. Immun.* 71: 5427–5431.

Yoshiyama H, Nakazawa T. 2000. Unique mechanism of *Helicobacter pylori* for colonizing the gastric mucus. *Microbes Infect.* 2: 55–60.

Zhang X-S, Blaser MJ. 2012. Natural transformation of an engineered *Helicobacter pylori* strain deficient in type II restriction endonucleases. *J. Bacteriol.* 194: 3407–16.

Chapter 3
Structure-function analysis of the H. pylori flagellum biogenesis protein HP0958

In preparation for submission to the journal Microbiology.

The work in this chapter includes contributions from other individuals:

- Heatplots described in Section 3.10 were generated by Dr. Ian Jeffery
- Statistical analyses and boxplots described in Section 3.14 were performed by Hugh Harris.

1 Abstract

Background: Motility is an essential feature of *Helicobacter pylori* infection. A yeast two-hybrid study investigating the proteome of *H. pylori* 26695 previously identified that the flagellum biogenesis protein HP0958 interacts with flagellar proteins FliH and RpoN (σ^{54}). HP0958 also interacts with the *flaA* mRNA transcript and may have a regulatory role in flagellum construction.

Materials and Methods: A panel of site-directed mutants of HP0958 was generated in order to elucidate the mechanisms of HP0958 function. A *hp0958*-null derivative strain of P79 was complemented with *hp0958* mutant alleles. GST pull-down, yeast two-hybrid and PXG assays were performed to investigate HP0958-FliH and HP0958-RpoN interactions. HP0958-*flaA* mRNA interaction was analysed by electrophoretic mobility shift assay.

Results: The previously reported HP0958-FliH (89-258) interaction could not be substantiated. Further, RpoN (74-210) also failed to interact with HP0958 at a detectable level when investigated using pull-down assay. The HP0958-RpoN (74-284) interaction was confirmed but was relatively weak by quantitative analysis yeast two-hybrid assay. Complementation of the *hp0958*-null P79 derivative with mutant alleles revealed that mutations in the coiled-coil have a more pronounced effect on motility than those in the zinc-finger. Many mutant derivative strains produced atypical flagellar extrusions from the cells at non-polar sites.

Conclusions: HP0958 does not interact with FliH. Residues 74-284 of RpoN are required for interaction with HP0958, predominantly along the coiled-coil domain. The zinc-finger domain of HP0958 is critical for interaction with the *flaA* mRNA transcript. We propose a novel function of HP0958 in localisation of flagellum biogenesis to the cell pole.

2 Introduction

Helicobacter pylori has been closely associated with humans throughout their evolution (Linz *et al.*, 2007). It currently infects approximately half of the global population, with higher prevalence in Asian and African countries (Linz *et al.*, 2007). Colonisation with this opportunistic pathogen is associated with many effects on the host, some positive and but mostly negative. Typically in later life, *H. pylori* infection can lead to development of duodenal and gastric ulcers, gastric cancer and MALT lymphoma in humans and *H. pylori* was identified as a Class I pathogen in 1994 (International Agency for Research on Cancer, 1994; Jemal *et al.*, 2011; Marshall and Warren, 1984; Pounder and Ng, 1995; The Eurogast Study Group, 1993; Warren and Marshall, 1983).

Motility is a key feature of *H. pylori* infection and is essential for colonisation (Eaton *et al.*, 1992). Flagellum biogenesis is a hierarchical and highly regulated process. In *H. pylori*, regulation of this process differs from that of the well described model systems of flagellum construction *e.g. Salmonella enterica* and *E. coli* (Anderson *et al.*, 2010; Chevance and Hughes, 2008; McCarter, 2006; Niehus *et al.*, 2004). Flagellar genes can be subdivided into three classes, the expression of which is under the control of specific sigma factors. Sigma 80 regulates the expression of Class I (*early)* genes which encode regulators and components of the basal body. RpoN (σ^{54}) control expression of Class II (*middle*) genes which encode components of the rod and hook, while σ^{28} controls expression of Class III (*late*) genes which encode the major filament protein FlaA (Niehus *et al.*, 2002).

HP0958 was identified as a hypothetical protein of unknown function in the genome of *H. pylori* 26695 (Tomb *et al.*, 1997). It was since identified as an essential component of flagellar construction, because inactivation of this gene generated aflagellate, non-motile cells (Pereira and Hoover, 2005; Ryan *et al.*, 2005a). Insertional mutation of the *hp0958* gene of *H. pylori* strain CCUG 17874 resulted in reduced levels of RpoN and lowered expression of Class II flagellar genes including *flgE* and *flaB,* deeming HP0958 a chaperone of RpoN. HP0958 also interacts with the mRNA transcript of the major flagellin-encoding gene, *flaA,* at a post-transcriptional level (Douillard *et al.*, 2008). The crystal structure of HP0958 revealed an N-terminal anti-parallel α-helical coiled-coil and a C-terminal zinc-finger domain (Caly *et al.*, 2010). Initial structure-function analysis of the HP0958-*flaA*

mRNA interaction indicated that the zinc-finger of HP0958 is involved in RNA binding (Caly *et al.*, 2010). However, little is known about what region of the mRNA transcript is required for this interaction.

In 2001, the protein-protein interaction map of *H. pylori* strain 26695 was predicted using yeast two-hybrid screens, identifying 1,200 potential interactions (Rain *et al.*, 2001). PIMRider was developed by Hybrigenics to view and analyse the output of the Rain study, and is accessible online (http://pim.hybrigenics.com). This study covered 46% of the proteome, including the predicted interaction network of HP0958. Statistically significant interactions were identified between HP0958 and the flagellar proteins RpoN and FliH, the negative regulator of FliI ATPase, as well as a number of other proteins of lower probability scores (Rain *et al.*, 2001). Douillard *et al.* proposed a model of the role of HP0958 in flagellum biogenesis. This model suggests that HP0958 acts as a chaperone to RpoN during the expression of Class II flagellar genes; upon the switch in specificity to Class III genes, HP0958 acts to guide the *flaA* transcript to the export apparatus through its interaction with FliH (Douillard *et al.*, 2008). However, the mechanism of binding in HP0958-FliH and HP0958-RpoN interactions has not been investigated at a structural level.

Protein-protein interactions (PPIs) are essential for cellular function. Transient PPIs, although short-lived, are extremely important for a variety of biological processes *e.g.* signalling cascades and transcription factors (Hahn and Kim, 2012; Ozbabacan *et al.*, 2011). PPIs can be detected using a number of biochemical and computational means including pull-down assay (Fields and Song, 1989; Geva and Sharan, 2011; Lane *et al.*, 2006; Stynen *et al.*, 2012; Tang and Bruce, 2009; Xia *et al.*, 2010; Zhang *et al.*, 2012). While the yeast two-hybrid system allows proteome analysis of a subject, it has a number of shortcomings. High rates of false-positives, incomplete coverage of the entire interactome, and the use of a eukaryotic system to investigate bacterial protein-protein interactions are limiting factors of this method (Stynen *et al.*, 2012); nevertheless, it is a valuable high throughput tool in identifying PPIs. The yeast two-hybrid performed on the proteome of *H. pylori* 26695 (Rain *et al.*, 2001) provided a valuable data set which can be used to create a more complete understanding of flagellum biogenesis in *H. pylori*. This study focused on the interactions of motility protein HP0958 with other flagellum biogenesis components including RpoN and FliH, identified from the previous yeast two-hybrid study.

3 Methods

3.1 Bacterial Strains and Culture Conditions

The bacterial and yeast strains used in this study are listed in Table 18. *H. pylori* strains were grown on Columbia Base Agar (CBA) solid medium, supplemented with 5% v/v heat-inactivated, defibrinated horse blood (Cruinn, Ireland) at 37°C, 5% CO_2 and sub-cultured every two days. For broth culture, cells were grown in brain heart infusion (BHI) broth (Sigma) supplemented with heat-inactivated foetal bovine serum (Sigma) and gently agitated in a microaerobic environment for 20 hrs. Mutant derivatives of *H. pylori* strain P79 were supplemented with chloramphenicol (10 µg/ml) and kanamycin (25 µg/ml) where required.

E. coli XLI-Blue Supercompetent cells (Stratagene, Agilent Technologies) were used as the host for molecular cloning of HP0958 site-directed mutants; *E. coli* Top 10 (Invitrogen, Carlsbad, CA) was used as the cloning system in all other cases. Proteins were over-expressed in *E. coli* Rosetta (Novagen, Darmstadt, Germany). *E. coli* XL1-Blue Supercompetent cells were grown in NZY^+ broth at 37°C with agitation. All other *E. coli* strains were cultured in Luria-Bertani (LB) media at 37°C or 18°C with agitation. Media was supplemented with ampicillin (100 µg/ml), erythromycin (50 µg/ml) and chloramphenicol (34 µg/ml) where required.

3.2 Molecular Cloning

All flagellar genes were amplified from *H. pylori* CCUG 17874 (Culture Collection University of Gothenburg, Gothenburg, Sweden). Genomic DNA was extracted from two-day old plates using DNeasy Blood and Tissue DNA Extraction Kit (Qiagen, Hilden, Germany) as previously described (Douillard *et al.*, 2008). PCR was performed on genomic template DNA using the primers listed in Appendix 16 at standard conditions for Velocity (Bioline, UK) and Taq DNA Polymerase (New England Biolabs, UK). PCR amplicons were cloned into restriction digested vectors and transformed into chemically competent *E. coli* host cells. In all cases, positive clones were selected through propagation on relevant agar supplemented with appropriate antibiotics (Table 18) and screened by colony PCR. Plasmid DNA was extracted from *E. coli* cells using Qiaprep Spin Miniprep kit (Qiagen, Hilden, Germany). DNA concentration and quality was estimated using Nanodrop 2000

(Thermo Scientific). Correct constructs were confirmed by sequencing performed by Eurofins MWG Operon (Ebersberg, Germany).

Yeast strains were made competent using a standard lithium acetate procedure and transformed with the relevant plasmids (Table 19) in the presence of salmon sperm carrier DNA according to the Clontech Yeast Protocols Handbook (Clontech Laboratories, USA). Briefly, 100 µg competent cells were incubated with 100 ng of the relevant plasmids and 0.1 mg salmon sperm DNA in the presence of 0.6 ml sterile polyethyleneglycol (PEG) 3350 and 1 X Tris-EDTA (TE) lithium acetate. Cells were incubated at 30ºC for 30 min, shaking. To each tube, 70 µl dimethyl sulfoxide (DMSO) was added, cells were heat shocked at 42ºC for 15 min and chilled on ice briefly before pelleting cells and resuspending in 500 ml 1 X TE. Cells were plated on appropriate media and incubated at 30ºC for 2 - 5 days. Positive clones were selected for through propagation on relevant drop out base YPD agar lacking different combinations of the amino acids tryptophan, leucine, adenine and histidine.

3.3 Site-Directed Mutagenesis

Point mutations of selected amino acids were generated using Quikchange II Site-Directed Mutagenesis kit (Stratagene, Agilent Technologies). Primers were designed according to the manufacturer's recommendations (Appendix 17) and synthesized by MWG Biotech (Ebersberg, Germany). Rationale for selection of targets within HP0968 to be mutated is described in Appendix 18. Plasmid DNA was isolated from an *E. coli* Top10 strain carrying the pDC006 plasmid using the Qiaprep® Spin Miniprep kit (Qiagen, Hilden, Germany). The *hp0958* gene present on pDC006 was used as the template DNA for mutagenesis. *Pfu* DNA polymerase amplified site-directed mutants from 10 ng plasmid DNA by thermal cycling as previously described (Caly *et al.*, 2010). *Dpn* I restriction digestion at 37°C for 1 hr was performed to remove template DNA. The resulting single-stranded plasmids were transformed into XL1-Blue Supercompetent cells and plasmids containing the correct mutation were screened by insert sequencing.

3.4 Allelic Exchange Mutagenesis

All genes were amplified from *Helicobacter pylori* CCUG 17874 using primers listed in Appendix 17 (manufactured by Eurofins MWG Operon (Germany)) and

standard Velocity polymerase cycling parameters. The promoter region of the alkyl hydroperoxide reductase (*ahpC*) gene, *php1563* was amplified to produce DNA with a 3' overhang complementary to the 5' of *hp0958.* Similarly, genes encoding wild-type or site-directed mutants of HP0958 were amplified to incorporate a 5' overhang complementary to the 3' of *php1563*. Splicing by overlapping extension (SOE) PCR was used to generate a single fused product *php1563_hp0958* as previously described (Douillard *et al.,* 2008; Heckman and Pease, 2007). SOE PCR products were ligated to shuttle vector pIR203K04 (a kind gift from D. J. McGee) following *Bam*HI and *Cla*I restriction digestion. This plasmid harbours a kanamycin resistance cassette and was designed to introduce DNA fragments into the intergenic region of *H. pylori* between genes *hp0203* and *hp0204* (Langford *et al.,* 2006).

3.5 Natural Transformation of *H. pylori*

Generation of the *hp0958* deletion mutant, *H. pylori* P79-0958KO, was previously described (Ryan *et al.,* 2005). *E. coli* Top10 was used as a cloning host before transformation into *H. pylori.* All constructs were confirmed by sequencing performed by Eurofins MWG Operon (Germany) and GATC (Germany). *H. pylori* P79-0958KO cells were transformed with shuttle vector pIR203K04 harbouring either wild type or site-directed mutants of *hp0958,* see Table 20 for details. Briefly, *H. pylori* P79-0958KO cells from one full 48 hr-old CBA plate were harvested in BHI broth supplemented with 0.5% FBS. The OD_{600} was corrected to 0.4 - 0.6 and recipient cells were incubated with 2 - 5 µg plasmid DNA at 37°C, 5% CO_2 for 2 hrs. The mixture was then plated on non-selective CBA. After 24 hours, cells were harvested in BHI broth and transferred to CBA agar supplemented with chloramphenicol (10 µg/ml) and kanamycin (25 µg/ml) and incubated for 3 - 4 days. Transformants were screened by motility assay and colony purified. Colony PCR and sequencing of the *hp0203-0204* intergenic region confirmed integration.

3.6 Motility Assay

Freshly prepared BHI soft agar plates containing 0.3% (w/v) agar supplemented with 10% heat-inactivated FBS (Sigma) and antibiotics, where appropriate, were inoculated with *H. pylori* strains and mutants. Cells from 48 hr-old CBA plates were harvested in BHI broth and OD_{600} was corrected to 0.4 - 0.6. Cells (5 µl) were

stabbed into the centre of each motility plate and incubated at 37°C, 5% CO_2. Plates were imaged after 4 days using the Gene Genius Bio-Imaging System (Syngene).

3.7 Electron Microscopy

Flagellum morphology was determined using transmission electron microscopy (TEM) to observe negatively stained *H. pylori* cell preparations. Liquid cultures were grown for 20 hrs and fixed with 2.5% glutaraldehyde solution (Sigma). Cells were allowed to sediment overnight, and gently resuspended in fresh 2.5% glutaraldehyde solution. One drop containing ~5 x 10^6 cells was applied to the surface of Formvar carbon-coated 200 mesh copper grids (Electron Microscopy Sciences, UK). Grids were quickly rinsed with H_2O and stained with 2% uranyl acetate (Sigma). Imaging was performed using a FEI Tecnai 120 transmission electron microscope operating at 120 kV (Biological Imaging Facility, Conway Institute of Biomolecular and Biomedical Research, University College Dublin).

3.8 Preparation of Whole Cell Fractions

H. pylori cells were harvested from 20 hr liquid cultures and pelleted at 13,000 rpm for 15 s. Pellets were washed with 1 ml sterile phosphate-buffered saline (PBS) and pelleted again. Supernatant was removed and pellets were resuspended gently in 500 µl fresh PBS. All cultures were corrected to an OD_{600} of 1.0 and centrifuged at 13,000 rpm for 15 s. Pellets were resuspended in Laemmli sample buffer, boiled at 100°C for 5 min and stored at -80°C.

3.9 Protein Electrophoresis and Western Blot

Standard protocols were used to separate and visualise proteins by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) (Sambrook *et al.*, 1989). Proteins were separated on 12.5% SDS acrylamide gels and transferred onto polyvinylidine fluoride (PVDF) membrane by electroblotting for 1 hr (Towbin *et al.*, 1979). Anti-hook and anti-flagellin polyclonal antibodies were used as primary antibodies during western blotting of *H. pylori* whole cell fractions (Kostrzynska *et al.*, 1991; O′Toole *et al.*, 1994). Anti-rabbit antibody raised in goat was coupled to horseradish-peroxidase (Sigma) and was used as the secondary antibody (Douillard *et al.*, 2008). Detection was performed with 4-chloro-1-nathphol and hydrogen peroxide.

3.10 Quantitative Analysis of Transcription by Real-Time PCR

Quantitative real-time PCR (qRT)-PCR was performed as described previously using primers designed with Primer 3 software (Appendix 19) (Douillard *et al.*, 2008; Untergasser *et al.*, 2012). Cells were grown in BHI broth supplemented with 10% FBS for 20 hrs and harvested in Bacteria RNA Protect (Qiagen). Cells were washed with PBS and lysed by bead-beating in Trizol® reagent (Ambion). RNA was purified using RNeasy Protect Bacteria Mini Kit (Qiagen) according to manufacturer's instructions and DNase-treated to remove residual DNA using TURBO DNA-*Free* (Ambion) as instructed. RNA was quantified by Nanodrop 2000 (Thermo Scientific) and RNA quality was assessed using Bioanalyser 2100 (Agilent Technologies) as directed by manufacturer. 200 ng of RNA was reverse transcribed to cDNA using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). All cDNA was diluted 50-fold before use in qRT-PCR. Briefly, 5 µl cDNA, 2 µl of 5 µM primer mix, 10 µl 2x mastermix (including Syber Green I polymerase) and 3 µl H_2O were mixed and qRT-PCR was performed in Roche LightCycler® 480 II. Reactions were performed in triplicate on at least three biological replicates and data was normalised to the *era* housekeeping gene. Relative fold-changes in gene expression were calculated as previously described (Pfaffl, 2001). Heat plots of normalised flagellar gene expression were generated, ranking strains according to *flaB* and *flgE* expression levels.

3.11 Protein Over-Expression and Purification

Proteins used in pull-down assay were expressed with an N-terminal glutathione sepharose (GST) tag and purified affinity purified as previously described (Caly *et al.*, 2010). An 8 residue N-terminal FLAG-tag (DYKDDDDK) was fused to bait proteins to facilitate immunoblotting. *E. coli* strains possessing the relevant plasmids (Table 19) were grown to OD_{600} 0.4 - 0.6 and protein expression was induced with 0.1 mM isopropylthiogalactoside (IPTG) for 16-20 hours at 18°C. Cells were harvested and lysed by passage through a French Press twice at 1,000 psi.

The soluble cytoplasmic fraction was incubated with Glutathione Sepharose 4B (GE Healthcare, UK) for 16 hours at 4°C. Purified proteins were released using PreScission protease (GE Healthcare, UK) in cleavage buffer (50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol (DTT), pH 7.5) at 4°C for 16 hrs

with gentle agitation. For elution of intact fusion proteins, resin was incubated with elution buffer containing 10 mM reduced glutathione according to the manufacturer's instructions. Eluted proteins were concentrated (≤ 2 mg/ml) and buffers dialysed using Amicon Centrifugal Filter Units (Millipore, Billerica, MA).

Proteins were further purified by anion exchange in a starting buffer of 20 mM ethanolamine, pH 9.0 on HiTrap FF 1 ml columns (GE Healthcare, UK) attached to an Äkta Purifier. Eluted purified proteins were concentrated and quantified using the Pierce Bicinchoninic Acid (BCA) assay (Thermo Scientific, USA) (Smith *et al.*, 1985).

3.12 GST Pull-Down Assay

The GST pull-down assay was adapted from Lane *et al.*, 2006 (Lane *et al.*, 2006). Briefly, 30 µl glutathione sepharose 4B was washed 4 times with 100 µl PBS at 4°C. GST-tagged bait protein (60 µg) was bound to resin in a total volume of 200 µl at room temperature for 30 min with gentle agitation (Table 19). Resin-bound protein was washed 4 times and incubated with FLAG-tagged prey protein at various prey : bait molar ratios. Samples were incubated at room temperature for 30 min with gentle agitation. Samples were washed twice with 100 µl PBS 0.5% Tween 20, 250 mM NaCl. Laemmli buffer was added to resin and samples were boiled for 5 min at 100°C, run on 12.5% SDS-PAGE gels and transferred onto PVDF membrane. A horseradish peroxidase coupled anti-FLAG monoclonal antibody was used to detect prey proteins. The membrane was incubated with Enhanced Chemiluminescence (ECL) Western Blot Detection Reagents and developed on Hyperfilm in darkness according to manufacturer's instructions (GE Healthcare, UK).

3.13 Yeast Two-Hybrid Assay

Saccharomyces cerevisiae strains AH109 and Y187 were used as hosts for yeast two hybrid assay (Y2H) (Table 19). *S. cerevisiae* wild type strains were grown on yeast extract peptone dextrose (YPD) agar or broth and supplemented with 0.003% (v/v) adenine-2-hemisulphate. *S. cerevisiae* strains possessing bait vector pGBKT7 were selected for on synthetically defined (SD) media lacking tryptophan; strains possessing prey vector pGADT7 were selected for on SD lacking leucine. SD media was supplemented with the following amino acids: 0.3 µM adenine-2-hemisulphate,

0.3 µM L-histidine-HCl, 1.67 µM L-leucine and 0.4 µM L-tryptophan. Cells were grown at 30ºC with agitation.

The Y2H strategy was based on the Clontech Matchmaker™ Gold Yeast Two Hybrid system (Clontech Laboratories, USA). See Table 19 and Appendix 16 for list of plasmids and primers used. Cells were made competent by standard lithium acetate procedure and transformed with relevant plasmids to investigate a given PPI. Transformed cells were plated on SD drop-out base with relevant amino acid supplements. Transformants were counted after 2 - 5 days and colony purified.

3.14 Plate X-Gal Assay

Y187-derivative strains possessing both bait and prey vectors were selected on SD media lacking tryptophan and leucine. Plate X-gal (PXG) assay was adapted from *Möckli et Auerbach* to assess protein-protein interactions (PPI) through activation of histidine-encoding reporter gene expression (Möckli and Auerbach, 2004). Five biological replicates per strain were assayed in triplicate. SD-T-L broth was inoculated at a starting $OD_{546} < 0.1$. Cells were grown at 30°C with agitation to an OD_{546} of 0.8 - 1. One absorbance unit of cells was transferred to a 96-well round-bottomed plate and pelleted. Cell lysis was achieved by 2 freeze thaw cycles: 3 minutes submerged in liquid nitrogen, 3 minutes at 37°C. Lysed pellets were resuspended in 20 µl sterile H_2O and transferred to a 96-well flat bottomed plate. Cells were incubated with 100 µl PBS, pH 7.4 containing 500 µg/ml X-gal, 0.3% (w/v) agarose and 0.05% (v/v) β-mercaptoethanol. Plates were incubated at room temperature in darkness. Time points were taken using a flatbed scanner and analysed by densitometry with ImageJ online software (Abramoff *et al.*, 2004).

3.15 RNA Secondary Structure Prediction and *in vitro* Transcription

Secondary structure prediction analysis of the full length *flaA* mRNA transcript was performed using RNAdraw which predicts structure based on McCaskill minimum free energy (Matzura and Wennborg, 1996) and RDM Circles which is based on maximum weight matching (Page, 2000). Truncated transcripts (regions 1, 2 and 3) were designed using RNAdraw and generated using the primers listed in Appendix 20. SOE-PCR was performed to generate the region 1 truncation of *flaA* mRNA which required deletion of the central portion of the transcript. PCR templates for *in vitro* transcription were concentrated using the Minielute PCR

Purification Kit (Qiagen). Biotin-labelled RNA was synthesised from an artificially fused 5' T7 polymerase binding site using the Riboprobe System T7 kit (Promega) and Biotin RNA Labelling Mix (Roche) as previously described (Caly *et al.*, 2010). Transcripts were DNase treated for 15 min at 37°C and concentrated by phenol/chloroform extraction followed by ethanol precipitation. Quality of RNA was assessed by agarose gel electrophoresis in a 3-(N-morpholino) propanesulfonic acid (MOPS) buffer followed by post-staining with ethidium bromide and imaging using the Gene Genius Bio-Imaging System (Syngene).

3.16 Electrophoretic Gel Migration Shift Assay

Electrophoretic gel migration shift assay (EMSA) was performed to investigate the nature of the interaction between HP0958 and *flaA* mRNA transcripts as previously described (Caly *et al.*, 2010). Briefly, 6 µg of purified HP0958 was incubated with 15 ng of full length *flaA* riboprobe (8 ng region 1/5 ng region 2/3.2 ng region 3) in a final volume of 15 µl binding buffer containing 20 mM Tris-acetate (pH 7.9), 50 mM potassium acetate, 10 mM magnesium acetate, 1 mM dithiothreitol (DTT), 200 µM S-(5-adenosyl)-L-methionine chloride (Sigma), and 40 U RNasin (Promega, USA). Samples were incubated at room temperature for 10 min followed by 5 minutes at 37°C. The RNA was then resolved by native agarose gel electrophoresis.

3.17 Northern Blotting

Resolved RNA was transferred from agarose gel to a Biodyne B nylon membrane (GE Healthcare, UK) overnight by capillary transfer. The membrane was rinsed with 5 X sodium chloride-sodium citrate (SSC) buffer and RNA was cross-linked to the membrane by UV-cross-linking at 120 mJ using the Stratalinker UV-linker (Stratagene, USA). The membrane was washed with Odyssey blocking buffer (Li-COR Biosciences) 1% SDS for 30 minutes at room temperature with gentle agitation. The blot was then incubated with Odyssey blocking buffer 1% SDS supplemented with streptavidin IRDye 680 (diluted 1:10,000) (Li-COR Biosciences) for 30 min at room temperature. The membrane was washed 3 times with PBS 0.1% Tween 20 and once with PBS before imaging using the Odyssey Infrared Imaging System (Li-COR Biosciences).

4 Results

4.1 The previously reported HP0958-FliH interaction could not be substantiated.

In 2001, a study was published in *Nature* which predicted by yeast two-hybrid that HP0958 interacts with the flagellar protein FliH, in addition to the sigma factor RpoN (Rain *et al.*, 2001). GST pull-down assay was performed with soluble proteins which were expressed in *E. coli* and purified by affinity and anion exchange chromatography. Concentrated, purified proteins were assessed by SDS-PAGE (data not shown). GST-tagged bait proteins were bound to glutathione sepharose resin and FLAG-tagged prey proteins were co-incubated to investigate predicted interactions. GST-HP0958 pull-down assay was performed against FliH (2-258) (full length) and FliH (89-258), the previously identified domain involved in this predicted interaction (Rain *et al.*, 2001). All assays failed to show any detectable interaction above background non-specific retention of the prey protein (lanes 5 and 6 *vs* lanes 2 and 3, Figure 22).

The HP0958-RpoN interaction also predicted in the Y2H study has since been confirmed and HP0958 was identified as a chaperone of σ^{54} (Pereira and Hoover, 2005; Rain *et al.*, 2001; Ryan *et al.*, 2005a). RpoN (74-284) was previously identified as the domain involved in the HP0958-RpoN interaction. However, we were unable to purify this protein in soluble form, and so GST pull-down analyses were performed with the soluble truncated protein RpoN (74-210) (Rain *et al.*, 2001). No interaction was detected between HP0958 and RpoN (74-210) under the conditions tested. HP0958 has been shown to interact with the *flaA* mRNA transcript (Douillard *et al.*, 2008). The soluble cytoplasmic fraction of motile culture of a *H. pylori* P79 lysate was also assayed for ability of any of its constituent proteins to bind GST-HP0958 (data not shown); however, no clear targets for interaction with HP0958 were identified.

To further investigate the results of the Rain *et al.* study regarding HP0958 protein-protein interactions, proteins were introduced by cloning respective genes into *S. cerevisiae* strains AH109 and Y187 and yeast two-hybrid assays were performed. Activation of expression of the reporter gene *lacZ* due to the interaction of bait and prey proteins produced β-galactosidase which was measured by the

adapted plate X-gal assay. β-galactosidase activity of Y187 derivative strains harbouring genes encoding a given PPI set was determined relative to β-galactosidase activity of an *E. coli* strain in which *lacZ* is constitutively expressed. PXG assay failed to detect any interaction between HP0958 and FliH (89-258) in both prey-bait combinations (Figure 23). The FliH (89-258)-FliI (2-91) interaction served as a positive control since this interaction set was also predicted in the Rain *et al.* study and was subsequently confirmed by biochemical means (Lane *et al.*, 2006; Rain *et al.*, 2001). Additionally, the HP0958-RpoN (74-284) interaction was verified as a weak interaction by Y2H assay, inducing ~7 fold less *lacZ* expression than Y187 derivative strains possessing FliH (89-258) and FliI (2-91) (Figure 23).

4.2 Complementation of *hp0958*-null derivative of P79 with HP0958 mutant alleles.

A panel of 18 target residues for site-directed mutagenesis of HP0958 were selected based on their potential contribution to HP0958 function during flagellum biogenesis (Appendix 18). A recent study suggested that conserved histidine residues may have a propensity to form stacking interactions with aromatic amino acids and so may be involved in PPIs (Liao *et al.*, 2013). Surface-exposed hydrophobic residues (leucine, isoleucine and phenylalanine) with a propensity to form interactions with other hydrophobic amino acids were also selected for mutation (Jones and Thornton, 1996). Conserved positively charged residues (arginine, lysine and histidine) were selected for mutation as they may be involved in protein-nucleic acid interactions with the negatively charged phosphate groups of nucleic acids (Ellis *et al.*, 2007; Iwakiri *et al.*, 2011).

Genes encoding mutant alleles of HP0958 were introduced into the chromosome of a *hp0958*-null derivative of *H. pylori* P79 at an intergenic site by natural transformation with the suicide vector pIR203K04 (Langford *et al.*, 2006) (Table 20). Mutant allele expression was under the control of the *ahpC* promoter Php1563 (Douillard *et al.*, 2008). To establish whether these mutant forms of HP0958 were capable of restoring motility to the non-motile derivative P79-0958KO, transformants were screened by motility assay on soft agar. Additional phenotypic analyses included microscopy, TEM, immunoblotting and qRT-PCR of selected flagellar genes. Flagellum biogenesis is a highly energy-consuming process for the cell and hence, without the highly selective pressure of its native environment,

strains grown in a lab setting had a tendency to revert to a non-motile state, as seen previously (Eaton *et al.*, 1992; Josenhans *et al.*, 2000). The re-introduction of the wild-type *hp0958* gene into a *hp0958*-null non-motile derivative strain of P79 was capable of restoring motility beyond wild-type P79 levels, as previously shown in a *H. pylori* 17874 mutant derivative (Douillard *et al.*, 2008). Therefore, it must be considered that the restoration of motility as determined by *ex-vivo* analyses is limited by the tendency for this highly genetically plastic pathogen to return to an aflagellate state, likely through reversible phase-variation of flagellar genes *e.g. fliP* (Josenhans *et al.*, 2000).

Electron microscopy revealed that derivative strains transformed with some HP0958 mutant alleles produced normal flagella, while others were unable to do so. Wild-type P79 cells generally possessed 1 - 2 polar flagella encased in a characteristic sheath (Figure 24; for further details see Appendix 21). P79-0958KO cells in which the *hp0958* gene has been insertionally inactivated were aflagellate and non-motile (Ryan *et al.*, 2005a) (Figure 24). Introduction of the wild-type *hp0958* gene under the control of the *ahpC* promoter restored flagellar production, as previously seen in a *hp0958*-null derivative of *H. pylori* strain CCUG 17874 (Douillard *et al.*, 2008) (Figure 24). The crystal structure of HP0958 revealed two structural domains: an N-terminal coiled-coil and a C-terminal Zn-finger. Mutation of residues in the Zn-finger generally produced flagellate mutant cells (Table 21). However, only 2 out of 11 mutations in the coiled-coil/hinge region produced derivative strains which were flagellate (Table 21).

Interestingly, several types of extrusions which did not resemble a typical *H. pylori* flagellum were observed by electron microscopy (Figure 24). Six mutations in the coiled-coil and 4 in the Zn-finger resulted in P79-0958KO complemented cells which produced a multi-bulb phenotype, so-called due to the protrusion of appendages which resembled multiple flagellar sheath distal bulbs without the presence of a flagellar filament. In some mutants (I99A and I204A), the strains produced singular or multiple enlarged bulbs (Figure 24). Surprisingly, 11 mutants (L47A, L58A, I99A, F161A, K195A, F203V, I204A, R205A, K209E, T222A and Y231F) spanning the two structural domains of HP0958 produced cells with appendages at non-polar sites, including 4 in the Zn-finger (F203V, K209E, T222A and Y231F) with fully-formed flagella at both poles/non-polar sites (Figure 24).

Thus, complementation of P79-0958KO strain with HP0958 mutant alleles produces strains which indicate HP0958 is either inactive or fully/partially active.

4.3 *H. pylori* P79-0958KO derivative strains complemented with coiled-coil mutant HP0958 alleles are non-motile while complementation with Zn-finger mutant alleles restores motility.

Motility of *H. pylori* strains was assessed by microscopy and soft agar assay. Strains which were flagellate according to TEM imaging also produced motility zones on 0.3% BHI agar (Figure 25). P79-0958KO cells were non-motile and only grew in the centre of the agar at the site of inoculation. Alanine substitution of leucine at position 47 (L47A) in the coiled-coil of the HP0958 protein was the only complemented mutant which appeared non-motile by motility assay (Figure 25). Eight mutant complemented strains (T3A, H4A, I99A, F161A, R181E, K187A, K195A and I204A), 6 of whose sequence changes are in the coiled-coil, did not produce a zone of motility within the agar. However, complemented strains harbouring these mutations produced a ring phenotype on the surface of the soft agar, possibly due to impaired motility (Figure 25). The L58A, F203V, T222A and Y231F complemented derivative strains produced halos similar to that of the P79-0958KO complemented strain with wild-type HP0958 allele (Figure 25). R205A, R205V and K209E complemented derivative strains produced halos of diameter similar to that of wild-type P79 and smaller than that of the complemented strain with wild-type HP0958 allele (Figure 25).

Previous studies of *hp0958* knock-out derivatives found that in the absence of HP0958, the Class II sigma factor, RpoN, is unstable which causes reduced expression of RpoN-dependent genes (Pereira and Hoover, 2005; Ryan *et al.*, 2005a). Additionally, HP0958 is thought to interact directly with the *flaA* mRNA transcript; in the absence of HP0958, FlaA expression levels are also impaired (Douillard *et al.*, 2008; Ryan *et al.*, 2005a). Therefore, expression levels of flagellin (FlaA and FlaB) and hook (FlgE) proteins were monitored as an indicator of HP0958 activity. Cell lysates from *H. pylori* strains grown in liquid to exponential phase were immunoblotted with anti-flagellin and anti-hook antibodies. In agreement with previous findings, the *hp0958* knock-out derivative of P79 had reduced flagellin protein levels and no FlgE was detected (Figure 26). The 9 mutant derivative strains

which were motile by motility assay (H13A, L58A, R184E, F203V, R205A, R205V, K209E, T222A and Y231F) all had a similar flagellin/hook profile to that of the wild-type P79 strain. Six mutant derivative strains (H4A, I99A, F161A, R181E, K187A and K195A) had expression profiles matching that of the P79-0958KO derivative of P79, indicating that these mutations impair HP0958 activity (Figure 26). Derivative strains harbouring the T3A and L47A mutations which affected the HP0958-RpoN interaction appear to produce more FlaB than the *hp0958*-null derivative of P79. The I204A complemented derivative has flagellin levels similar to P79-0958KO but does produce FlgE at a detectable level, albeit less than that of the wild-type complemented derivative (Figure 26).

The mRNA expression levels of 5 flagellar genes were monitored relative to the housekeeping *era* gene. Figure 27 shows a heatplot of the qRT-PCR data, ranking the HP0958 mutants by expression of the RpoN regulon genes, *flaB* and *flgE*. In agreement with Douillard *et al.*, *fliA* and *rpoN* experience the least fluctuation in expression when *hp0958* is deleted or mutated compared to the P79 wild-type. Overall, the expression profiles for motile and non-motile strains form two separate clusters, with the following exceptions. In the case of non-motile derivative strains harbouring the H4A and F161A mutations in HP0958, these restored *flaA* expression levels close to that of the wild-type complement, with higher *flgE* expression and lower *flaB* expression. Motile derivative strains K209E and R205A are found within the non-motile cluster (Figure 27), however, these cells produced reduced zones of motility by soft agar assay (Figure 25). Seven of the 10 mutant derivative strains which produced multi-bulb extrusions cluster together between the motile and aflagellate cells. Therefore, these strains have an intermediate expression profile of flagellar genes indicating some level of function of the HP0958 mutant alleles.

4.4 Structure/function analysis of the HP0958-RpoN (74-284) interaction reveals involvement of the HP0958 coiled-coil domain.

Though the HP0958-RpoN interaction has been confirmed by biochemical means, little is known about the mechanism of this interaction. The Rain *et al.* study predicted that residues 30-218 of HP0958 were required for the interaction with RpoN (74-284) (Rain *et al.*, 2001). In order to investigate this, 14 site-directed mutants of HP0958 were screened by Y2H and PXG analysis to determine their effect upon the HP0958-RpoN (74-284) interaction. β-galactosidase induction due to

interactions involving these mutants was expressed as a proportion of the wild-type HP0958-RpoN (74-284) β-galactosidase activity (Figure 28).

Interactions involving R181E and R184E mutants in the coiled-coil, and F203V, K209E and T222A in the Zn-finger behaved similarly to wild-type HP0958 (Figure 28). Alanine substitution of residues T3, H13, L47, F161 and K187 in the coiled-coil and hinge region resulted in β-galactosidase levels which differed significantly from the wild-type. The I99A mutant failed to support any measurable level of reporter gene expression, indicating that alanine substitution of I99 in the coiled-coil abolished the interaction between HP0958 and RpoN (74-284). H13A, L47A, K195A and 1204A mutations all decreased the strength of interaction between HP0958 and RpoN (74-284), while T3A, F161A and K187A all increased the strength of interaction.

4.5 The full length *flaA* transcript is required for full-strength HP0958 interaction.

Secondary structural analysis of the *flaA* mRNA transcript revealed a predicted structure which forms three distinct regions (Figure 29). Region 1 contains nucleotides 1-317 (including the ribosomal binding site) and nucleotides 1202-1633 (Figure 29). This truncated RNA transcript was predicted to have almost identical secondary structure to that of the full length sub-region, except for the presence of an additional loop replacing the deleted middle section (regions 2 and 3). Region 2 contains nucleotides 326-817 of the *flaA* mRNA transcript and region 3 contains nucleotides 869-1198, both of which were truncated at these points to retain the same predicted secondary structure as those regions within the full length *flaA* mRNA.

EMSA analysis was performed to identify the region(s) of the *flaA* transcript which interact(s) with HP0958 during flagellum biogenesis (Figure 30). Full length *flaA* mRNA in complex with HP0958 migrated more slowly than unbound transcript and produced a diffuse gel-shift band, as previously shown (Caly *et al.*, 2010; Douillard *et al.*, 2008). None of the truncated *flaA* transcripts (regions 1-3) produced a gel-shift similar to the full length mRNA-HP0958 complex and hence were impaired in their ability to interact with HP0958. A very faint band was visible at the same position as the full length gel-shift for the *flaA* region 1 (lane 6). A very weak gel-shift with a higher mobility was produced by the HP0958-*flaA* region 2 complex.

Region 3 was unable to produce any detectable gel-shift through interactions with HP0958.

4.6 Structure/function analysis of the HP0958-*flaA* mRNA interaction.

Caly *et al.* previously tested 14 site-directed mutants of HP0958 to determine the impact on the HP0958-*flaA* mRNA interaction (Caly *et al.*, 2010). While many of these impacted upon the interaction, none of the mutations abolished the interaction. In order to augment the structure-function analysis of HP0958, 14 additional site-directed mutants of the flagellum biogenesis protein were generated in this study. The targets for mutagenesis spanned the two structural domains of HP0958 with 6 in the coiled-coil, 1 in the hinge region and 7 in the Zn-finger domain. These mutants were selected based on their conservation in HP0958 homologues across ε-proteobacteria and their positively charged or aromatic characteristics (Jones *et al.*, 2001). The previous study found that R181A, R184A and K209A mutations had an observable effect on the complex gel-shift relative to wild-type HP0958 (Caly *et al.*, 2010). Glutamic acid substitution was performed for these 3 residues in order to determine if this mutation could exacerbate the effect caused by alanine mutation.

Alanine substitution of T3A, H4A, H13A and Y231F had little or no effect on the migration of the HP0958-*flaA* transcript complex when compared to the gel-shift produced by wild-type HP0958 (Figure 31). L58A and K187A mutations produced a HP0958-*flaA* mRNA complex gel-shift similar to that of the wild-type HP0958. The T222A mutant apparently strengthened the protein-RNA interaction. K195A, F203V and R205V mutants formed complexes with *flaA* mRNA that migrated slightly faster than the wild-type complex, resulting in a slightly lower gel-shift position. R181E, R184E and K209E mutations did not alter the migration of *flaA* mRNA and hence abolished the HP0958-*flaA* mRNA interaction. L58A and T222A in complex with the *flaA* mRNA transcript produced gel-shifts which were slightly less diffuse than that of wild-type HP0958-*flaA* mRNA complex.

5 Discussion

HP0958 is an essential component of flagellum biogenesis which is involved in multiple interactions during the assembly process (Caly *et al.*, 2010; Douillard *et al.*, 2008; Ryan *et al.*, 2005a). The protein-protein interaction network of *H. pylori* provided a platform for investigation into the role of HP0958, identifying interactions with two key flagellar components, FliH and RpoN (Rain *et al.*, 2001). In this study, we present a detailed structure-function analysis of HP0958 through investigation of previously identified interactions with RpoN, FliH and *flaA* mRNA.

Y2H and GST pull-down analyses failed to confirm the previously identified HP0958-FliH interaction. This interaction as indicated by Rain (Rain *et al.*, 2001) appears to be a false positive within a large scale analysis of protein-protein interactions. Although measures have been taken to reduce the level of false positives wrongly identified as interaction pairs in yeast two-hybrid assays, our study affirms the necessity to confirm Y2H data by biochemical methods.

Complementation of a non-motile *hp0958*-null derivative of P79 with site-directed mutant alleles of *hp0958* resulted in derivative strains with wild-type, partial or no HP0958 activity. Mutations which abolished or significantly decreased HP0958-RpoN interactions resulted in aflagellate cells while all mutant derivative strains which produced flagella resembling the wild-type were capable of motility. Thus, not unexpectedly, HP0958 contributes solely to flagellum assembly, not flagellum function. In general, mutations in the Zn-finger resulted in cells which produced flagella, while those in the coiled-coil lacked flagella resembling the wild-type. This indicates that the RpoN interaction site in HP0958 is predominantly localised to the coiled-coil, supporting the Y2H analysis performed in this study.

Complementation of the non-motile P79-0958KO strain with 2 mutant proteins which abolished HP0958-*flaA* mRNA interaction (R184E and K209E; assessed by EMSA) but did not significantly affect the RpoN interaction produced cells with diminished motility. Complementation of the *hp0958*-null derivative of P79 with R205A and R205V mutant proteins also resulted in cells with reduced motility when compared to cells complemented with wild-type HP0958; these mutations formed HP0958-*flaA* mRNA complexes which migrated differently to that of the wild-type in complex (Figure 31). Thus, the data indicates that HP0958 functions can be separated based the on activities of two distinct structural domains (coiled-coil and

Zn-finger) through the combined structure-function analysis of HP0958-*flaA* mRNA and HP0958-RpoN interactions, supporting the hypothesis proposed by Douillard *et al.* (Caly *et al.*, 2010; Douillard *et al.*, 2008) (Figure 32).

Recently, Iwakiri *et al.* performed structural analysis of 91 protein-RNA interactions for which 3D information was available and found that aspartic acid is often present at protein-RNA interfaces where it is proposed to be involved in RNA loop recognition (Iwakiri *et al.*, 2011). The presence of two conserved aspartic acid residues (D208 and D219) between the cystine knuckles of the Zn-finger in HP0958 supports the involvement of this domain in *flaA* mRNA interactions (Appendix 22). Detailed structure-function analysis of the HP0958-*flaA* mRNA interaction, together with the previously published work of Caly *et al.* has identified key residues in the Zn-finger of HP0958 involved in RNA contact (Caly *et al.*, 2010). Coiled-coil mutations at the N-terminal (T3A, H4A, H13A) did not affect the interaction. Positively charged amino acids (R181, R184, K209 and K195) which are associated with protein-RNA interactions were found to be involved in *flaA* mRNA binding, likely through electrostatic interactions with negatively charged phosphate groups of RNA (Ellis *et al.*, 2007; Iwakiri *et al.*, 2011). Residue I204 of HP0958 may have a dual function as mutation significantly affected both RpoN and *flaA* mRNA interactions.

Y2H analysis confirmed that HP0958 interacts with the domain spanning residues 74-284 of RpoN. Analysis of the HP0958-RpoN interaction revealed many contact points along the structure of HP0958: 5 in the coiled-coil, 1 in the hinge region and 2 in the Zn-finger. This indicates that the RpoN protein is likely to be in an extended conformation while interacting with HP0958. I99A and K187A mutations abolish the HP0958-RpoN (74-284) interaction in a Y2H interaction model. H13A, L47A, K195A and I204A mutations all result in a significant decrease in the strength of the RpoN interaction. Interestingly, T3A, F161A and K187A all significantly increased the interaction strength but none of these mutant alleles were capable of restoring motility when transformed into P79-0958KO. We hypothesise that by enhancing the binding of HP0958 to RpoN beyond that of the wild-type interaction, this can inhibit activity of this sigma factor by reducing its interaction with the core RNA polymerase.

Mutations disturbing the RpoN regulon had a more dramatic effect on flagellum biogenesis as seen by TEM analysis when compared to mutations which disturbed

the HP0958-*flaA* mRNA interaction. It may be that the *flaA* mRNA interacts at more residues which are in close proximity which can compensate for single site-mutations, whereas residues involved in RpoN interactions are dispersed across a much larger surface area which more easily destabilise the interaction. It is also possible that some mutants which impede *flaA* incorporation into the filament can compensate for this by producing filaments with higher FlaB composition than the wild-type and hence produce flagella which are still capable of motility.

H. pylori are lophotrichous and generally possess 2 - 6 polar sheathed flagella with a characteristic bulbed tip (Geis *et al.*, 1993; Goodwin *et al.*, 1985). FlhF and FlhG have been implicated in localisation of flagellum biogenesis to the bacterial cell pole and in control of flagellum number in *H. pylori*, *Campylobacter jenuni* and *Vibrio cholerae* (Balaban and Hendrixson, 2011; Balaban *et al.*, 2009; Lertsethtakarn *et al.*, 2011). TEM analysis revealed that many derivative strains of P79-0958KO complemented with HP0958 mutant proteins presented flagellar-type extrusions from the cell surface at non-polar positions which did not resemble typical flagella. Furthermore, F203V and Y231F mutations resulted in complemented cells which produced wild-type flagella at both poles; T222A complemented cells produced wild-type flagella which protruded from the side of the bacterial cells. Therefore, we propose a novel function of HP0958 during flagellum biogenesis: localisation of flagellum biogenesis to a single cell pole. The occurrence of non-polar flagellar extrusions from cells complemented with mutated alleles spanning all secondary structural elements of HP0958 suggests that this role may involves both the coiled-coil and Zn-finger domains. These extrusions are likely to either be empty flagellar sheaths or sheaths encasing abnormal flagellar sub-structures, similar to the empty sheaths produced by *fliD* mutant derivative cells (Kim *et al.*, 1999). Transcriptional analysis of strains producing such extrusions indicates partial restoration of HP0958 function/flagellar gene expression, indicating some flagellum biogenesis activity within the cells.

In *Caulobacter crescentus*, the flagellin genes *fljK* and *fljL* are transcribed but not translated until the hook/basal-body complex has been completed. FlbT binds and destabilises the transcript to prevent premature translation and secretion, much as we hypothesise HP0958 may act on the *flaA* mRNA transcript in *H. pylori* (Anderson and Gober, 2000). FlbT, the post-transcriptional regulator of flagellin synthesis in *Caulobacter crescentus*, interacts with the 5' untranslated region of flagellin mRNA.

Deletion of the 5' untranslated regon of the *H. pylori flaA* mRNA transcript did not have any effect upon HP0958 binding (Douillard *et al.*, 2008). Therefore, a secondary structure-based approach was adopted in order to investigate the nature of this interaction from the RNA perspective. Region 1 and Region 2 truncated mRNA transcripts had significantly reduced capacity to bind HP0958. However, the observation of very faint gel-shifts does indicate the involvement of these branches of the full length predicted secondary structure. While Region 3 of the *flaA* mRNA transcript alone may not be capable of interacting with HP0958, these results indicate that the *flaA* transcript as a whole is required for efficient protein-RNA interaction. This may be mediated by sequence-specific interactions or recognition of the secondary structure of the complete mRNA transcript.

In conclusion, this study presents an in depth structure-function analysis of the role of HP0958 during flagellum biogenesis in *H. pylori*. Taken together, these data support the previously proposed mechanism of HP0958 function with one exception. HP0958 was proposed to target the *flaA* mRNA transcript to the export apparatus through its interaction with FliH (Douillard *et al.*, 2008). With the elimination of the FliH interaction from this model, there is a need for further refinement of our understanding of the role of HP0958 in flagellum assembly. One possibility is that the HP0958 has additional interaction partners which have not yet been identified. It cannot be excluded that HP0958 and the identified interaction partners discussed in this study may require additional flagellar components to form a fully functional complex. Analysis of potential interactions between HP0958 and components of the basal body and export apparatus such as FlhA may provide the key for what targets the transcript in complex with HP0958 for efficient export. Purification of a soluble form of the RpoN (74-284) would facilitate analyses to determine the potential role of HP0958 in the switch between expression of Class II and Class III flagellar genes. The presence of flagellar extrusions from non-polar sites in derivative strains of P79-0958KO complemented with mutant alleles suggests a novel role of HP0958 in localisation of flagellum biogenesis to a single cell pole. Further investigation into this function is warranted to further elucidate flagellum assembly of *H. pylori.*

6 Acknowledgements and Disclosures

This project was supported by a Research Frontiers Programme award (09_RFP_GEN2443) from Science Foundation Ireland to PWOT, and by an Embark scholarship from IRCSET to CDC. Yeast strains and Clontech Matchmaker™ Gold Yeast Two Hybrid plasmids were a kind gift from Prof. Paul Young. The plasmid pIR203K04 was generously provided by Dr. David McGee. TEM imaging was performed at the Biological Imaging Facility directed by Dr. Dimitri Scholtz, Conway Institute of Biomolecular and Biomedical Research, University College Dublin.

7 Tables and Figures

Table 18 List of strains used in this study

Strain	Relevant characteristics	Source
H. pylori		
CCUG 17874	Wild type strain	CCUG, Sweeden
P79	P1 Str^r	(Heuermann and Haas, 1998)
P79-0958KO[1]	P79 Δ*hp0958*::Cm^r	(Douillard *et al.*, 2008)
P79-0958/pIR203K04	P79 Δ*hp0958*::Cm^r with pIR203K04 (Kan^r)	(Douillard *et al.*, 2008)
E. coli		
XL1-Blue Supercompetent cells	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F^- proAB lacIqZΔM15 Tn10 (TetR)]	Stratagene, USA
One shot Top 10	F^- mcrA _ (mrr-hsdRMS-mcrBC) _80lacZ_M15 _lacX74 nupG recA1 araD139 (ara-leu)7697 galE15 galK16 rpsL (StrR) endA1	Invitrogen, CA
DH5α	F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1	Invitrogen, CA
Rosetta 2(DE3) pLysS	F^- ompT hsdSB (rB–mB–) gal dcm (DE3) pLysSRARE2 (CamR)	Novagen, Darmstadt, Germany

S. cerevisiae

AH109	MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2::$GAL1_{UAS}$-$GAL1_{TATA}$-HIS3, $GAL2_{UAS}$-$GAL2_{TATA}$-ADE2, URA3::$MEL1_{UAS}$-$MEL1_{TATA}$-lacZ	Dr. Paul Young, UCC
Y187	MATα, ura3-52, his3-200, ade2-101, trp1-901, leu2-3, 112, gal4Δ, met-, gal80Δ, URA3::GAL1UAS-GAL1TATA-lacz	Dr. Paul Young, UCC

[1]KO, knockout.

Table 19 List of plasmids used for yeast two-hybrid and protein expression

Plasmids	Relevant characteristics	Source
pGEX-6p-3	N-terminally GST-tagged expression vector	GE Healthcare, UK
pDC006	pGEX-6p-3 *hp0958*	(Caly *et al.*, 2010)
pFliH01	pGEX-6p-3 *fliH 2-258*	This study
pFliH02	pGEX-6p-3 *fliH 89-258*	This study
pRpoN01	pGEX-6p-3 *rpoN 74-284*	This study
pRpoN02	pGEX-6p-3 *rpoN 74-210*	This study
pMut2	pGEX-6p-3 HP0958 mutant T3A	This study
pMut3	pGEX-6p-3 HP0958 mutant H4A	This study
pMut4	pGEX-6p-3 HP0958 mutant H13A	This study
pMut6	pGEX-6p-3 HP0958 mutant L58A	This study
pMut9	pGEX-6p-3 HP0958 mutant R181E	This study
pMut10	pGEX-6p-3 HP0958 mutant R184E	This study
pMut11	pGEX-6p-3 HP0958 mutant K187A	This study
pMut12	pGEX-6p-3 HP0958 mutant K195A	This study
pMut13	pGEX-6p-3 HP0958 mutant F203V	This study
pMut15	pGEX-6p-3 HP0958 mutant R205A	This study
pMut16	pGEX-6p-3 HP0958 mutant R205V	This study
pMut17	pGEX-6p-3 HP0958 mutant K209E	This study
pMut18	pGEX-6p-3 HP0958 mutant T222A	This study
pMut19	pGEX-6p-3 HP0958 mutant Y231F	This study
pMAD	Em^r cassette; β-galactosidase gene under constitutive promoter	(Arnaud *et al.*, 2004)
pGBKT7	Kan^r for selection in *E. coli*; *Trp1* nutritional marker for selection in *S. cerevisiae*	Dr. Paul Young
pGADT7	Amp^r for selection in *E. coli*; *Leu2* nutritional marker for selection in *S. cerevisiae*	Dr. Paul Young
pCC01	pGBKT7 *hp0958*	This study
pCC02	pGADT7 *fliH 89-258*	This study
pCC03	pGBKT7 *fliH 89-258*	This study
pCC04	pGADT7 *hp0958*	This study
pCC05	pGADT7 *rpoN 74-284*	This study
pCC06	pGADT7 *rpoN 74-210*	This study
pCC07	pGADT7 *fliI 2-91*	This study
pMut2k	pGBKT7 HP0958 mutant T3A	This study
pMut3k	pGBKT7 HP0958 mutant H4A	This study
pMut4k	pGBKT7 HP0958 mutant H13A	This study
pMut5k	pGBKT7 HP0958 mutant L47A	This study
pMut6k	pGBKT7 HP0958 mutant L58A	This study
pMut7k	pGBKT7 HP0958 mutant I99A	This study

pMut8k	pGBKT7 HP0958 mutant F161A	This study
pMut9k	pGBKT7 HP0958 mutant R181E	This study
pMut10k	pGBKT7 HP0958 mutant R184E	This study
pMut11k	pGBKT7 HP0958 mutant K187A	This study
pMut12k	pGBKT7 HP0958 mutant K195A	This study
pMut13k	pGBKT7 HP0958 mutant F203V	This study
pMut14k	pGBKT7 HP0958 mutant I204A	This study
pMut17k	pGBKT7 HP0958 mutant K209E	This study
pMut18k	pGBKT7 HP0958 mutant T222A	This study
pMut19k	pGBKT7 HP0958 mutant Y231F	This study

Table 20 List of plasmids transformed into *H. pylori* strain P79-0958KO

Plasmids	Relevant characteristics	Source
pIR203K04	Kan^r suicide vector	(Langford *et al.*, 2006)
pIR0958	pIR203K04 with the *hp0958* gene under the control of the *hp1563* promoter	(Douillard *et al.*, 2008)
pIRmut2	pIR0958 mutant T3A	This study
pIRmut3	pIR0958 mutant H4A	This study
pIRmut4	pIR0958 mutant H13A	This study
pIRmut5	pIR0958 mutant L47A	This study
pIRmut6	pIR0958 mutant L58A	This study
pIRmut7	pIR0958 mutant I99A	This study
pIRmut8	pIR0958 mutant F161A	This study
pIRmut9	pIR0958 mutant R181E	This study
pIRmut10	pIR0958 mutant R184E	This study
pIRmut11	pIR0958 mutant K187A	This study
pIRmut12	pIR0958 mutant K195A	This study
pIRmut13	pIR0958 mutant F203V	This study
pIRmut14	pIR0958 mutant I204A	This study
pIRmut15	pIR0958 mutant R205A	This study
pIRmut16	pIR0958 mutant R205V	This study
pIRmut17	pIR0958 mutant K209E	This study
pIRmut18	pIR0958 mutant T222A	This study
pIRmut19	pIR0958 mutant Y231F	This study

Table 21 Overview of structure-function analysis of HP0958 by analysis of site-directed mutant proteins

Compiled results of biochemical assays and complementation data from the current study, with previously published structure-function analysis by *Caly et al.* (Caly *et al.*, 2010) where: (*) positive; (-) negative; empty cells denote no data available; (M) motile; (N) non-motile; (S) swarming. "Location" refers to the secondary structure within HP0958 at that residue selected for mutation where: (α) α-helix; (β) β-sheet; (Kn) knuckle co-ordinating zinc atom; (hinge) linker region between coiled-coil and Zn-finger domains.

Mutation	**Location**	**RpoN interaction (Y2H)**	***flaA* interaction (EMSA)**	**Motility assay***	**TEM**			
					Flagella	Multi-bulb	Large bulbs	Non-polar
P79 WT				M	*	-	-	-
KO				N	-	-	-	-
Complement				M	*	-	-	-
T3A	α1a	Increased	Little or no effect	S	-	*	-	-
H4A	α1a	Same as WT	Little or no effect	S	-	-	-	-
H13A	α1a	Decreased	Little or no effect	M	-	*	-	-
L47A^{+}	α1b	Decreased	Same as WT	N	-	*	-	*
L58A	α1b		Same as WT	M	*	*	-	*
I99A^{+}	α2a	Abolished	Same as WT	S	-	-	*	*
F161A^{+}	α2b	Increased	Same as WT	S	-	*	-	*
F178A^{+}	α3		Observable effect					
Y179A^{+}	α3							
R181A^{+}	α3		Observable effect					
R181E	α3	Same as WT	Abolished	S	-	*	-	-
R184A^{+}	α3		Little or no effect					
R184E	α3	Same as WT	Abolished	M	*	-	-	-
W185A^{+}	α3		Little or no effect					
K187A	hinge	Increased	Same as WT	S	-	-	-	-
T189A^{+}	hinge		Observable effect					

K195A	Zn ribbon	Decreased	Same as WT	S	-	*	-	*
K196A[+]	Zn ribbon		Little or no effect					
Q197A[+]	Zn ribbon		Observable effect					
C199A[+]	Zn ribbon, Kn1		Observable effect					
F203V	Zn ribbon, Kn1	Same as WT	Observable effect	M	*	-	-	*
I204A[+]		Decreased	Observable effect	S	-	*	*	*
R205A	Zn ribbon		Observable effect	M	*	*	-	*
R205V°	Zn ribbon		Observable effect	M	-	-	-	-
K209A[+]	Zn ribbon, α4		Little or no effect					
K209E	Zn ribbon, α4	Same as WT	Abolished	M	*	*	-	*
Y211A[+]	Zn ribbon, α4		Little or no effect					
T222A	Zn ribbon, Kn2	Same as WT	Increased	M	*	-	-	*
R228A[+]	Zn ribbon, Kn2		Observable effect					
Y231F	Zn ribbon, β2		Little or no effect	M	*	-	-	*

[*]Wild-type and P79 derivative strains.

[+]HP0958 site-directed mutants generated for EMSA screen by *Caly et al.* (Caly *et al.*, 2010).

°TEM of this strain will be repeated as images from this culture include artefacts (Appendix 21).

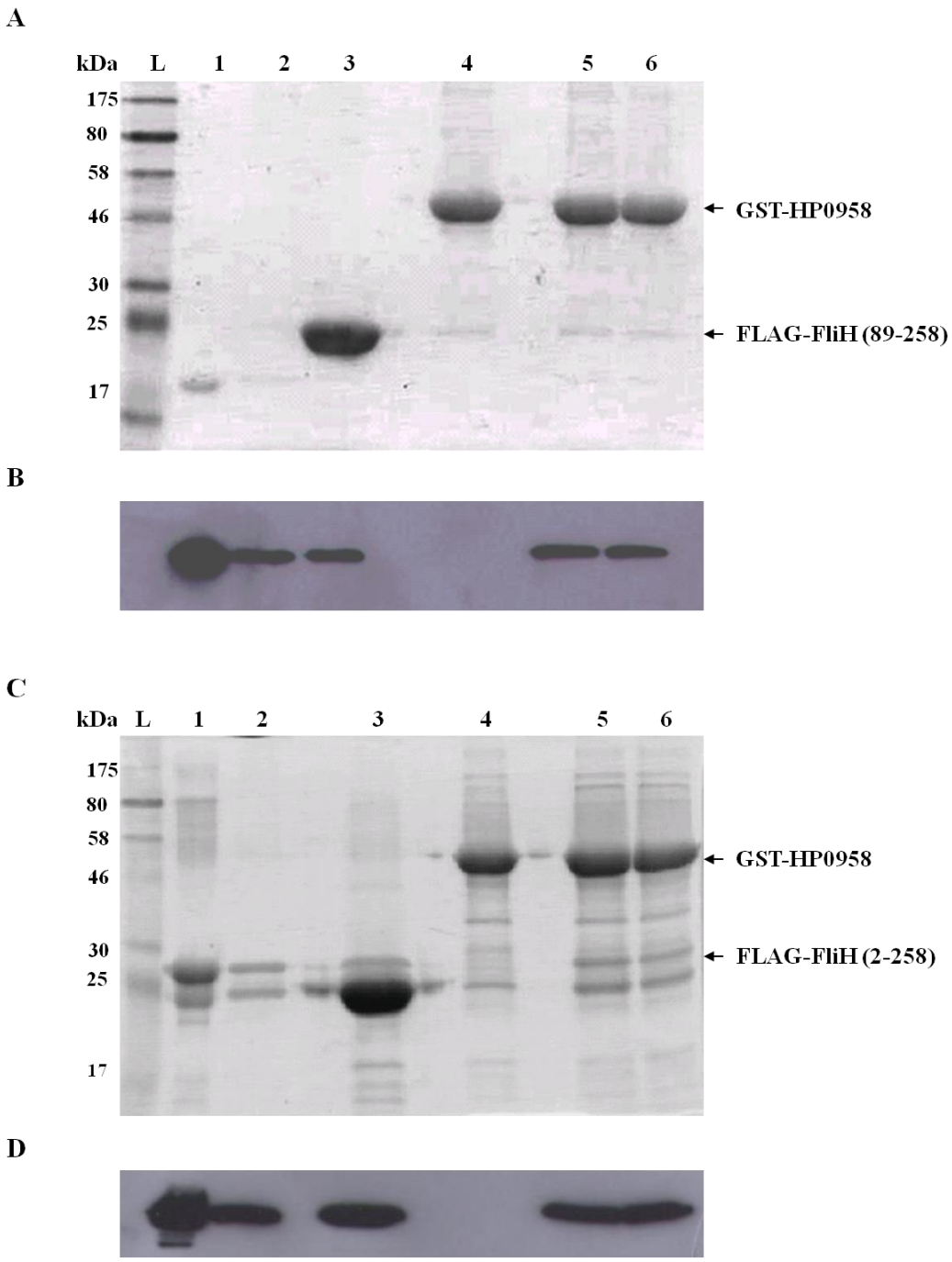


Figure 22 GST pull-down assay investigating the previously proposed HP0958-FliH interaction.

GST pull-down assay of HP0958 and FLAG fusion FliH proteins. (**A**) SDS-PAGE of HP0958-FliH (89-258) GST pull-down assay; (**B**) corresponding immunoblot with anti-FLAG antibody; (**C**) SDS-PAGE of HP0958-FliH (89-258) GST pull-down assay; (**D**) corresponding immunoblot with anti-FLAG antibody. Loading was identical for (**A - D**) where: L = Prestained Broad Range protein ladder; 1 = FLAG fusion FliH; 2 = glutathione sepharose B incubated with FLAG-FliH; 3 = glutathione sepharose B resin bound GST incubated with FLAG-FliH; 4 = glutathione sepharose resin bound GST-HP0958; 5 and 6 = glutathione sepharose B resin bound GST-HP0958 incubated with FLAG-FliH.

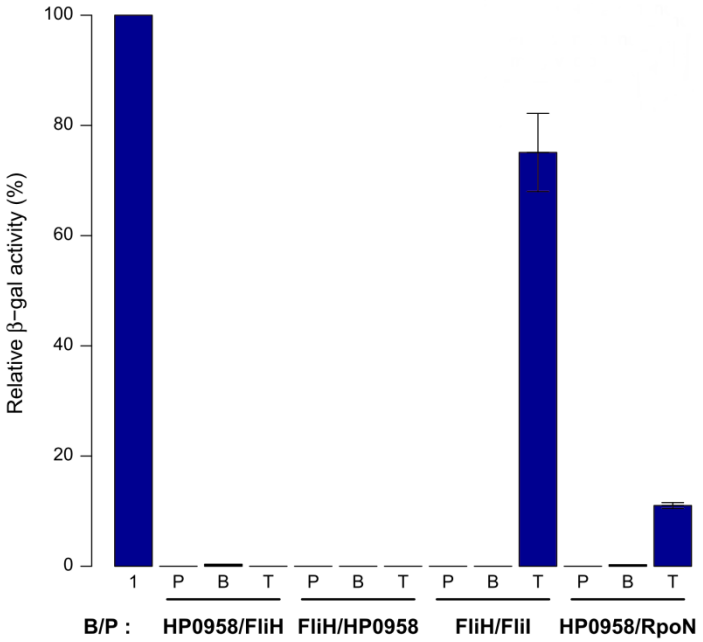


Figure 23 Investigation of previously proposed HP0958 PPIs by yeast two-hybrid assay.

β-galactosidase activity relative to positive control strain *E. coli* Top10 harbouring the pMAD plasmid for constitutive expression of the enzyme. B = strains possessing bait protein expressed as fusion protein with the GAL4 DNA-binding domain; P = prey protein expressed as fusion protein with the GAL4 transcription activation domain; T = strains possessing both vectors harbouring bait and prey genes for a given interaction pair.

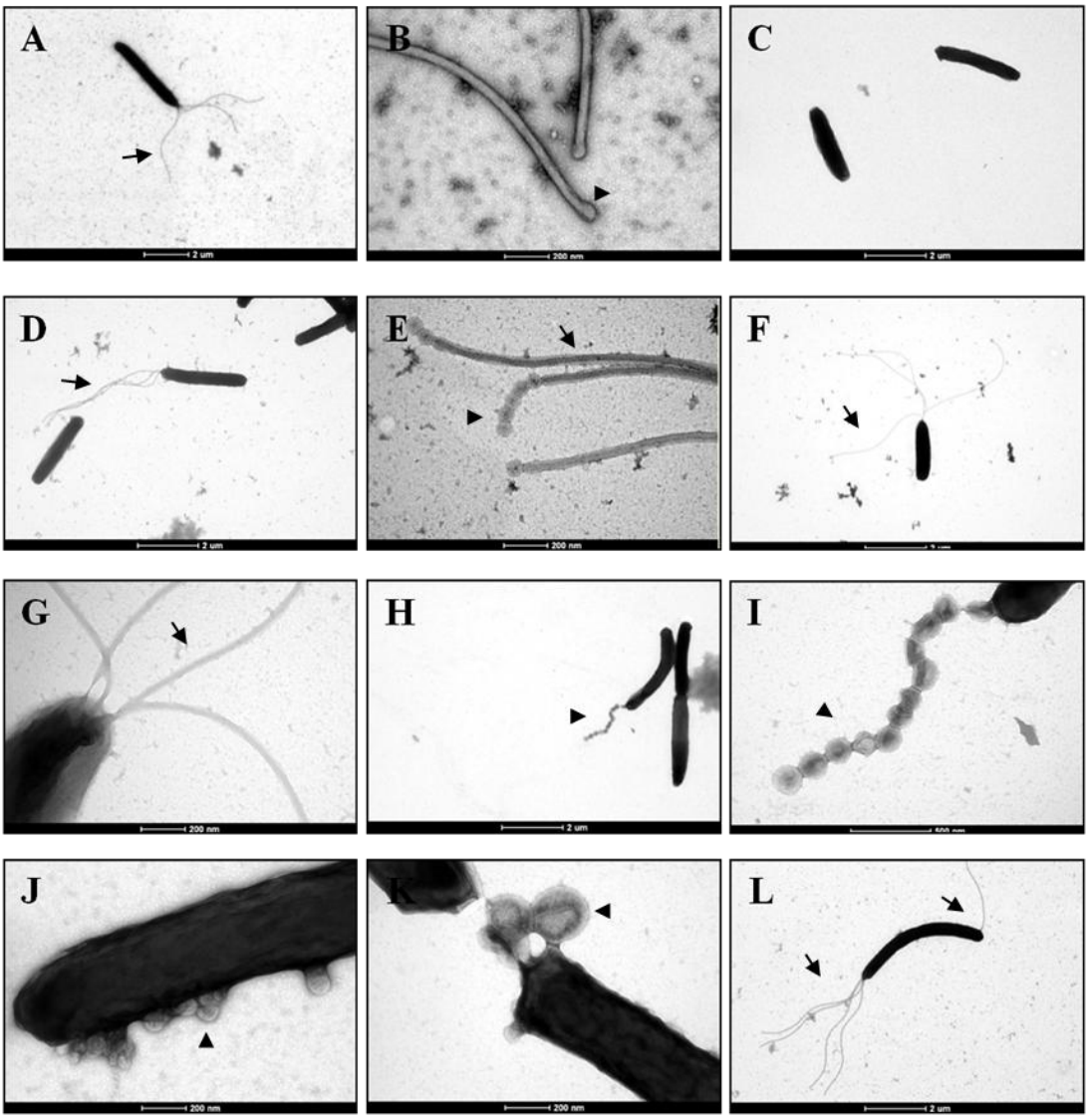
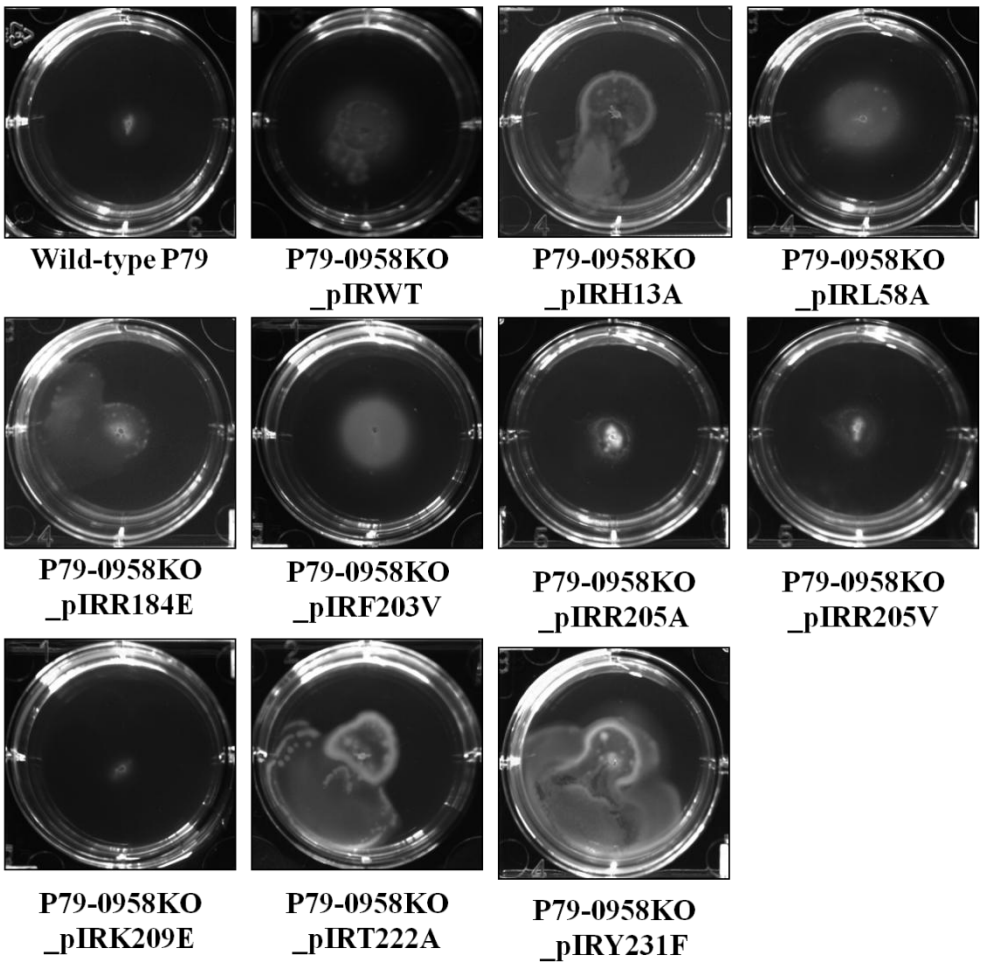


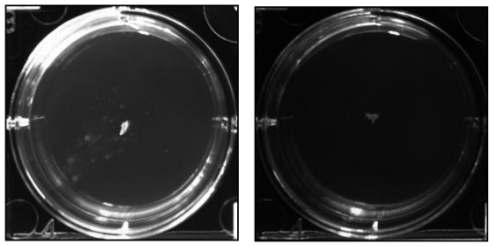
Figure 24 Flagellum production by *H. pylori* P79 and derivatives.

Transmission electron micrographs of *H. pylori* cells stained with uranyl acetate. In each panel, arrows mark flagella and arrowheads mark bulb structures. (A) Wild-type P79; (B) detail of P79 wild-type flagella; (C) *hp0958*-null derivative of P79; (D) complemented *hp0958*-null derivative of P79; (E) detail of complemented derivative; (F) representative example of a flagellate cell from HP0958 mutant complemented with mutated allele; (G) detail of flagellate cell from HP0958 mutant complemented with mutated allele; (H) representative example of a cell with multi-bulb phenotype from HP0958 mutant complemented with mutated allele; (I) detail of multi-bulb phenotype; (J) representative example of a cell with non-polar bulb phenotype from HP0958 mutant complemented with mutated allele; (K) representative example of a cell with large polar bulbs from HP0958 mutant complemented with mutated allele; (L) representative example of a cell with amphitricious flagella from HP0958 mutant complemented with mutated allele.

A Motile



B Non-motile



C Swarming

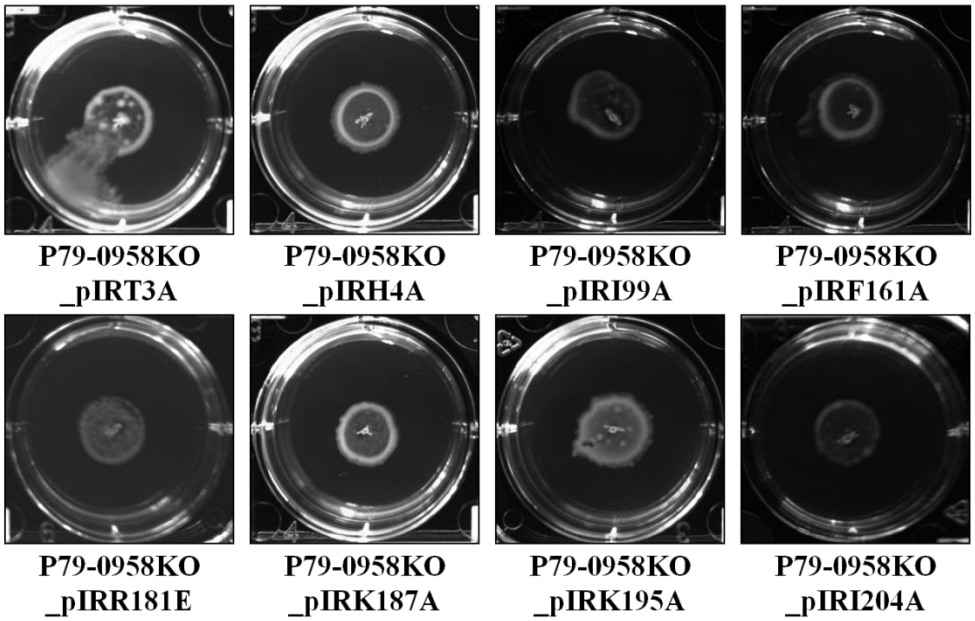


Figure 25 Motility screens of *H. pylori* P79 and derivatives.

Cells were inoculated in 0.3% soft agar and incubated for 4 days at 37°C, 5% CO_2. Halo formation within agar indicates motility while non-motile cells remain at the site of inoculation. Growth outwards from the point of inoculation on the surface only (not in the agar) results in ring pattern due to swarming.

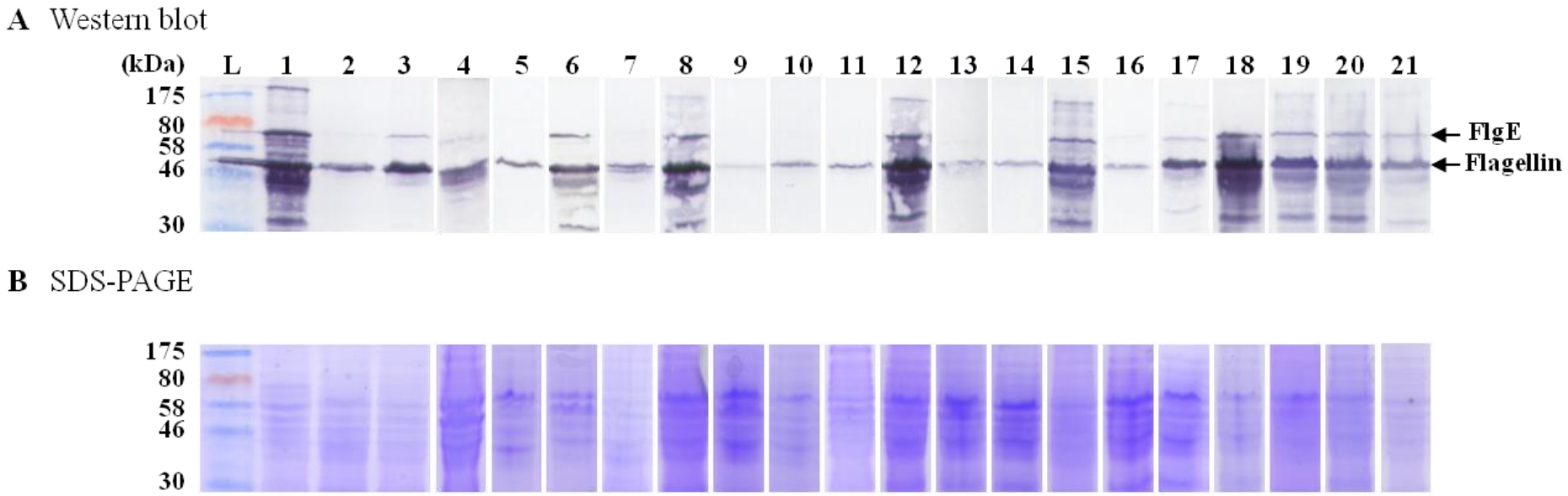


Figure 26 Western blot analysis of flagellum protein expression in P79 and its derivatives.

Flagellin and hook protein levels of cells corrected to OD_{600} 1.0 after 20 hrs liquid culture. (L) ColorPlus™ prestained protein ladder, broad range (7 - 175 kDa); (1) wild-type P79; (2) P79-0958KO; (3) P79-0958KO_pIRWT; (4) P79-0958KO_pIRT3A; (5) P79-0958KO_pIRH4A; (6) P79-0958KO_pIRH13A; (7) P79-0958KO_pIRL47A; (8) P79-0958KO_pIRL58A; (9) P79-0958KO_pIRI99A; (10) P79-0958KO_pIRF161A; (11) P79-0958KO_pIRR181E; (12) P79-0958KO_pIRR184E; (13) P79-0958KO_pIRK187A; (14) P79-0958KO_pIRK195A; (15) P79-0958KO_pIRF203V; (16) P79-0958KO_pIRI204A; (17) P79-0958KO_pIRR205A; (18) P79-0958KO_pIRR205V; (19) P79-0958KO_pIRK209E; (20) P79-0958KO_pIRT222A; (21) P79-0958KO_pIRY231F.

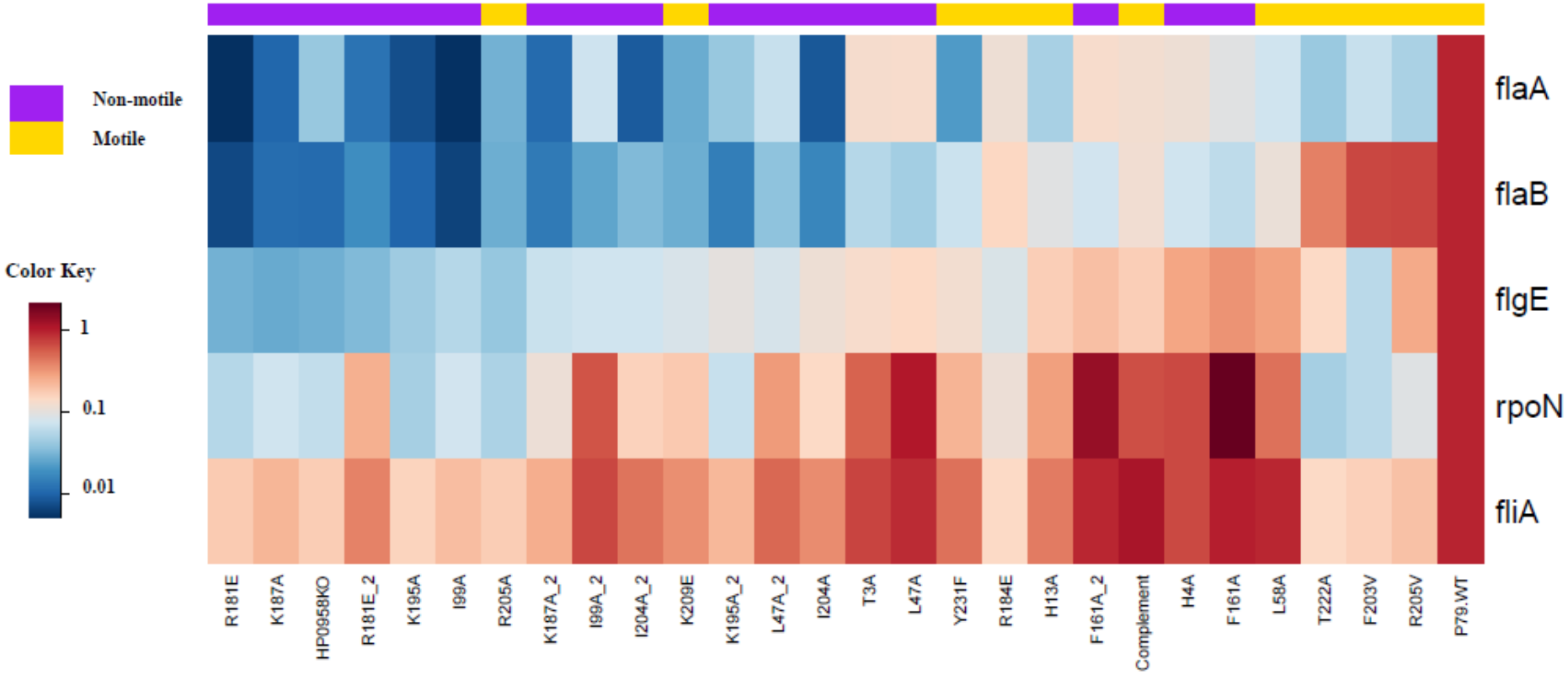


Figure 27 Flagellar gene expression of hp0958-*null* P79 derivative complemented with mutated alleles.

mRNA levels of 5 flagellar genes relative to the wild-type P79 transcription profile. Gene names are listed on the y-axis and strains analysed are listed on the x-axis by the *hp0958* mutant allele they possess. "HP0958KO" is a control for expression in the absence of HP0958 and "complement" is the P79 mutant derivative complemented with the wild-type HP0958. 1 = wild-type level; 0.1 = 10 fold reduction in expression; 0.01 = 100 fold reduction in expression. All gene expression was normalised to the housekeeping gene, *era*. Non-motile strains are coloured purple and motile strains are coloured yellow.

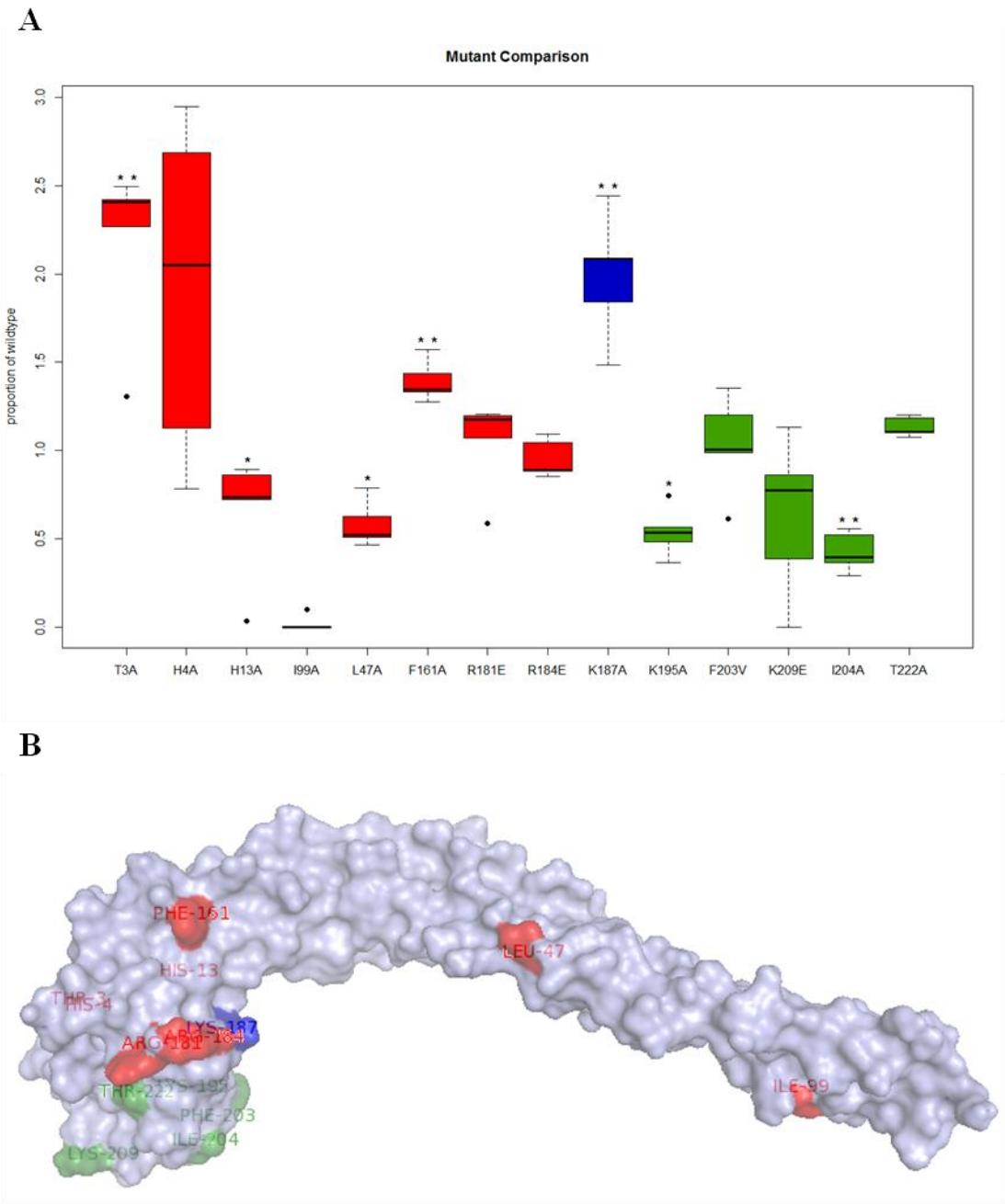
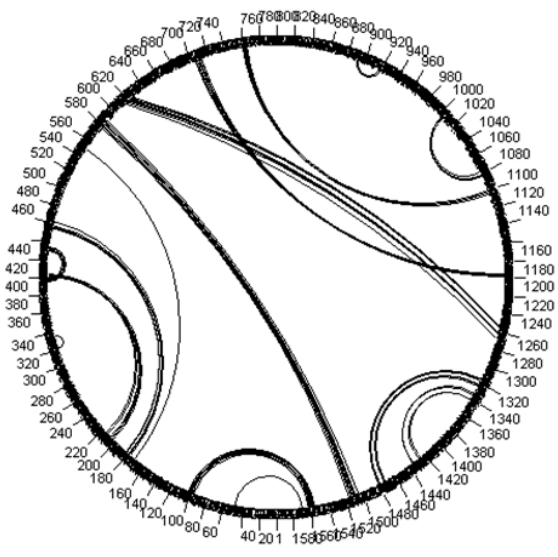


Figure 28 Identification of residues in HP0958 which are involved in the interaction with RpoN (74-284).

(A) β-galactosidase activity of *S. cerevisiae* Y187 derivative strains expressing HP0958 site-directed mutant proteins and RpoN (74-284) represented as a proportion of the signal from the derivative strain harbouring wild-type HP0958/RpoN (74-284) interaction. (**B**) Illustration of the structure of wild-type HP0958 protein highlighting residues tested by Y2H and β-galactosidase assay. Image generated with Pymol (DeLano Scientific, CA). Colours in both (**A**) and (**B**) refer to the secondary structure of residues selected for mutation of HP0958 for each mutant where: red = α-helix; blue = hinge region; green = Zn-finger. Mann-Whitney pairwise statistical test was performed for each pair-wise comparison. *P≤0.05; ** P≤0.01; ***P≤0.001.

A



B

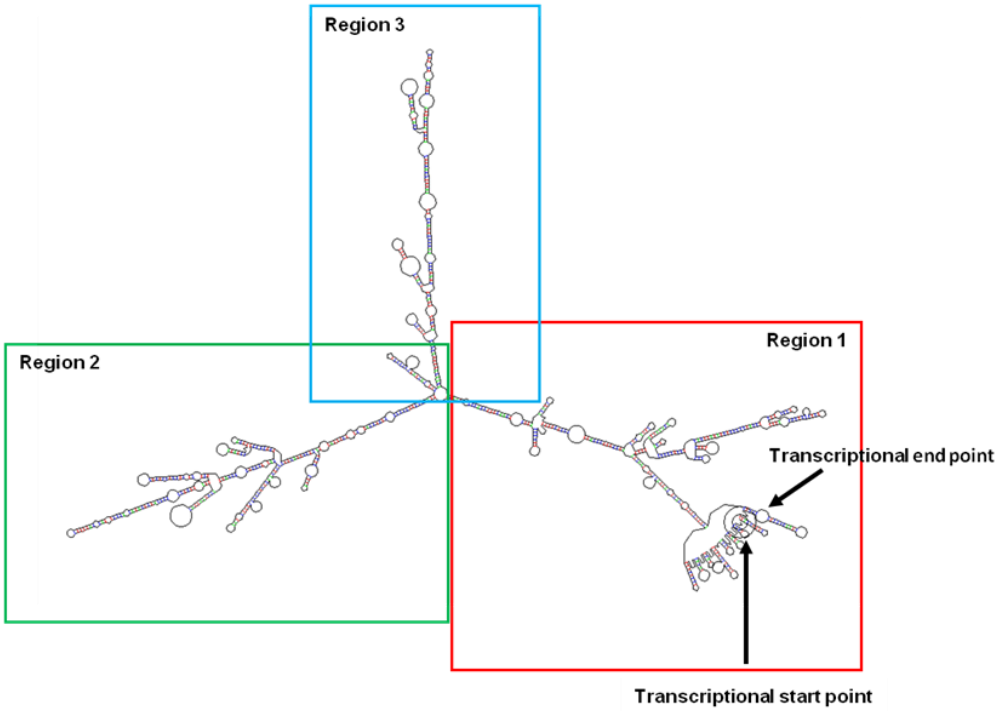


Figure 29 Predicted secondary structure of the *flaA* mRNA transcript.

Predicted secondary structure of *flaA* mRNA transcript of *H. pylori* strain 17874. (**A**) Structure generated in Circles based on maximum weight matching; (**B**) Structure generated in RNAdraw based on McCaskill minimum free energy.

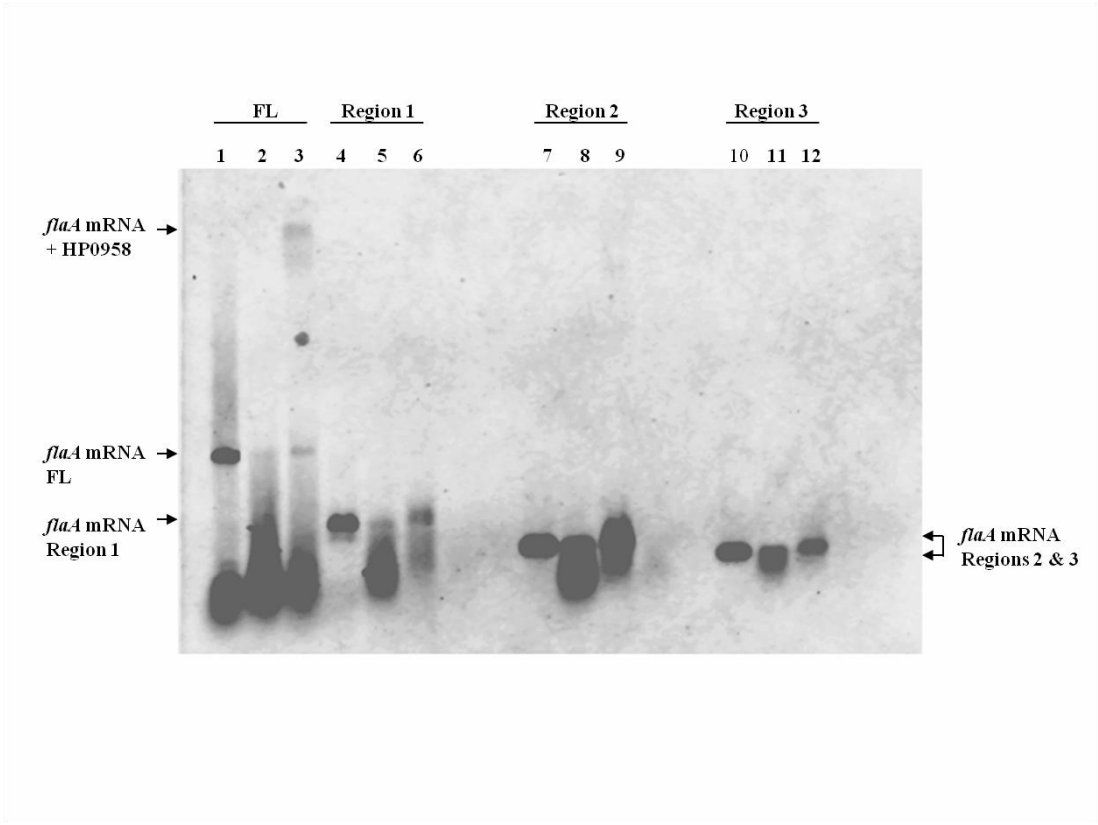


Figure 30 Gel shift assay of HP0958 binding to full length and shortened derivatives of the *flaA* mRNA transcript.

15 ng of biotin-labelled riboprobe was incubated with 6 µg wild type HP0958. RNA corresponding to truncated *flaA* transcripts is labelled Region 1, 2 or 3. Arrows indicate the position of *flaA* transcripts after gel electrophoresis and the different migration of the HP0958/*flaA* mRNA complex. Order as follows: full length *flaA* mRNA (1) RNA load; (2) co-incubation with GST control protein; (3) co-incubation with HP0958; region 1 *flaA* mRNA (4) RNA load; (5) co-incubation with GST control protein; (6) co-incubation with HP0958; region 2 *flaA* mRNA (7) RNA load; (8) co-incubation with GST control protein; (9) co-incubation with HP0958; region 3 *flaA* mRNA (10) RNA load; (11) co-incubation with GST control protein; (12) co-incubation with HP0958.

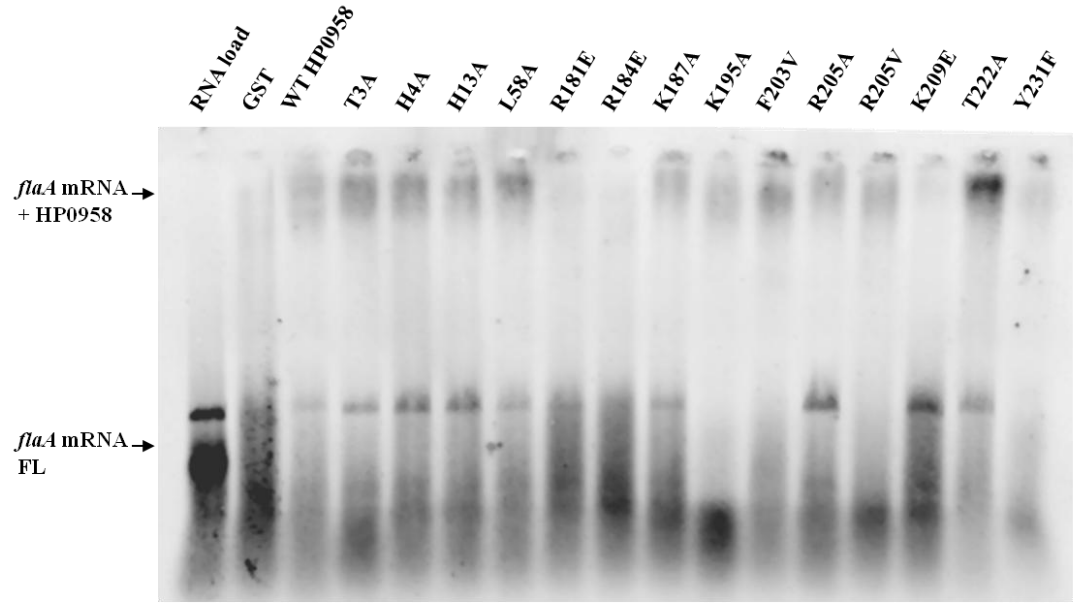


Figure 31 Mobility shift assay screen of HP0958 mutants.

15 ng of biotin-labelled full-length *flaA* riboprobe was incubated with 6 µg HP0958 (wild-type or site-directed mutants). Arrows indicate the position of *flaA* transcripts after gel electrophoresis and the different migration of the HP0958/*flaA* mRNA complex. Controls and site-directed mutants are labelled on x-axis.

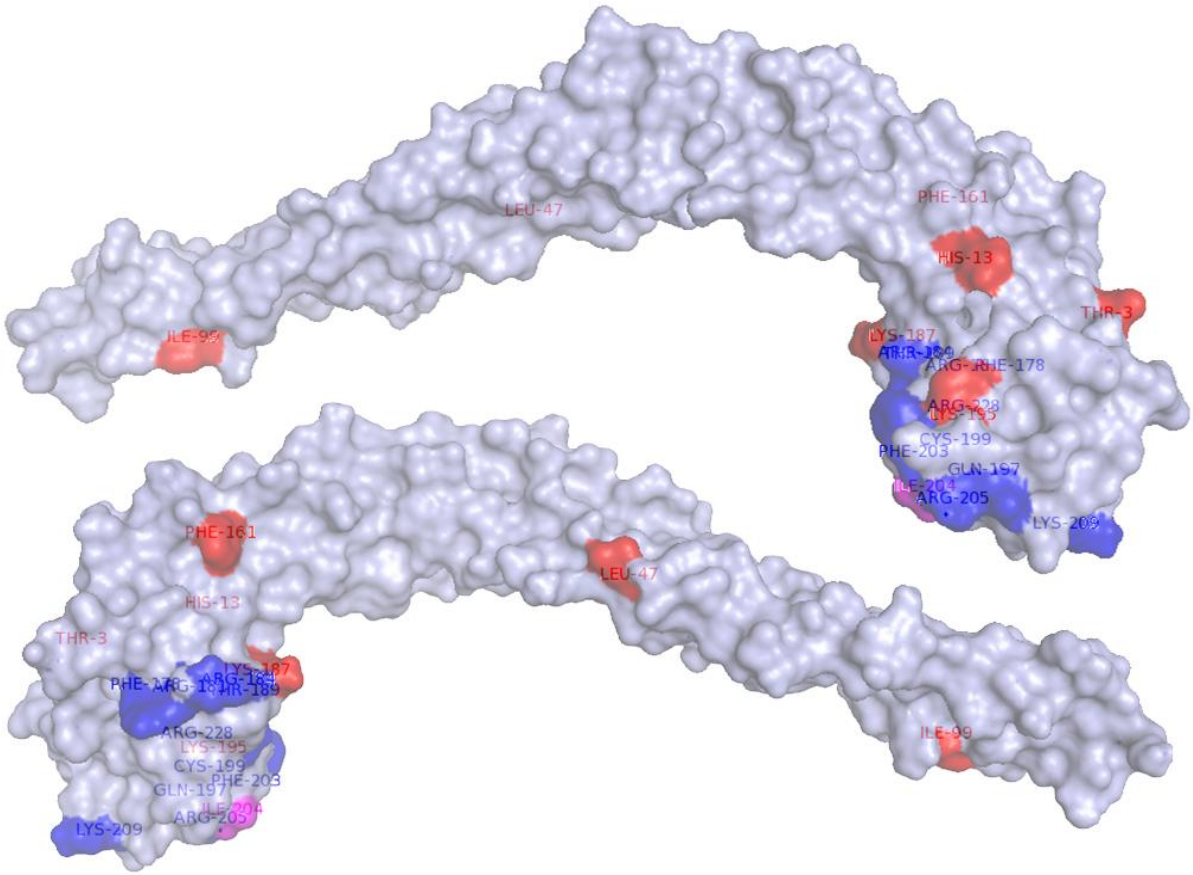


Figure 32 Residues within HP0958 protein involved in interactions with RpoN and/or *flaA* mRNA.

Residues which are involved in the HP0958-RpoN (74-284) interaction are coloured red; those involved in the HP0958-*flaA* mRNA interaction are coloured blue; residue I204A which is involved in both interactions is coloured magenta. Image generated with Pymol (DeLano Scientific, CA) and includes previously published data on the HP0958-*flaA* mRNA interaction by Caly *et al.* (Caly *et al.*, 2010).

8 References

Abramoff MD, Magalhaes PJ, Ram SJ. 2004. Image processing with ImageJ. *Biophotonics Int.* 11: 36–42.

Anderson JK, Smith TG, Hoover TR. 2010. Sense and sensibility: flagellum-mediated gene regulation. *Trends Microbiol.* 18:30–7.

Anderson PE, Gober JW 2000. FlbT, the post-transcriptional regulator of flagellin synthesis in *Caulobacter crescentus*, interacts with the 5' untranslated region of flagellin mRNA. *Mol. Microbiol.* 38: 41–52.

Arnaud M, Chastanet A, Débarbouillé M. 2004. New vector for efficient allelic replacement in naturally nontransformable, low-GC-content, Gram-positive bacteria. *App. Env. Microbiol.* 70: 6887–6891.

Balaban M, Hendrixson DR. 2011. Polar flagellar biosynthesis and a regulator of flagellar number influence spatial parameters of cell division in *Campylobacter jejuni*. *PLoS Pathog.* 7: e1002420.

Balaban M, Joslin SN, Hendrixson DR. 2009. FlhF and its GTPase activity are required for distinct processes in flagellar gene regulation and biosynthesis in *Campylobacter jejuni*. *J. Bacteriol.* 191: 6602–11.

Caly DL, O'Toole PW, Moore SA. 2010. The 2.2-Å structure of the HP0958 protein from *Helicobacter pylori* reveals a kinked anti-parallel coiled-coil hairpin domain and a highly conserved Zn-ribbon domain. *J. Mol. Biol.* 403: 405–19.

Chevance FF V, Hughes KT. 2008. Coordinating assembly of a bacterial macromolecular machine. *Nat. Rev. Microbiol.* 6: 455–65.

Douillard FP, Ryan KA, Caly DL, Hinds J, *et al.* 2008. Posttranscriptional regulation of flagellin synthesis in *Helicobacter pylori* by the RpoN chaperone HP0958. *J. Bacteriol.* 190: 7975–84.

Eaton KA, Morgan DR, Krakowka S. 1992. Motility as a factor in the colonisation of gnotobiotic piglets by *Helicobacter pylori*. *J. Med. Microbiol.* 37: 123–7.

Ellis JJ, Broom M, Jones S. 2007. Protein–RNA interactions: structural analysis and functional classes. *Proteins Struct. Funct. Bioinforma.* 66: 903–911.

Fields S, Song O. 1989. A novel genetic system to detect protein-protein interactions. *Nature* 340: 245–246.

Geis G, Suerbaum S, Forsthoff B, Leying H, Opferkuch W. 1993. Ultrastructure and biochemical studies of the flagellar sheath of *Helicobacter pylori*. *J. Med. Microbiol.* 38: 371–7.

Geva G, Sharan R. 2011. Identification of protein complexes from co-immunoprecipitation data. *Bioinformatics* 27: 111–7.

Goodwin CS, McCulloch RK, Armstrong JA, Wee SH. 1985. Unusual cellular fatty acids and distinctive ultrastructure in a new spiral bacterium (*Campylobacter pyloridis*) from the human gastric mucosa. *J. Med. Microbiol.* 19: 257–267.

Hahn S, Kim D. 2012. Transient protein-protein interaction of the SH3-peptide complex via closely located multiple binding sites. *PLoS One* 7: e32804.

Heckman KL, Pease LR. 2007. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protoc.* 2: 924–32.

Heuermann D, Haas R. 1998. A stable shuttle vector system for efficient genetic complementation of *Helicobacter pylori* strains by transformation and conjugation. *Mol. Gen. Genet.* 257: 519–28.

International Agency for Research on Cancer. 1994. WORLD HEALTH ORGANIZATION IARC Monographs on the evaluation of carcinogenic risks to humans: schistosomes , liver flukes and *Helicobacter pylori*. 61: 177–240.

Iwakiri J, Tateishi H, Chakraborty A, Patil P, Kenmochi N. 2011. Dissecting the protein-RNA interface: the role of protein surface shapes and RNA secondary structures in protein-RNA recognition. *Nucleic Acids Res.* 40: 3299–306.

Jemal A, Bray F, Ferlay J. 2011. Global cancer statistics. *CA Cancer J Clin* 61: 69–90.

Jones S, Daley DT, Luscombe NM, Berman HM, Thornton JM. 2001. Protein-RNA interactions: a structural analysis. *Nucleic Acids Res.* 29: 943–54.

Jones S, Thornton JM. 1996. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* 93: 13–20.

Josenhans C, Eaton KA, Thevenot T, Suerbaum S. 2000. Switching of flagellar motility in *Helicobacter pylori* by reversible length variation of a short homopolymeric sequence repeat in *fliP*, a gene encoding a basal body protein. *Infect. Immun.* 68: 4598–603.

Kim JS, Chang JH, Chung SI, Yum JS. 1999. Molecular cloning and characterization of the *Helicobacter pylori fliD* gene, an essential factor in flagellar structure and motility. *J. Bacteriol.* 181: 6969–6976.

Kostrzynska M, Betts JD, Austin JW, Trust TJ. 1991. Identification, characterization, and spatial localization of two flagellin species in *Helicobacter pylori* flagella. *J. Bacteriol.* 173: 937–46.

Lane MC, O'Toole PW, Moore SA. 2006. Molecular basis of the interaction between the flagellar export proteins FliI and FliH from *Helicobacter pylori*. *J. Biol. Chem.* 281: 508–17.

Langford ML, Zabaleta J, Ochoa AC, Testerman TL, McGee DJ. 2006. *In vitro* and *in vivo* complementation of the *Helicobacter pylori* arginase mutant using an intergenic chromosomal site. *Helicobacter* 11: 477–493.

Lertsethtakarn P, Ottemann KM, Hendrixson DR. 2011. Motility and chemotaxis in *Campylobacter* and *Helicobacter*. *Annu. Rev. Microbiol.* 65: 389–410.

Liao S-M, Du Q-S, Meng J-Z, Pang Z-W, Huang R-B. 2013. The multiple roles of histidine in protein interactions. *Chem. Cent. J.* 7: 44-55.

Linz B, Balloux F, Moodley Y, Manica A, *et al.* 2007. An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* 445: 915–8.

Marshall BJ, Warren JR. 1984. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* 1: 1311–5.

Matzura O, Wennborg A. 1996. RNAdraw: an integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows. *Comput. Appl. Biosci.* 12: 247–249.

McCarter LL. 2006. Regulation of flagella. *Curr. Opin. Microbiol.* 9: 180–6.

Möckli N, Auerbach D. 2004. Quantitative beta-galactosidase assay suitable for high-throughput applications in the yeast two-hybrid system. *Biotechniques* 36: 872–6.

Niehus E, Gressmann H, Ye F, Schlapbach R, *et al.* 2004. Genome-wide analysis of transcriptional hierarchy and feedback regulation in the flagellar system of *Helicobacter pylori*. *Mol. Microbiol.* 52: 947–61.

Niehus E, Ye F, Suerbaum S, Josenhans C. 2002. Growth phase-dependent and differential transcriptional control of flagellar genes in *Helicobacter pylori*. *Microbiology* 148: 3827–37.

O′Toole PW, Kostrzynska M, Trust TJ. 1994. Non-motile mutants of *Helicobacter pylori* and *Helicobacter mustelae* defective in flagellar hook production. *Mol. Microbiol.* 14: 691–703.

Ozbabacan SEA, Engin HB, Gursoy A, Keskin O. 2011. Transient protein-protein interactions. *Protein Eng. Des. Sel.* 24: 635–48.

Page RDM. 2000. Circles: automating the comparative analysis of RNA secondary structure. *Bioinformatics* 16: 1042–1043.

Pereira L, Hoover TR. 2005. Stable accumulation of sigma 54 in *Helicobacter pylori* requires the novel protein HP0958. *J. Bacteriol.* 187: 4463–4469.

Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29: e45.

Pounder RE, Ng D. 1995. The prevalence of *Helicobacter pylori* infection in different countries. *Aliment. Pharmacol. Ther.* 9: 33–9.

Rain JC, Selig L, De Reuse H, Battaglia V, *et al.* 2001. The protein-protein interaction map of *Helicobacter pylori*. *Nature* 409: 211–5.

Ryan KA, Karim N, Worku M, Moore SA, *et al.* 2005. HP0958 is an essential motility gene in *Helicobacter pylori*. *FEMS Microbiol. Lett.* 248: 47–55.

Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: a laboratory manual.* 2nd ed. Cold Spring Harbor Laboratory, New York, NY.

Smith PK, Krohn RI, Hermanson GT, Mallia AK, *et al.* 1985. Measurement of protein using bicinchoninic acid. *Anal. Biochem.* 150: 76–85.

Stynen B, Tournu H, Tavernier J, Dijck P Van. 2012. Diversity in genetic in vivo methods for protein-protein interaction studies: from the yeast two-hybrid system to the mammalian split-luciferase system. *Microbiol. Mol. Biol. Rev.* 76: 331–82.

Tang X, Bruce JE. 2009. Chemical cross-linking for protein-protein interaction studies. In *Mass Spectrometry of Proteins and Peptides* (M.S. Lipton and L. Paša-Toliceds), Vol. 492 of, pp. 283–293, Springer.

The Eurogast Study Group. 1993. Epidemiology of, and risk factors for, *Helicobacter pylori* infection among 3194 asymptomatic subjects in 17 populations. *Gut* 34: 1672–6.

Tomb JF, White O, Kerlavage AR, Clayton RA, *et al.* 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388: 539–47.

Towbin H, Staehelin T, Gordon J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. U. S. A.* 74: 4350–4354.

Untergasser A, Cutcutache I, Koressaar T, Ye J, *et al.* 2012. Primer3-new capabilities and interfaces. *Nucleic Acids Res.* 40: e115.

Warren JR, Marshall BJ. 1983. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. Lancet 321: 1273–1275.

Xia J-F, Zhao X-M, Huang D-S. 2010. Predicting protein-protein interactions from protein sequences using meta predictor. *Amino Acids* 39: 1595–9.

Zhang QC, Petrey D, Deng L, Qiang L, *et al.* 2012. Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature* 490: 556–60.

General Discussion

1 General Discussion

Over the last ~30 years, research into the gastric human pathogen *Helicobacter pylori* has generated major advancements in our understanding of its biology, pathogenicity and intimate host associations. The classification of *H. pylori* as a human carcinogen, emerging antibiotic resistances and globally high levels of gastric diseases highlight its continuing clinical relevance throughout the decades (International Agency for Research on Cancer, 1994; Marshall *et al.*, 1984; Parkin, 2006). Genome mining has lead to the identification of strain-specific genes, biomarkers of disease and provided insights into the regulatory network of gene expression in *H. pylori*. Due to close evolution with its human host, *H. pylori* has developed a genome which is extremely efficient for performing critical processes during its ecological cycle such as infection, colonisation and evasion of the host immune response (Backert and Clyne, 2011; Linz *et al.*, 2007). The genome of *H. pylori* appears to be in a constant state of flux, with recombination events, phase variation and mutation rates enabling the extreme plasticity (Olbermann *et al.*, 2010; Suerbaum and Josenhans, 2007).

In Chapter 2 the draft genome sequences of two *H. pylori* strains, CCUG 17874 and P79, were described as a published Genome Announcement (Clancy *et al.*, 2012). Comparative analysis of strains 17874 and P79 identified key differences in predicted pseudogenes, genes encoding outer membrane proteins, restriction modification (RM) systems and competence genes. Regarding the difference in transformation efficiency between the strains, it is likely that the restriction barrier is the main factor which makes plasmid introduction into 17874 more difficult, while variations in the *comB* locus may also contribute. The development of methylomics has recently led to the identification of novel recognition sequences for RM systems, as well as highlighting remarkable levels of interstrain variation (Krebes *et al.*, 2013; Murray *et al.*, 2012). The mere absence or presence of a particular RM gene cannot be accepted as sufficient evidence for interstrain variation of biological relevance since RM system components have previously been shown to be present but inactive in some strains of *H. pylori* (Xu *et al.*, 2000). The relatively recently described derivative strain of 26695 *H. pylori* which was deficient in type II RM systems had increased transformation efficiency, highlighting an exciting new method of enhancing the tractability of lab strains which are difficult to transform (Zhang and

Blaser, 2012). Application of this strategy to the motile type-strain 17874, together with the publically available draft genome sequence for this strain, would greatly enhance motility studies requiring genetic manipulations.

Interestingly, each person infected with *H. pylori* possesses unique strain(s), with the exception of close family members, highlighting the adaptability of this human pathogen (Fialho *et al.*, 2010; Raymond *et al.*, 2008; Schwarz *et al.*, 2008). The extreme genetic diversity of *H. pylori* and compounding factors such as host genetics and environmental factors have hampered efforts to identify clear disease markers (Chung *et al.*, 2010; Kodaman *et al.*, 2014; Neal *et al.*, 2013; Wroblewski *et al.*, 2010). Presence of cytotoxin associated gene A (*cagA*) and vacuolating cytotoxin A (*vacA*) gene are two of the best described virulence determinants associated with *H. pylori*-related disease development (Backert and Clyne, 2011). However, strains lacking the standard characteristed virulence factors including SabA and BabA have been isolated from patients suffering from gastric neoplasia (Thiberge *et al.*, 2010). This indicates that *H. pylori* may harbour virulence determinants which are currently unknown.

Refinement of the core genome and pan-genome of *H. pylori* will contribute to global understanding of the mechanisms of infection and persistence in its host. In this study, the core genome of *H. pylori* was further reduced in comparison to recent analyses (Lu *et al.*, 2013). Comparative analyses of the species core genome with the core genome of strains isolated from individuals suffering from *H. pylori*-related diseases ("disease-core") has resulted in a panel of genes which can be probed in further analyses. Many of the "disease-core" genes encoded hypothetical proteins whose characterisation is warranted based on the potential for these to be novel biomarkers of disease or virulence factors. It is likely that these hypothetical proteins have functions which are advantageous to these strains either in colonisation or induction of disease-onset, as *H. pylori* is an extremely efficient organism in terms of gene conservation. Many of these genes may encode proteins involved in LPS biosynthesis, outer membrane proteins or transporter proteins, all of which are contributors of virulence, effective colonisation and persistence. For instance, HP1286 was originally annotated as encoding a hypothetical protein in the genome of strain 26695, but has since been identified as a stress-response factor homologous to YceI . The high number of genes encoding hypothetical proteins annotated in the genomes of strains 17874 and P79 is striking, considering the relatively small

genome size of *H. pylori* and the large number of strains for which genome sequence data is available. Clearly, the presence of so many (often strain-specific) genes encoding proteins of unknown function undermines our understanding of the complexity of *H. pylori* as a persistent human pathogen. Systematic characterisation of these genes would undoubtedly enhance future studies regarding infection, colonisation and the on-set of disease triggered by *H. pylori*. Furthermore, the identification of homopolymeric repeats in the genomes of strains 17874 and P79 highlights the role of phase variation in *H. pylori* regulation of gene expression of factors such as motility and surface-exposed antigens. The link has been made between phase variation and epigenetic regulation of gene expression in *H. pylori* (Srikhanta *et al.*, 2011). This mechanism of regulation adds another layer of complexity to the issue of pathogenesis. Therefore, a challenge now exists to develop genome data mining tools which are capable of accurately annotating and filtering the sequence information to enhance targeted approaches to future analyses. The coupling of genomics, transcriptomics and the relatively recent methylomics to biological analyses will help to elucidate the intricacies of gene expression and regulation, marking an exciting new era in *H. pylori* research (Murray *et al.*, 2012; Sharma *et al.*, 2010).

In Chapter 3, the focus of this study was shifted from a broad scale genome analysis to a detailed structure-function analysis of the essential *H. pylori* flagellum biogenesis protein, HP0958. Although structural elements of the *H. pylori* flagellum are closely conserved when compared to model organisms *E. coli* and *Salmonella*, regulation of flagellar assembly deviates from these (Anderson *et al.*, 2010; Lertsethtakarn *et al.*, 2011; Niehus *et al.*, 2004). Elucidation of the regulatory mechanisms of flagellum biogenesis in *H. pylori* will help to understand this complex process and may contribute to knowledge of spatial regulation of flagella in motile bacteria as *H. pylori* are lophotrichously flagellated while the model organisms are peritrichously flagellated.

In this study, the previously proposed interaction of HP0958 with FliH, the negative regulator of flagellar ATPase FliI, was found to be a false positive from a yeast-two hybrid (Y2H) analysis of the proteome of *H. pylori* (Rain *et al.*, 2001). Confirmation of the HP0958-RpoN (74-284) interaction by Y2H analysis and subsequent site-directed mutagenesis identified the involvement of the coiled-coil domain of HP0958. HP0958 was previously identified as a chaperone of RpoN

(Douillard *et al.*, 2008; Pereira and Hoover, 2005). The HP0958-RpoN interaction was identified as a relatively weak interaction compared to FliH-FliI binding. One possibility is that the absence of currently unidentified additional interaction partners prevents the formation of a stable complex in the Y2H system. Another scenario is that the weak, transient nature of the HP0958-RpoN interaction is biologically favourable, allowing the interaction of RpoN with the core RNA polymerase (residues ~70-200 in *Aquifex aeolicus*) to stimulate expression of Class II flagellar genes (Hong *et al.*, 2009). The role of HP0958 in this case may be to protect RpoN from intracellular proteases, prevent aggregation or induce RpoN to take on an extended conformation which exposes the core RNA polymerase binding site. While it is known that HP0958 improves the stability of RpoN in the cytoplasm, the potential involvement of HP0958 in RpoN interactions with the core RNA polymerase has yet to be investigated (Pereira and Hoover, 2005).

Complementation of a non-motile *hp0958*-null derivative of P79 with HP0958 mutant alleles by natural transformation revealed that mutations in the coiled-coil generally resulted in non-motile cells, while those in the Zn-finger generally restored motility either partially or fully. Interestingly, electron microscopy of complemented mutant strains revealed presence of flagellar-type extrusions from the cell surface at both poles and/or non-polar sites. This suggests that HP0958 may have an additional role in localisation of flagellum biogenesis to a single cell pole; this process likely involves both the coiled-coil and Zn-finger domains of HP0958. Little is known about the mechanism of flagellum localisation in *H. pylori*. FlhF and FlhG have been implicated in localisation of flagellum biogenesis to the bacterial cell pole and control of flagellum number in *H. pylori*, *Campylobacter jenuni* and *Vibrio* (Balaban and Hendrixson, 2011; Balaban *et al.*, 2009; Lertsethtakarn *et al.*, 2011). Balaban *et al.* proposed that in *C. jejuni*, FlhF is involved in the activation of RpoN-dependent gene expression in an FlgS/R-independent mechanism (Balaban and Hendrixson, 2011). Recently, a novel role for flagellar-associated autotoxin A (FaaA) has been proposed in localising flagellum production to the cell pole in *H. pylori* (Radin *et al.*, 2013). Site-directed mutagenesis of HP0958 enabled refinement of the current model for HP0958-*flaA* mRNA interaction. The previously published structure-function analysis of the HP0958-*flaA* mRNA interaction was augmented by analysis of a further 14 side-directed mutants combined with analysis of the RNA subdomains required for interaction (Caly *et al.*, 2010). The full length transcript of the major

flagellin was found to be required for direct full-strength interaction with the Zn-finger domain of HP0958.

Given that FliH does not interact with HP0958, the currently accepted model for the role of HP0958 during flagellum biogenesis must be re-evaluated (Douillard *et al.*, 2008). Components of the basal body and export apparatus may fill the previously proposed role of FliH in this model. Elucidation of the role of HP0958 in flagellar localisation to the cell pole may be the key. One scenario could involve interaction between RpoN-bound HP0958 and FlhF. FlhF has been implicated in localising the MS ring components to the cell pole in *C. jenuni*, in addition to its influence on the activity of RpoN (Balaban *et al.*, 2009). Therefore, future studies to identify potential interactions between HP0958, RpoN, FlhF and potentially FlhG may identify the missing link which targets the HP0958-*flaA* mRNA complex to the export apparatus. Furthermore, together with FaaA, these proteins may be involved in a cascade which ensures the tight spatial regulation of flagellum biogenesis at a single cell pole. In *Vibrio alginolyticus*, FlhG has been identified to negatively regulate the activity of FlhF (Kusumoto *et al.*, 2008). Elucidation of the potential interplay between these flagellar proteins will serve to improve the model of flagellum construction in *H. pylori*.

Additionally, many of the P79-0958KO derivative strains complemented with HP0958 mutant proteins produced truncated flagellar-type structures which resembled flagellar distal bulbs. Mutation of the *fliD* gene also produced mutant cells which possessed atypical flagellar sheaths, which are thought to be of similar composition to that of the bacterial outer membrane (Geis *et al.*, 1993; Kim *et al.*, 1999). FaaA localises to the flagellar sheath in *H. pylori* and deletion of the gene resulted in cells with abnormal cell localisation of the flagellum (Radin *et al.*, 2013). There have been relatively few developments in our understanding of the biosynthesis of the flagellar sheath. Geis *et al.* found that the sheath contains fatty acids, LPS and low molecular weight proteins, and suggested that it is an extension of the outer membrane (Geis et al., 1993). In addition, only two proteins have since been identified as localising to the flagellar sheath: the autotransporter FaaA and HpaA (Jones et al., 1997; Luket and Penn, 1995; Radin et al., 2013). Therefore, the panel of derivative strains possessing HP0958 mutant proteins from this study provides a platform for analysis of flagellar sheath biogenesis. Future analysis of the essential flagellum biogenesis protein HP0958 will provide understanding of the

deviations of flagellum construction from the model organisms and potentially enhance our understanding of flagellar localisation and sheath production.

2 References

Anderson JK, Smith TG, Hoover TR. 2010. Sense and sensibility: flagellum-mediated gene regulation. *Trends Microbiol.* 18: 30–7.

Backert S, Clyne M. 2011. Pathogenesis of *Helicobacter pylori* infection. *Helicobacter* 16: 19–25.

Balaban M, Hendrixson DR. 2011. Polar flagellar biosynthesis and a regulator of flagellar number influence spatial parameters of cell division in *Campylobacter jejuni*. *PLoS Pathog.* 7: e1002420.

Balaban M, Joslin SN, Hendrixson DR. 2009. FlhF and its GTPase activity are required for distinct processes in flagellar gene regulation and biosynthesis in *Campylobacter jejuni*. *J. Bacteriol.* 191: 6602–11.

Caly DL, O'Toole PW, Moore SA. 2010. The 2.2-Å structure of the HP0958 protein from *Helicobacter pylori* reveals a kinked anti-parallel coiled-coil hairpin domain and a highly conserved Zn-ribbon domain. *J. Mol. Biol.* 403: 405–19.

Chung C, Olivares A, Torres E, Yilmaz O, *et al.*. 2010. Diversity of VacA intermediate region among *Helicobacter pylori* strains from several regions of the world. *J. Clin. Microbiol.* 48: 690–6.

Clancy CD, Forde BM, Moore SA, O'Toole PW. 2012. Draft genome sequences of *Helicobacter pylori* strains 17874 and P79. *J. Bacteriol.* 194: 2402.

Douillard FP, Ryan KA, Caly DL, Hinds J, *et al.*. 2008. Posttranscriptional regulation of flagellin synthesis in *Helicobacter pylori* by the RpoN chaperone HP0958. *J. Bacteriol.* 190: 7975–84.

Fialho AMN, Braga ABC, Braga Neto MB, Carneiro JG, *et al.*. 2010. Younger siblings play a major role in *Helicobacter pylori* transmission among children from a low-income community in the Northeast of Brazil. *Helicobacter* 15: 491–6.

Geis G, Suerbaum S, Forsthoff B, Leying H, Opferkuch W. 1993. Ultrastructure and biochemical studies of the flagellar sheath of *Helicobacter pylori*. *J. Med. Microbiol.* 38: 371–7.

Hong E, Doucleff M, Wemmer DE. 2009. Structure of the RNA polymerase core-binding domain of sigma(54) reveals a likely conformational fracture point. *J. Mol. Biol.* 390: 70–82.

International Agency for Research on Cancer. 1994. WORLD HEALTH ORGANIZATION IARC Monographs on the evaluation of carcinogenic risks to Humans: schistosomes , liver flukes and *Helicobacter pylori*. 61: 177–240.

Jones AC, Logan RP, Foynes S, Cockayne A, *et al.*. 1997. A flagellar sheath protein of *Helicobacter pylori* is identical to HpaA, a putative hemagglutinin, but is not an adhesin for AGS cells. *J. Bacteriol.* 179: 5643–5647.

Kim JS, Chang JH, Chung S Il, Yum JS. 1999. Molecular cloning and characterization of the *Helicobacter pylori fliD* gene , an essential factor in flagellar structure and motility. *J. Bacteriol.* 181: 6969–6976.

Kodaman N, Pazos A, Schneider BG, Piazuelo MB, *et al.*. 2014. Human and *Helicobacter pylori* coevolution shapes the risk of gastric disease. *Proc. Natl. Acad. Sci. U. S. A.* 111: 1455–60.

Krebes J, Morgan RD, Bunk B, Spröer C, *et al.*. 2013. The complex methylome of the human gastric pathogen *Helicobacter pylori*. *Nucleic Acids Res.* 1–18.

Kusumoto A, Shinohara A, Terashima H, Kojima S, *et al.*. 2008. Collaboration of FlhF and FlhG to regulate polar-flagella number and localization in *Vibrio alginolyticus*. *Microbiology* 154: 1390–9.

Lertsethtakarn P, Ottemann KM, Hendrixson DR. 2011. Motility and chemotaxis in *Campylobacter* and *Helicobacter* . *Annu. Rev. Microbiol.* 65: 389–410.

Linz B, Balloux F, Moodley Y, Manica A, *et al.*. 2007. An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* 445: 915–8.

Lu W, Wise MJ, Tay CY, Windsor HM, *et al.*. 2013. Comparative analysis of the full genome of the *Helicobacter pylori* isolate, Sahul64, identifies genes of high divergence. *J. Bacteriol.* 2: doi:10.1128/JB.01021–13.

Luket CJ, Penn CW. 1995. Identification of a 29 kDa flagellar sheath protein in *Helicobacter pylori* using a murine monoclonal antibody. *Microbiology* 141: 597–604.

Marshall BJ, Royce H, Annear D. 1984. Original isolation of *Campylobacter pyloris* from human gastric mucosa. *Microbios Lett* 25: 83–88.

Murray IA, Clark TA, Morgan RD, Boitano M, *et al.*. 2012. The methylomes of six bacteria. *Nucleic Acids Res.* 40: 11450–62.

Neal JT, Peterson TS, Kent ML, Guillemin K. 2013. *H. pylori* virulence factor CagA increases intestinal cell proliferation by Wnt pathway activation in a transgenic zebrafish model. *Dis. Model. Mech.* 6: 802–10.

Niehus E, Gressmann H, Ye F, Schlapbach R, *et al.*. 2004. Genome-wide analysis of transcriptional hierarchy and feedback regulation in the flagellar system of *Helicobacter pylori*. *Mol. Microbiol.* 52: 947–61.

Olbermann P, Josenhans C, Moodley Y, Uhr M, *et al.*. 2010. A global overview of the genetic and functional diversity in the *Helicobacter pylori* cag pathogenicity island. *PLoS Genet.* 6: e1001069.

Parkin DM. 2006. The global health burden of infection-associated cancers in the year 2002. *Int. J. Cancer* 118: 3030–44.

Pereira L, Hoover TR. 2005. Stable accumulation of sigma 54 in *Helicobacter pylori* requires the novel protein HP0958. *J. Bacteriol.* 187: 4463–4469.

Radin JN, Gaddy JA, González-Rivera C, Loh JT, *et al.*. 2013. Flagellar localization of a *Helicobacter pylori* autotransporter protein. *MBio* 4: e00613-12.

Rain JC, Selig L, Reuse H De, Battaglia V, *et al.*. 2001. The protein-protein interaction map of *Helicobacter pylori*. *Nature* 409: 211–5.

Raymond J, Thiberge J-M, Kalach N, Bergeret M, *et al.*. 2008. Using macro-arrays to study routes of infection of *Helicobacter pylori* in three families. *PLoS One* 3: e2259.

Schwarz S, Morelli G, Kusecek B, Manica A, *et al.*. 2008. Horizontal versus familial transmission of *Helicobacter pylori*. *PLoS Pathog.* 4: e1000180.

Sharma CM, Hoffmann S, Darfeuille F, Reignier J, *et al.*. 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* 464: 250–5.

Srikhanta YN, Gorrell RJ, Steen JA, Gawthorne JA, *et al.*. 2011. Phasevarion mediated epigenetic gene regulation in *Helicobacter pylori*. *PLoS One* 6: e27569.

Suerbaum S, Josenhans C. 2007. *Helicobacter pylori* evolution and phenotypic diversification in a changing host. *Nat. Rev. Microbiol.* 5: 441–52.

Thiberge J-M, Boursaux-Eude C, Lehours P, Dillies M-A, *et al.*. 2010. From array-based hybridization of *Helicobacter pylori* isolates to the complete genome sequence of an isolate associated with MALT lymphoma. *BMC Genomics* 11: 368-379.

Wroblewski LE, Peek RM, Wilson KT. 2010. *Helicobacter pylori* and gastric cancer: factors that modulate disease risk. *Clin. Microbiol. Rev.* 23: 713–39.

Xu Q, Morgan RD, Roberts RJ, Blaser MJ. 2000. Identification of type II restriction and modification systems in *Helicobacter pylori* reveals their substantial diversity among strains. *Proc. Natl. Acad. Sci. U. S. A.* 97: 9671–6.

Zhang X-S, Blaser MJ. 2012. Natural transformation of an engineered *Helicobacter pylori* strain deficient in type II restriction endonucleases. *J. Bacteriol.* 194: 3407–16.

Appendices

Appendix 1 List of methylases in the genome of *H. pylori* CCUG 17874

Locus tag	Contig	Product
HP17_00556	13	Type II adenine methyltransferase
HP17_00561	13	Cytosine specific DNA methyltransferase
HP17_00576	13	Adenine/cytosine DNA methyltransferase
HP17_00811	15	Type II adenine methyltransferase
HP17_04995	88	Type II DNA modification methyltransferase
HP17_05010	88	Type II DNA modification (methyltransferase)
HP17_05065	88	Flagellin N-methylase, FliB
HP17_05280	91	DNA methylase
HP17_07432	144	N5-glutamine S-adenosyl-L-methionine-dependent methyltransferase
HP17_07164	140	Type II DNA modification enzyme (methyltransferase)
HP17_07154	140	Type II DNA modification enzyme
HP17_05865	100	rRNA methyltransferase
HP17_05645	99	Adenine-specific DNA methyltransferase
HP17_06342	110	Putative type II cytosine specific methyltransferase
HP17_08696	165	N6-adenine-specific methylase
HP17_02549	41	Adenine-specific DNA methyltransferase
HP17_02150	35	Ribosomal protein L11 methyltransferase
HP17_02065	34	Pore-forming cytolysin (rRNA methylase)
HP17_003479	68	M.HpyAVIII, type II cytosine specific DNA methyltransferase
HP17_04079/84	73/74	Adenine-specific DNA methyltransferase
HP17_02808	44	Adenine-specific DNA methyltransferase
HP17_01438	24	DNA adenine methylase
HP17_01448	24	Adenine-specific DNA methylase
HP17_01453	24	Adenine-specific DNA methylase
HP17_07027	133	Adenine-specific DNA methyltransferase
HP17_06947	131	Type III restriction enzyme M protein
HP17_06942	131	Adenine-specific DNA methylase
HP17_03874	72	Ubiquinone/menaquinone biosynthesis methyltransferase
HP17_03959	73	Type II methylase
HP17_03964	73	Type II adenine specific DNA methyltransferase
HP17_04104	75	Type III R-M system modification enzyme
HP17_04286	80	Putative type II methylase protein

Appendix 2 List of endonucleases in the genome of *H. pylori* CCUG 17874

Locus tag	Contig	Product
HP17_00566	13	Restriction endonuclease
HP17_05690	99	Endonuclease III
HP17_05650	99	Type III restriction enzyme R protein (Res)
HP17_05590	98	3-methyladenine DNA glycosylase
HP17_08516	164	HP0790-like protein
HP17_08626	165	Hypothetical protein
HP17_08144	154	Type I restriction enzyme subunit S
HP17_03189	59	Putative type I restriction-modification enzyme specificity subunit S
HP17_03354	65	Holliday junction resolvase
HP17_02464	39	Type I restriction/modification specificity protein
HP17_02095	34	Similar to archaeal Holliday junction resolvase and Mrr protein
HP17_01888	31	Holliday junction resolvase-like protein
HP17_06877	130	Putative endonuclease G
HP17_04169	75	Restriction modification system DNA specificity domain protein
HP17_06677	125	Mulitfunctional nucleoside diphosphate kinase/apyrimidinic endonuclease/ 3'-phosphodiesterase
HP17_05005	88	Type II R-M system restriction endonuclease
HP17_08484	162	Type II restriction endonuclease
HP17_07159	140	Type II restriction endonuclease
HP17_02554	41	Restriction endonuclease Hpy8I
HP17_01443	24	Type II restriction endonuclease
HP17_07037	133	Type II R-M system restriction endonuclease

Appendix 3 List of all methylases in the genome of *H. pylori* P79

Locus tag	Contig	Product
HP79_04107	111	tRNA mo(5)U34 methyltransferase
HP79_04916	128	Adenine-specific DNA methyltransferase
HP79_04911	128	Cytosine specific DNA methyltransferase
HP79_04891	128	Adenine/cytosine DNA methyltransferase
HP79_04677	126	Type II restriction enzyme M protein (hsdM)
HP79_08238	199	rRNA large subunit methyltransferase
HP79_08602	203	Adenine-specific DNA methyltransferase
HP79_08622	204	Adenine-specific DNA methyltransferase
HP79_08677	204	Flagellin N-methylase, FliB
HP79_05531	139	N5-glutamine S-adenosyl-L-methionine-dependent methyltransferase
HP79_02349	68	RNA methyltransferase
HP79_02124	64	Adenine-specific DNA methyltransferase
HP79_03041	86	N6-adenine-specific methylase
HP79_01635	43	Adenine-specific DNA methyltransferase
HP79_07126	173	Ribosomal protein L11 methyltransferase
HP79_07026	171	Haemolysin (tly), rRNA methylase
HP79_07151	173	16S rRNA methyltransferase GidB
HP79_09002	214	Cytosine specific DNA methyltransferase (BSP6IM)
HP79_03386	95	tRNA (guanine-N(7)-)-methyltransferase
HP79_03698	102	Adenine-specific DNA methyltransferase
HP79_05764	142	Adenine-specific DNA methy ltransferase
HP79_05779	145	Adenine-specific DNA methyltransferase
HP79_05854	147	Type IIS restriction enzyme M1 protein (mod)
HP79_05859	147	Type IIS restriction enzyme M2 protein (mod)
HP79_05864	147	Adenine-specific DNA methyltransferase
HP79_05869	147	Adenine-specific DNA methyltransferase
HP79_06346	154	tRNA (guanine-N(1)-)-methyltransferase
HP79_06411	157	Adenine-specific DNA methyltransferase
HP79_06976	169	Adenine-specific DNA methyltransferase
HP79_00105	5	Ribosomal RNA large subunit methyltransferase N
HP79_00370	14	Methylated-DNA--protein-cysteine methyltransferase
HP79_00848	23	16S ribosomal RNA methyltransferase RsmE
HP79_00988	28	Type III DNA modification enzyme (methyltransferase)
HP79_00995	29	Type III R-M system modification enzyme
HP79_01020	29	Type IIS restriction enzyme R and M protein (ECO57IR)
HP79_01200	31	Ubiquinone/menaquinone biosynthesis methyltransferase
HP79_05511	137	N-6 Adenine-specific DNA methylase

Appendix 4 List of endonucleases in the genome of *H. pylori* P79

Locus tag	Contig	Product
HP79_07645	187	Mulitfunctional nucleoside diphosphate kinase/apyrimidinic endonuclease/3'-phosphodiesterase
HP79_02179	66	Endonuclease III
HP79_00025	102	Ulcer-associated gene restriction endonuclease (iceA)
HP79_00617	20	Holliday junction resolvase-like protein
HP79_04142	113	Type I restriction enzyme S protein (hsdS)
HP79_02129	64	Type III restriction enzyme R protein (res)
HP79_02054	61	3-methyladenine DNA glycosylase
HP79_03146	88	Anti-codon nuclease masking agent (prrB)
HP79_02956	82	Hypothetical protein
HP79_02804	78	Type I restriction enzyme S protein (hsdS)
HP79_07061	172	Similar to archaeal Holliday junction resolvase and Mrr protein
HP79_06089	150	Type I restriction enzyme S protein (hsdS)
HP79_01000	29	Type III restriction enzyme R protein (res)

Appendix 5 List of sequenced genomes used in core genome analysis of *H. pylori*

Strain	Size (Mbp)*	Location	Disease	Status
SNT49	1.61 (3.2kb p)	India	Asymptomatic	complete
26695	1.67	UK	Gastritis	complete
HPAG1	1.6 (0.01p)	Sweeden	Chronic atrophic gastritis	complete
Shi470	1.61	Peru	Gastritis	complete
Puno120	1.62 (0.01 p)	Peru	Gastritis	complete
Puno135	1.65	Peru	Gastritis	complete
F16	1.58	Japan	Gastritis	complete
v225d	1.59 (0.01 p)	Venezuela	Acute superficial gastritis	complete
SJM180	1.66	Peru	Gastritis	complete
B8	1.67 (0.01p)		Gastric ulcer	complete
B128	1.65		Gastric ulcer	73 contigs
J99	1.64	USA	Duodenal ulcer	complete
51	1.59	South Korea	Duodenal ulcer	complete
908	1.55	France	Duodenal ulcer	complete
2017	1.55	France	Duodenal ulcer	complete
2018	1.56	France	Duodenal ulcer	complete
F30	1.57 (0.01 p)	Japan	Duodenal ulcer	complete
P12	1.67 (0.01 p)	Germany	Duodenal ulcer	complete
PeCan4	1.63 (0.01p)	Peru	Gastric cancer	complete
ELS37	1.66 (0.01 p)	El Salvador	Gastric cancer	complete
F32	1.58 (2.6kb p)	Japan	Diffuse type gastric cancer	complete
F57	1.61	Japan	Diffuse type gastric cancer	complete
PeCan18	1.66	Peru	Gastric cancer	complete
XZ274	1.63 (0.02 p)	China	Gastric cancer	complete
98-10	1.57	Japan	Gastric cancer	51 contigs
B45	1.6 (0.02 phage)$^{+}$	France	MALT lymphoma	63 contigs
B38	1.58	France	MALT lymphoma	complete
G27	1.65 (0.01 p)	Italy	Endoscopy patient	complete

52	1.57	South Korea	complete
HUP-B14	1.6 (0.01 p)	Spain	complete
35A	1.57		complete
India7	1.68	India	complete
83	1.62		complete
Lithuania75	1.62 (0.02 p)	Lithuania	complete
Gambia94/24	1.71 (2.5kb p)	Gambia	complete
Sat464	1.56 (0.01 p)	Peru	complete
Shi417	1.67	Peru	complete
Cuz20	1.64	Peru	complete
SouthAfrica7	1.65 (0.03 p)	South Africa	complete
Shi112	1.66	Peru	complete
Shi169	1.62	Peru	complete

*Figures in brackets denote the presence and size (Mb unless otherwise stated) of plasmid DNA (p). [+]Strain harbours 0.02 Mb phage DNA.

Strains used in core genome analysis in addition to 17874 and P79 (not listed). Strain and genome information was collected from the National Center for Biotechnology Information (NCBI, 2013).

Appendix 6 Revised core genome of *H. pylori

Locus tag	Product
HP17_00015	Phosphoenolpyruvate synthase
HP17_00025	Threonyl-tRNA synthetase
HP17_00030	InfC translation initiation factor IF-3
HP17_00035	RpmI 50S ribosomal protein
HP17_00040	RplT 50S ribosomal protein L20
HP17_00045	Outer membrane protein
HP17_00060	Hypothetical protein
HP17_00070	L-serine deaminase
HP17_00075	L-serine transporter
HP17_00080	Phospho-2-dehydro-3-deoxyheptonate aldolase
HP17_00090	Bacterioferritin comigratory protein
HP17_00095	Hypothetical protein
HP17_00110	Fe-S oxidoreductase
HP17_00120	L-lactate permease
HP17_00140	Cbb3-type cytochrome c oxidase subunit I
HP17_00145	Cbb3-type cytochrome c oxidase subunit II
HP17_00150	Cytochrome c oxidase, cbb3-type, CcoQ subunit
HP17_00155	Cytochrome c oxidase, cbb3-type, subunit III
HP17_00160	Hypothetical protein
HP17_00170	Hypothetical protein
HP17_00185	Hypothetical protein
HP17_00200	Hypothetical protein
HP17_00205	Hypothetical protein
HP17_00210	AroK shikimate kinase
HP17_00215	Hypothetical protein
HP17_00225	Hypothetical protein
HP17_00235	Hypothetical protein
HP17_00240	Delta-aminolevulinic acid dehydratase
HP17_00255	Response regulator OmpR
HP17_00260	Hypothetical protein
HP17_00270	Collagenase
HP17_00275	Hypothetical protein
HP17_00285	PrfB peptide chain release factor 2
HP17_00295	Flagellar biosynthesis protein FliR
HP17_00300	Hypothetical protein
HP17_00315	Excinuclease ABC subunit B
HP17_00325	Adenylosuccinate lyase
HP17_00335	Pyruvate ferredoxin oxidoreductase, beta subunit
HP17_00340	PorA pyruvate flavodoxin oxidoreductase subunit alpha

HP17_00345	PorD pyruvate flavodoxin oxidoreductase subunit delta
HP17_00350	Pyruvate flavodoxin oxidoreductase subunit gamma
HP17_00355	Outer membrane protein HorH; signal peptide
HP17_00380	Glk glucokinase
HP17_00385	6-phosphogluconolactonase
HP17_00390	Glucose-6-phosphate 1-dehydrogenase
HP17_00395	Phosphogluconate dehydratase
HP17_00400	Multifunctional KHG/KDPG aldolase
HP17_00405	Putative beta-lactamase HcpC
HP17_00415	Putative outer membrane protein
HP17_00425	UDP-glucose 4-epimerase
HP17_00430	TruA tRNA pseudouridine synthase A
HP17_00435	Hypothetical protein
HP17_00440	Pcm protein-L-isoaspartate O-methyltransferase
HP17_00445	NrdF ribonucleotide-diphosphate reductase subunit beta
HP17_00475	Biotin carboxylase
HP17_00496	Dcd deoxycytidine triphosphate deaminase
HP17_00506	16S ribosomal RNA methyltransferase RsmE
HP17_00511	Hypothetical protein
HP17_00521	Thiol:disulfide interchange protein
HP17_00541	Hydrogenase expression/formation protein
HP17_00551	Hypothetical protein
HP17_00581	Symporter, SSS family (Proline Permease); membrane protein
HP17_00586	Delta-1-pyrroline-5-carboxylate dehydrogenase
HP17_00591	Hypothetical protein
HP17_00676	Urease accessory protein UreH
HP17_00681	Urease accessory protein UreG
HP17_00701	Urease accessory protein UreI
HP17_00721	Urease subunit alpha
HP17_00726	LspA lipoprotein signal peptidase
HP17_00731	GlmM phosphoglucosamine mutase
HP17_00736	RpsT 30S ribosomal protein S20
HP17_00741	PrfA peptide chain release factor 1
HP17_00751	Hypothetical protein
HP17_00761	RpsI 30S ribosomal protein S9
HP17_00766	RplM 50S ribosomal protein L13
HP17_00771	Hypothetical protein
HP17_00776	Hypothetical protein
HP17_00781	Hypothetical protein
HP17_00786	RNA polymerase sigma factor RpoD
HP17_00791	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase
HP17_00796	Malonyl CoA-acyl carrier protein transacylase

HP17_00830	Hypothetical protein
HP17_00835	2-hydroxyacid dehydrogenase
HP17_00840	Hypothetical protein
HP17_00855	Hypothetical protein
HP17_00865	Hypothetical protein
HP17_00875	2',3'-cyclic-nucleotide 2'-phosphodiesterase
HP17_00880	S-ribosylhomocysteinase
HP17_00885	Cystathionine gamma-synthase/cystathionine beta- lyase
HP17_00905	DnaK molecular chaperone DnaK
HP17_00910	Heat shock protein GrpE
HP17_00920	Hypothetical protein
HP17_00925	Hypothetical protein
HP17_00930	Hypothetical protein
HP17_00935	Flagellin B
HP17_00950	Hypothetical protein
HP17_00955	4-oxalocrotonate tautomerase
HP17_00960	RecR recombination protein RecR
HP17_00970	Heat shock protein HtpX
HP17_00975	FolE GTP cyclohydrolase I
HP17_00980	Geranyltranstransferase (Farnesyl-diphosphate synthase) (FPP synthase)
HP17_00985	SurE 5'(3')-nucleotidase/polyphosphatase
HP17_00990	Hypothetical protein
HP17_00995	6-carboxy-5,6,7,8-tetrahydropterin synthase
HP17_01000	Hypothetical protein
HP17_01005	Hypothetical protein
HP17_01025	Amino acid ABC transporter permease
HP17_01030	Putative polar amino acid transport system substrate-binding protein
HP17_01035	Alanine racemase
HP17_01045	D-alanine glycine permease
HP17_01050	D-amino acid dehydrogenase
HP17_01055	Hypothetical protein
HP17_01128	Nickel cobalt outer membrane efflux protein
HP17_01133	GlyS glycyl-tRNA synthetase subunit beta
HP17_01143	Phosphoglyceromutase
HP17_01148	GatC aspartyl/glutamyl-tRNA amidotransferase subunit C
HP17_01153	Adenosylmethionine--8-amino-7-oxononanoate transaminase
HP17_01158	Peptidyl-prolyl cis-trans isomerase D
HP17_01163	Cell division protein FtsA
HP17_01168	Cell division protein FtsZ
HP17_01238	DNA polymerase III subunit delta
HP17_01243	Ribonuclease R
HP17_01268	Tryptophanyl-tRNA synthetase

HP17_01273	Biotin biosynthesis protein BioC
HP17_01278	SecG preprotein translocase subunit SecG
HP17_01283	Frr ribosome recycling factor
HP17_01288	PyrE orotate phosphoribosyltransferase
HP17_01293	Conserved hypothetical mitochondrial protein- like protein 4
HP17_01303	NADH dehydrogenase subunit A
HP17_01308	NADH dehydrogenase subunit B
HP17_01313	NADH dehydrogenase subunit C
HP17_01323	Putative NADH oxidoreductase I
HP17_01338	NADH:ubiquinone oxidoreductase subunit H
HP17_01343	NADH dehydrogenase subunit I
HP17_01348	NADH:ubiquinone oxidoreductase subunit J
HP17_01353	NADH:ubiquinone oxidoreductase subunit K
HP17_01358	NADH:ubiquinone oxidoreductase subunit L
HP17_01368	NADH:ubiquinone oxidoreductase subunit N
HP17_01398	TrpA tryptophan synthase subunit alpha
HP17_01403	Tryptophan synthase subunit beta
HP17_01413	TrpD anthranilate phosphoribosyltransferase
HP17_01418	Anthranilate synthase component II
HP17_01423	Anthranilate synthase component I
HP17_01463	Hypothetical protein
HP17_01468	Hypothetical protein
HP17_01478	Transcriptional regulator (tenA)
HP17_01498	Nicotinamide mononucleotide transporter
HP17_01513	RplQ 50S ribosomal protein L17
HP17_01518	DNA-directed RNA polymerase subunit alpha
HP17_01523	RpsD 30S ribosomal protein S4
HP17_01528	30S ribosomal protein S11
HP17_01538	RpsM 30S ribosomal protein S13
HP17_01548	InfA translation initiation factor IF-1
HP17_01553	Methionine aminopeptidase
HP17_01558	SecY preprotein translocase subunit SecY
HP17_01563	RplO 50S ribosomal protein L15
HP17_01568	RpsE 30S ribosomal protein S5
HP17_01573	RplR 50S ribosomal protein L18
HP17_01578	RplF 50S ribosomal protein L6
HP17_01583	RpsH 30S ribosomal protein S8
HP17_01593	RplE 50S ribosomal protein L5
HP17_01603	RplN 50S ribosomal protein L14
HP17_01608	RpsQ 30S ribosomal protein S17
HP17_01613	50S ribosomal protein L29
HP17_01618	RplP 50S ribosomal protein L16

HP17_01623	RpsC 30S ribosomal protein S3
HP17_01628	RplV 50S ribosomal protein L22
HP17_01633	RpsS 30S ribosomal protein S19
HP17_01643	RplB 50S ribosomal protein L2
HP17_01648	RplW 50S ribosomal protein L23
HP17_01653	RplD 50S ribosomal protein L4
HP17_01658	RplC 50S ribosomal protein L3
HP17_01663	RpsJ 30S ribosomal protein S10
HP17_01673	Hypothetical protein
HP17_01683	RnhB ribonuclease HII
HP17_01698	FumC fumarate hydratase
HP17_01703	Hypothetical protein
HP17_01708	Hypothetical protein
HP17_01713	Putative cobalt-zinc-cadmium resistance protein CzcB
HP17_01723	Hypothetical protein
HP17_01728	Branched-chain amino acid transport protein
HP17_01733	Chaperone protein DnaJ
HP17_01738	Hypothetical protein
HP17_01748	MnmA tRNA-specific 2-thiouridylase MnmA
HP17_01753	Hypothetical protein
HP17_01763	Nickel responsive regulator
HP17_01793	Putative heme iron utilization protein
HP17_01803	ArgS arginyl-tRNA synthetase
HP17_01813	Gmk guanylate kinase
HP17_01828	Nuclease NucT
HP17_01833	Outer membrane protein HorC; signal peptide
HP17_01838	FlgH flagellar basal body L-ring protein
HP17_01858	LpxK tetraacyldisaccharide 4'-kinase
HP17_01863	NAD synthetase
HP17_01868	Ketol-acid reductoisomerase
HP17_01873	Cell division inhibitor
HP17_01878	MinE cell division topological specificity factor MinE
HP17_01883	Hypothetical protein
HP17_01888	Holliday junction resolvase-like protein
HP17_01910	Hypothetical protein
HP17_01960	HP17_01960 single-stranded-DNA-specific exonuclease
HP17_01965	PyrG CTP synthetase
HP17_01970	HP17_01970 hypothetical protein
HP17_01980	FliG flagellar motor switch protein G
HP17_01985	FliH flagellar assembly protein H
HP17_01990	1-deoxy-D-xylulose-5-phosphate synthase
HP17_01995	GTP-binding protein LepA

HP17_02025	Flagellar basal-body rod protein
HP17_02030	General substrate transporter, MFS superfamily
HP17_02050	Transketolase
HP17_02060	Bifunctional riboflavin kinase/FMN adenylyltransferase
HP17_02070	Hypothetical protein
HP17_02075	PyrB aspartate carbamoyltransferase catalytic subunit
HP17_02095	Similar to archaeal Holliday junction resolvase
HP17_02100	High-affinity nickel-transport protein
HP17_02105	Hypothetical protein
HP17_02130	CDP-diacylglycerol--serine O- phosphatidyltransferase
HP17_02145	Cell division protein FtsH
HP17_02150	PrmA ribosomal protein L11 methyltransferase
HP17_02160	Outer membrane protein HorD
HP17_02165	Hypothetical protein
HP17_02175	GidB 16S rRNA methyltransferase GidB
HP17_02180	QueA S-adenosylmethionine:tRNA ribosyltransferase- isomerase
HP17_02185	Sec-independent protein translocase protein tat
HP17_02190	Sec-independent translocase
HP17_02195	RuvB Holliday junction DNA helicase RuvB
HP17_02200	PanB 3-methyl-2-oxobutanoate hydroxymethyltransferase
HP17_02210	Hypothetical protein
HP17_02215	Hypothetical protein
HP17_02225	MinC septum formation inhibitor
HP17_02230	LpxC UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase
HP17_02235	Hypothetical protein
HP17_02240	Homoserine kinase
HP17_02262	RbfA ribosome-binding factor A
HP17_02267	Ribosome maturation factor rimP
HP17_02277	Phosphodiesterase domain-containing protein
HP17_02282	Transcriptional regulator
HP17_02292	Hypothetical protein
HP17_02297	FlhA flagellar biosynthesis protein FlhA
HP17_02302	RpsO 30S ribosomal protein S1
HP17_02312	3-dehydroquinate dehydratase
HP17_02317	X-Pro aminopeptidase
HP17_02322	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase
HP17_02327	FlhF flagellar biosynthesis regulator FlhF
HP17_02332	Flagellar biosynthesis protein FlhG
HP17_02342	FliA flagellar biosynthesis sigma factor
HP17_02347	FliM flagellar motor switch protein FliM
HP17_02352	Flagellar motor switch protein FliY
HP17_02357	Hypothetical protein

HP17_02362	Hypothetical protein
HP17_02367	Ferric uptake regulation protein
HP17_02372	Recombination factor protein RarA
HP17_02377	Putative transcriptional regulator, MerR family protein
HP17_02397	Response regulator
HP17_02402	IspDF bifunctional 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase/2-C-methyl-D-erythritol 2,4- cyclodiphosphate synthase protein
HP17_02407	Protease DO
HP17_02422	Phosphatidylglycerophosphate synthase
HP17_02427	Hypothetical protein
HP17_02432	7-alpha-hydroxysteroid dehydrogenase
HP17_02437	Dihydrodipicolinate synthase
HP17_02447	Putative zinc protease
HP17_02452	Dihydroorotate dehydrogenase
HP17_02457	Polyphosphate kinase
HP17_02484	Glyceraldehyde-3-phosphate dehydrogenase
HP17_02489	Integral membrane protein
HP17_02499	Hypothetical protein
HP17_02524	Outer membrane protein HopB
HP17_02529	Outer membrane porin and adhesin HopC; signal peptide
HP17_02564	FlgD flagellar basal body rod modification protein
HP17_02594	Hypothetical protein
HP17_02638	Hypothetical protein
HP17_02648	Maf-like protein
HP17_02668	Carbamoyl phosphate synthase small subunit
HP17_02673	Hypothetical protein
HP17_02678	Integral membrane protein
HP17_02683	Membrane transport protein
HP17_02688	Hypothetical protein
HP17_02703	Hypothetical protein
HP17_02708	Aspartate kinase
HP17_02713	RNA pyrophosphohydrolase
HP17_02723	Cytochrome c-553
HP17_02728	Coproporphyrinogen III oxidase
HP17_02733	Camphor resistance protein CrcB
HP17_02738	HemD uroporphyrinogen-III synthase
HP17_02743	Hypothetical protein
HP17_02758	Undecaprenyl pyrophosphate synthase
HP17_02768	Glycinamide ribonucleotide synthetase
HP17_02773	Hypothetical protein
HP17_02778	Organic solvent tolerance protein
HP17_02783	Phosphoribosyltransferase

HP17_02788	Polynucleotide phosphorylase/polyadenylase
HP17_02793	F0F1 ATP synthase subunit C
HP17_02798	Serine acetyltransferase
HP17_02808	Site-specific DNA-methyltransferase
HP17_02813	Hypothetical protein
HP17_02818	ABC transporter ATP-binding protein
HP17_02823	Elongation factor Tu
HP17_02833	SecE preprotein translocase subunit SecE
HP17_02838	nusG transcription antitermination protein NusG
HP17_02843	rplK 50S ribosomal protein L11
HP17_02848	rplA 50S ribosomal protein L1
HP17_02853	rplJ 50S ribosomal protein L10
HP17_02858	rplL 50S ribosomal protein L7/L12
HP17_02887	rpsL 30S ribosomal protein S12
HP17_02892	HP17_02892 30S ribosomal protein S7
HP17_02902	HP17_02902 elongation factor G
HP17_02932	HP17_02932 ADP-heptose--LPS heptosyltransferase II
HP17_02947	HP17_02947 aspartate-semialdehyde dehydrogenase
HP17_02967	GltX glutamyl-tRNA synthetase
HP17_02992	Hypothetical protein
HP17_03002	Oligoendopeptidase F
HP17_03007	Hypothetical protein
HP17_03012	Hypothetical protein
HP17_03022	Hypothetical protein
HP17_03027	Hypothetical protein
HP17_03042	Alpha-carbonic anhydrase
HP17_03047	Putative arabinose transporter
HP17_03052	Hypothetical protein
HP17_03057	Na+/H+ antiporter
HP17_03067	Putative PP-loop family ATPase
HP17_03072	Multidrug-efflux transporter
HP17_03077	Nucleoside transporter
HP17_03087	Purine-nucleoside phosphorylase
HP17_03114	Hypothetical protein
HP17_03129	tRNA-dihydrouridine synthase B
HP17_03134	Hypothetical protein
HP17_03204	Integral membrane protein
HP17_03214	ABC transporter, ATP-binding protein
HP17_03229	Phosphoheptose isomerase
HP17_03239	ADP-heptose synthase
HP17_03244	ADP-L-glycero-D-manno-heptose-6-epimerase
HP17_03249	D,D-heptose 1,7-bisphosphate phosphatase

HP17_03254	Hypothetical protein
HP17_03259	Pantothenate kinase
HP17_03269	Hypothetical protein
HP17_03274	Dut deoxyuridine 5'-triphosphate nucleotidohydrolase
HP17_03279	GreA transcription elongation factor GreA
HP17_03284	Ipid-A-disaccharide synthase
HP17_03289	Hypothetical protein
HP17_03294	HypA hydrogenase nickel incorporation protein
HP17_03319	CDP-diacylglycerol pyrophosphatase
HP17_03324	Alkylphosphonate uptake protein
HP17_03329	Hypothetical protein
HP17_03334	Hypothetical protein
HP17_03339	Catalase
HP17_03344	Iron-regulated outer membrane protein
HP17_03354	RuvC Holliday junction resolvase
HP17_03384	RuvA Holliday junction DNA helicase RuvA
HP17_03389	Hypothetical protein
HP17_03399	CysS cysteinyl-tRNA synthetase
HP17_03409	Iron(III) dicitrate transport system ATP-binding protein
HP17_03414	Iron(III) dicitrate ABC transporter permease protein
HP17_03419	Short-chain oxidoreductase
HP17_03424	Hypothetical protein
HP17_03454	Cysteine-rich protein X
HP17_03464	Gamma-glutamyltranspeptidase
HP17_03469	FlgK flagellar hook-associated protein FlgK
HP17_03474	Hypothetical protein
HP17_03479	M. HpyAVIII, a type II cytosine specific DNA methyltransferase
HP17_03484	Hypothetical protein
HP17_03494	FKBP-type peptidyl-prolyl cis-trans isomerase slyD
HP17_03499	Hypothetical protein
HP17_03504	Peptidoglycan-associated lipoprotein precursor
HP17_03509	TolB translocation protein TolB
HP17_03519	Biopolymer transport protein ExbD/TolR
HP17_03524	Biopolymer transport protein
HP17_03529	AtpC F0F1 ATP synthase subunit epsilon
HP17_03534	F0F1 ATP synthase subunit beta
HP17_03539	F0F1 ATP synthase subunit gamma
HP17_03544	F0F1 ATP synthase subunit alpha
HP17_03549	F0F1 ATP synthase subunit delta
HP17_03554	F0F1 ATP synthase subunit B
HP17_03559	F0F1 ATP synthase subunit B'
HP17_03564	Plasmid replication-partition-like protein

HP17_03579	Biotin--protein ligase
HP17_03624	Hypothetical protein
HP17_03629	Peptidyl-prolyl cis-trans isomerase B, cyclosporine-type rotamase (ppi)
HP17_03634	Carbon storage regulator
HP17_03639	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase
HP17_03644	SmpB SsrA-binding protein
HP17_03669	Hypothetical protein
HP17_03674	HP17_03674 membrane protein insertase
HP17_03684	TrmE tRNA modification GTPase TrmE
HP17_03709	Hypothetical protein
HP17_03714	Hypothetical protein
HP17_03719	Membrane-associated lipoprotein
HP17_03724	Collagen-binding surface adhesin SpaP
HP17_03729	Thioredoxin
HP17_03734	Ribosomal large subunit pseudouridine synthase
HP17_03744	Cytochrome c551 peroxidase
HP17_03759	Hypothetical protein
HP17_03764	ABC transport system substrate binding protein
HP17_03784	Outer membrane protein; signal peptide
HP17_03794	Branched-chain amino acid aminotransferase
HP17_03799	Outer membrane protein HorJ; signal peptide
HP17_03804	DNA polymerase I
HP17_03819	Hypothetical protein
HP17_03824	Tmk thymidylate kinase
HP17_03829	CoaD phosphopantetheine adenylyltransferase
HP17_03834	3-octaprenyl-4-hydroxybenzoate carboxy-lyase
HP17_03844	FlgA flagellar basal body P-ring biosynthesis protein FlgA
HP17_03849	DNA helicase II (UvrD)
HP17_03854	hypothetical protein
HP17_03859	seryl-tRNA synthetase
HP17_03864	Hypothetical protein
HP17_03869	Exodeoxyribonuclease VII small subunit
HP17_03874	UbiE ubiquinone/menaquinone biosynthesis methyltransferase
HP17_03879	Hypothetical protein
HP17_03884	X-Pro dipeptidase
HP17_03889	Antibiotic transport system permease protein
HP17_03894	Hypothetical protein
HP17_03899	Hypothetical protein
HP17_03904	Lipase-like protein
HP17_03909	Hemolysin domain-containing protein
HP17_03924	Putative nifU-like protein
HP17_03929	Hypothetical protein

HP17_03934	MurE UDP-N-acetylmuramoylalanyl-D-glutamate--2, 6-diaminopimelate ligase
HP17_03939	Transaldolase
HP17_03944	50S ribosomal protein L25/general stress protein Ctc
HP17_03949	Peptidyl-tRNA hydrolase
HP17_03954	Permease; membrane protein
HP17_03999	Hypothetical protein
HP17_04004	Riboflavin biosynthesis protein
HP17_04009	Sodium/glutamate symport carrier protein/glutamate permease
HP17_04014	Saccharopine dehydrogenase
HP17_04024	Putative glycerol-3-phosphate acyltransferase PlsY
HP17_04029	Hypothetical protein
HP17_04034	FrpB-like protein
HP17_04049	Selenocysteine synthase
HP17_04054	NusA transcription elongation factor NusA
HP17_04109	ATP-dependent DNA helicase RecG
HP17_04114	Hypothetical protein
HP17_04119	Hypothetical protein
HP17_04124	Exodeoxyribonuclease III
HP17_04129	Hypothetical protein
HP17_04149	Hypothetical protein
HP17_04154	Glucosamine--fructose-6-phosphate aminotransferase
HP17_04184	Hypothetical protein
HP17_04199	Arginase
HP17_04204	Alanine dehydrogenase
HP17_04226	Outer membrane protein
HP17_04251	Hypothetical protein
HP17_04256	Hypothetical protein
HP17_04276	Fructose-1,6-bisphosphatase
HP17_04281	Hypothetical protein
HP17_04296	Ubiquinol cytochrome c oxidoreductase, cytochrome c1 subunit
HP17_04301	Ubiquinol-cytochrome c reductase cytochrome b subunit
HP17_04306	Ubiquinol-cytochrome c reductase, iron-sulfur subunit
HP17_04311	Transcription-repair coupling factor
HP17_04326	Hypothetical protein
HP17_04331	Folylpolyglutamate synthase
HP17_04336	Hypothetical protein
HP17_04341	LeuS leucyl-tRNA synthetase
HP17_04346	Integral membrane protein
HP17_04351	SecF preprotein translocase subunit SecF
HP17_04356	SecD preprotein translocase subunit SecD
HP17_04371	YajC preprotein translocase subunit YajC
HP17_04376	NhaA pH-dependent sodium/proton antiporter

HP17_04381	Putative recombination protein RecB
HP17_04386	RpsB 30S ribosomal protein S2
HP17_04391	Tsf elongation factor Ts
HP17_04396	HP17_04396 cell division protein
HP17_04401	FliE flagellar hook-basal body protein FliE
HP17_04406	FlgC flagellar basal body rod protein FlgC
HP17_04411	FlgB flagellar basal body rod protein FlgB
HP17_04431	Alkyl hydroperoxide reductase
HP17_04436	Outer membrane protein
HP17_04441	Penicillin-binding protein 2
HP17_04446	Hypothetical protein
HP17_04451	EngB GTP-binding protein YsxC
HP17_04456	Hypothetical protein
HP17_04461	Hypothetical protein
HP17_04466	3-deoxy-D-manno-octulosonate 8-phosphate phosphatase
HP17_04471	Rare lipoprotein A
HP17_04476	Regulatory protein DniR
HP17_04481	DNAse
HP17_04486	Riboflavin synthase subunit alpha
HP17_04511	Hypothetical protein
HP17_04536	Pyridoxine 5'-phosphate synthase
HP17_04541	PdxA 4-hydroxythreonine-4-phosphate dehydrogenase
HP17_04556	FlgG flagellar basal body rod protein FlgG
HP17_04581	Hypothetical protein
HP17_04596	NusB transcription antitermination protein NusB
HP17_04606	2-dehydro-3-deoxyphosphooctonate aldolase
HP17_04616	Orotidine 5'-phosphate decarboxylase
HP17_04621	PanC pantoate--beta-alanine ligase
HP17_04646	GroEL chaperonin GroEL
HP17_04651	GroES co-chaperonin GroES
HP17_04656	DnaG DNA primase
HP17_04666	Hypothetical protein
HP17_04671	Hypothetical protein
HP17_04676	Hypothetical protein
HP17_04681	ATPase/DNA transfer protein
HP17_04691	Chemotaxis protein
HP17_04696	Carboxynorspermidine decarboxylase
HP17_04701	Lipid A 1-phosphatase
HP17_04706	Lipid A phosphoethanolamine transferase
HP17_04731	Isocitrate dehydrogenase
HP17_04741	Dethiobiotin synthetase
HP17_04751	Putative universal stress global response regulator UspA

HP17_04756	ATP-dependent Clp protease adapter protein ClpS
HP17_04761	ATP-dependent C1p protease (clpA)
HP17_04776	Hypothetical protein
HP17_04791	Hypothetical protein
HP17_04796	ComB9 competence protein
HP17_04801	ComB10 competence protein
HP17_04806	Mannose-1-phosphate guanyltransferase
HP17_04811	GDP-D-mannose dehydratase
HP17_04816	Nodulation protein (nolK)
HP17_04831	3-deoxy-manno-octulosonate cytidylyltransferase
HP17_04836	Disulphide isomerase
HP17_04841	Hypothetical protein
HP17_04851	Hypothetical protein
HP17_04856	Cysteine-rich protein E; beta-lactamase HcpE precursor
HP17_04861	Hypothetical protein
HP17_04866	HemC porphobilinogen deaminase
HP17_04871	Prolyl-tRNA synthetase
HP17_04876	HemA glutamyl-tRNA reductase
HP17_04881	Octaprenyl-diphosphate synthase (Octaprenyl pyrophosphate synthetase) (OPP synthetase)
HP17_04886	Hypothetical protein
HP17_04891	Hypothetical protein
HP17_04896	HP17_04896 Neutrophil activating protein NapA (bacterioferritin)
HP17_04901	Histidine kinase sensor protein
HP17_04906	Hypothetical protein
HP17_04911	FlgI flagellar basal body P-ring protein
HP17_04916	ATP-dependent RNA helicase
HP17_04933	Hypothetical protein
HP17_04938	Oligopeptide permease
HP17_04965	Hypothetical protein
HP17_04970	Adenylosuccinate synthetase
HP17_04975	Hypothetical protein
HP17_04985	Hypothetical protein
HP17_04990	XseA exodeoxyribonuclease VII large subunit
HP17_05015	ATP-dependent protease binding subunit/heat shock protein
HP17_05025	Dihydroorotase
HP17_05030	Chlorohydrolase
HP17_05035	Hypothetical protein
HP17_05040	(dimethylallyl)adenosine tRNA methylthiotransferase
HP17_05045	Hypothetical protein
HP17_05050	Hypothetical protein
HP17_05055	Hypothetical protein
HP17_05060	Hypothetical protein

HP17_05065	Hypothetical protein
HP17_05070	ATP-dependent nuclease
HP17_05075	Hypothetical protein
HP17_05080	Putative 4Fe-4S ferredoxin-type protein
HP17_05085	Guanosine pentaphosphate phosphohydrolase
HP17_05090	Lipopolysaccharide heptosyltransferase-1
HP17_05095	Lipid A biosynthesis lauroyl acyltransferase
HP17_05100	Tgt queuine tRNA-ribosyltransferase
HP17_05115	AroB 3-dehydroquinate synthase
HP17_05120	hypothetical protein
HP17_05125	hypothetical protein
HP17_05130	Cell division protein FtsH; signal peptide
HP17_05135	Hypothetical protein
HP17_05140	Hypothetical protein
HP17_05165	Chorismate mutase
HP17_05180	AmiE acylamide amidohydrolase
HP17_05185	FlgL flagellar hook-associated protein FlgL
HP17_05190	RplU 50S ribosomal protein L21
HP17_05195	RpmA 50S ribosomal protein L27
HP17_05200	Peptide ABC transporter substrate-binding protein
HP17_05205	Peptide ABC transporter permease
HP17_05230	ObgE GTPase CgtA
HP17_05240	Hypothetical protein
HP17_05245	Glutamate-1-semialdehyde aminotransferase
HP17_05250	Hypothetical protein
HP17_05255	Hypothetical protein
HP17_05260	Putative N-carbamoyl-D-amino acid amidohydrolase
HP17_05285	Hypothetical protein
HP17_05290	Conserved ATP/GTP binding protein
HP17_05315	Hypothetical protein
HP17_05320	AspA aspartate ammonia-lyase
HP17_05325	UDP-N-acetylglucosamine 1- carboxyvinyltransferase
HP17_05330	Hypothetical protein
HP17_05335	UTP--glucose-1-phosphate uridylyltransferase subunit
HP17_05340	Soluble lytic murein transglycosylase
HP17_05350	Glutamylglutaminyl-tRNA synthetase
HP17_05370	Polynucleotide adenylyltransferase; poly(A) polymerase
HP17_05390	Hypothetical protein
HP17_05400	Hypothetical protein
HP17_05405	Quinone-reactive Ni/Fe hydrogenase (hydD)
HP17_05410	Ni/Fe-hydrogenase, b-type cytochrome subunit
HP17_05415	Nickel-dependent hydrogenase, large subunit

HP17_05425	Hydrogenase (NiFe) small subunit HydA
HP17_05445	Cysteine-rich protein F 977146:978204 reverse
HP17_05450	Tetrahydrodipicolinate N-succinyltransferase
HP17_05455	IspG 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase
HP17_05460	Hypothetical protein
HP17_05465	MurC UDP-N-acetylmuramate--L-alanine ligase
HP17_05470	Hypothetical protein
HP17_05480	Inorganic pyrophosphatase
HP17_05495	Adk adenylate kinase
HP17_05500	AspS aspartyl-tRNA synthetase
HP17_05505	Chemotaxis protein (cheV)
HP17_05510	LigA NAD-dependent DNA ligase LigA
HP17_05515	Hypothetical protein
HP17_05550	Hypothetical protein
HP17_05560	Membrane fusion protein (mtrC)
HP17_05565	Outer-membrane protein of the hefABC efflux system
HP17_05570	HemE uroporphyrinogen decarboxylase
HP17_05590	3-methyladenine DNA glycosylase
HP17_05595	Flagellin A 1013546:1015003 reverse
HP17_05610	Methyl-accepting chemotaxis transmembrane sensory protein
HP17_05615	8-amino-7-oxononanoate synthase
HP17_05625	Tumor necrosis factor alpha-inducing protein
HP17_05635	Dsbb-like protein
HP17_05640	Hypothetical protein
HP17_05660	OorB 2-oxoglutarate-acceptor oxidoreductase subunit OorB
HP17_05670	OorD 2-oxoglutarate-acceptor oxidoreductase subunit OorD
HP17_05675	Aminodeoxychorismate lyase (pabC)
HP17_05680	Hypothetical protein
HP17_05690	Endonuclease III
HP17_05695	Flagellar motor switch protein
HP17_05715	Dihydroorotase
HP17_05720	Hypothetical protein
HP17_05725	Hypothetical protein
HP17_05735	Bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase
HP17_05740	Signal peptidase I (lepB)
HP17_05745	Hypothetical protein
HP17_05750	Ribose-5-phosphate isomerase B
HP17_05755	Hypothetical protein
HP17_05760	Adenine phosphoribosyltransferase
HP17_05765	Hypothetical protein
HP17_05770	Multifunctional aminopeptidase A
HP17_05775	GTP-binding protein YchF

HP17_05780	Hypothetical protein
HP17_05790	DapF diaminopimelate epimerase
HP17_05795	Hypothetical protein
HP17_05800	Hypothetical protein
HP17_05805	Hypothetical protein
HP17_05810	RpsU 30S ribosomal protein S21
HP17_05815	FabG 3-ketoacyl-(acyl-carrier-protein) reductase
HP17_05820	AcpP acyl carrier protein
HP17_05830	3-oxoacyl-(acyl carrier protein) synthase II
HP17_05865	RNA methyltransferase
HP17_05875	RpmE 50S ribosomal protein L31
HP17_05880	Rho transcription termination factor Rho
HP17_05885	Glutamate racemase
HP17_05990	Hypothetical protein
HP17_05995	Bifunctional phosphopantothenoylcysteine decarboxylase
HP17_06010	Hypothetical protein
HP17_06015	Hypothetical protein
HP17_06025	DNA-binding protein HU
HP17_06035	Hypothetical protein
HP17_06040	SpeE spermidine synthase
HP17_06045	CoaE dephospho-CoA kinase
HP17_06057	GatA aspartyl/glutamyl-tRNA amidotransferase subunit A
HP17_06062	Inosine 5'-monophosphate dehydrogenase
HP17_06067	F0F1 ATP synthase subunit A
HP17_06152	Phosphoserine phosphatase
HP17_06157	Ferritin
HP17_06162	Hypothetical protein
HP17_06177	Hypothetical protein
HP17_06182	Processing zinc-metalloprotease
HP17_06187	GatB aspartyl/glutamyl-tRNA amidotransferase subunit B
HP17_06207	RnhA ribonuclease H
HP17_06212	Rnc ribonuclease III
HP17_06217	Chorismate synthase
HP17_06222	Hypothetical protein
HP17_06227	Coproporphyrinogen III oxidase
HP17_06232	Glycerol-3-phosphate dehydrogenase
HP17_06247	Outer membrane protein
HP17_06252	Aspartate aminotransferase
HP17_06262	Phage integrase family site specific recombinase
HP17_06267	Methylated-DNA--protein-cysteine methyltransferase
HP17_06282	Oxidoreductase
HP17_06287	Ribonucleotide-diphosphate reductase subunit

HP17_06312	GlmU bifunctional N-acetylglucosamine-1-phosphate uridyltransferase/glucosamine-1-phosphate acetyltransferase
HP17_06337	Hypothetical protein
HP17_06357	Acetyl-CoA acetyltransferase
HP17_06362	Succinyl-CoA-transferase subunit A
HP17_06387	Outer membrane protein
HP17_06407	Acetone carboxylase subunit alpha
HP17_06412	Hypothetical protein
HP17_06417	Hypothetical protein
HP17_06422	Diacylglycerol kinase
HP17_06437	DNA gyrase subunit A
HP17_06442	Hypothetical protein
HP17_06472	Outer membrane protein HopE
HP17_06482	16S rRNA m(4)C1402 methyltranserfase
HP17_06487	Hypothetical protein
HP17_06492	Hypothetical protein
HP17_06507	Hypothetical protein
HP17_06522	RNA polymerase factor sigma-54 1186162:1187406
HP17_06527	ABC-type transport system, ATP binding protein
HP17_06562	Hypothetical protein
HP17_06597	LysS lysyl-tRNA synthetase
HP17_06602	Serine hydroxymethyltransferase
HP17_06607	Hypothetical protein
HP17_06612	Hypothetical protein
HP17_06627	Hypothetical protein
HP17_06637	FrdB fumarate reductase iron-sulfur subunit
HP17_06642	Fumarate reductase flavoprotein subunit
HP17_06647	Fumarate reductase cytochrome b-556 subunit
HP17_06652	TpiA triosephosphate isomerase
HP17_06657	Enoyl-(acyl carrier protein) reductase
HP17_06662	LpxD UDP-3-O-[3-hydroxymyristoyl] glucosamine N- acyltransferase
HP17_06667	S-adenosylmethionine synthetase
HP17_06677	Ndk mulitfunctional nucleoside diphosphate kinase/apyrimidinic endonuclease/3'-phosphodiesterase
HP17_06682	Hypothetical protein
HP17_06697	3-oxoacyl-(acyl carrier protein) synthase III
HP17_06702	Hypothetical protein
HP17_06707	Hypothetical protein
HP17_06722	ATP-binding protein (mpr)
HP17_06737	Outer membrane protein
HP17_06742	Heat shock protein 90
HP17_06747	Cysteine-rich protein A
HP17_06752	Succinyl-diaminopimelate desuccinylase

HP17_06762	Sodium-dependent transporter (huNaDC-1)
HP17_06772	1-deoxy-D-xylulose 5-phosphate reductoisomerase
HP17_06782	Hypothetical protein
HP17_06787	Hypothetical protein
HP17_06792	Cysteine desulfurase
HP17_06797	NifU-like protein
HP17_06812	DNA repair protein RadA
HP17_06817	Bifunctional methionine sulfoxide reductase A/B protein
HP17_06832	Hypothetical protein
HP17_06882	Prephenate dehydrogenase
HP17_06887	ATP-dependent protease L
HP17_06902	Flagellar assembly protein FliW
HP17_06907	FabZ (3R)-hydroxymyristoyl-ACP dehydratase
HP17_06912	UDP-N-acetylglucosamine acyltransferase
HP17_06917	ClpX ATP-dependent protease ATP-binding subunit ClpX
HP17_06922	Rod shape-determining protein MreB
HP17_06932	Rod shape-determining protein MreC
HP17_06967	Replicative DNA helicase
HP17_06977	UbiA prenyltransferase
HP17_06992	Hypothetical protein
HP17_06997	Phosphatidylserine decarboxylase
HP17_07002	Quinolinate synthetase
HP17_07007	Nicotinate-nucleotide pyrophosphorylase
HP17_07042	Carboxyl-terminal protease
HP17_07047	Hypothetical protein
HP17_07052	1-acyl-sn-glycerol-3-phosphate acyltransferase
HP17_07057	Uracil-DNA glycosylase
HP17_07067	Glyceraldehyde-3-phosphate dehydrogenase
HP17_07072	Pgk phosphoglycerate kinase
HP17_07077	Magnesium and cobalt transport protein
HP17_07082	Hypothetical protein
HP17_07129	Outer membrane protein HofD; signal peptide
HP17_07134	Outer membrane protein HofC
HP17_07169	GTP-binding protein TypA
HP17_07214	Arginine decarboxylase
HP17_07219	Polysaccharide biosynthesis protein
HP17_07229	Hypothetical protein
HP17_07239	Hypothetical protein
HP17_07249	Cyclopropane fatty acid synthase
HP17_07259	Hypothetical protein
HP17_07264	Putative neuraminyllactose-binding hemagglutinin- like protein
HP17_07274	GuaA GMP synthase

HP17_07279	Hypothetical protein
HP17_07289	Hypothetical protein
HP17_07294	Cysteine desulfurase
HP17_07304	PheS phenylalanyl-tRNA synthetase subunit alpha
HP17_07322	PheT phenylalanyl-tRNA synthetase subunit beta
HP17_07327	3-phosphoshikimate 1-carboxyvinyltransferase
HP17_07332	IspH 4-hydroxy-3-methylbut-2-enyl diphosphate reductase
HP17_07337	RpsA 30S ribosomal protein S1
HP17_07342	Hypothetical protein
HP17_07347	D-3-phosphoglycerate dehydrogenase
HP17_07352	3-octaprenyl-4-hydroxybenzoate carboxy-lyase
HP17_07357	Hypothetical protein
HP17_07367	CheA-MCP interaction modulator
HP17_07372	Autophosphorylating histidine kinase
HP17_07377	Purine-binding chemotaxis protein CheW
HP17_07382	Adhesin-thiol peroxidase
HP17_07392	Hypothetical protein
HP17_07402	Hypothetical protein
HP17_07407	Hypothetical protein
HP17_07412	Hypothetical protein
HP17_07417	Hypothetical protein
HP17_07432	N5-glutamine S-adenosyl-L-methionine-dependent methyltransferase
HP17_07447	Putative potassium channel protein
HP17_07457	Flagellar sheath adhesin
HP17_07462	MraY phospho-N-acetylmuramoyl-pentapeptide- transferase
HP17_07477	Hypothetical protein
HP17_07502	DNA polymerase III subunit beta
HP17_07507	GyrB DNA gyrase subunit B
HP17_07532	Hypothetical protein
HP17_07537	UDP-sugar diphosphatase
HP17_07557	Dihydrodipicolinate reductase
HP17_07572	Glutamine synthetase
HP17_07587	RplI 50S ribosomal protein L9
HP17_07592	ATP-dependent protease subunit HslV
HP17_07602	Era GTPase Era
HP17_07607	Hypothetical protein
HP17_07722	KsgA 16S ribosomal RNA methyltransferase KsgA/Dim1 family protein
HP17_07727	Ribonuclease J
HP17_07732	Polysialic acid capsule expression protein
HP17_07737	Ribosomal RNA large subunit methyltransferase N
HP17_07747	Hypothetical protein

HP17_07762	Type IV secretion system ATPase
HP17_07772	FliQ flagellar biosynthesis protein FliQ
HP17_07777	MurB UDP-N-acetylenolpyruvoylglucosamine reductase
HP17_07807	7-cyano-7-deazaguanine reductase
HP17_07842	Hypothetical protein
HP17_07847	Putative nucleotide phosphoribosyltransferase
HP17_07852	Putative aminotransferase
HP17_07857	Phosphatidyl glycerophosphatase A
HP17_07867	2-hydroxy-6-oxohepta-2,4-dienoate hydrolase
HP17_07872	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D- alanine ligase; putative membrane protein
HP17_07877	Hypothetical protein
HP17_07882	Ribose-phosphate pyrophosphokinase
HP17_07902	Hypothetical protein
HP17_07907	Hypothetical protein
HP17_07912	tRNA (guanine-N(7)-)-methyltransferase 1 2
HP17_07917	Cell division protein
HP17_07927	Hypothetical protein
HP17_07932	Flagellar protein FlaG
HP17_07937	FliD flagellar capping protein
HP17_07942	FliS flagellar protein FliS
HP17_07947	Hypothetical protein
HP17_07952	Molybdopterin biosynthesis protein
HP17_07962	Carbon-nitrogen hydrolase
HP17_07967	Hypothetical protein
HP17_07972	Hypothetical protein
HP17_07987	Hypothetical protein
HP17_07992	Signal recognition particle-docking protein FtsY
HP17_08049	2-nitropropane dioxygenase
HP17_08054	Tyrosyl-tRNA synthetase
HP17_08059	Guanosine-3', 5'-bis(diphosphate)3'- pyrophosphohydrolase /Guanosine-3',5'-Bis(diphosphate) synthetase II (ppGpp-3'-pyrophosphohydrolase/ppGpp synthetase II)
HP17_08064	RpoZ DNA-directed RNA polymerase subunit omega
HP17_08069	PyrH uridylate kinase
HP17_08074	Hypothetical protein
HP17_08104	Hypothetical protein
HP17_08109	LolA lipoprotein chaperone
HP17_08114	Preprotein translocase subunit SecA
HP17_08119	Hypothetical protein
HP17_08124	Hypothetical protein
HP17_08164	Glucose/galactose transporter
HP17_08169	Hypothetical protein

HP17_08174	Glutamine ABC transporter periplasmic glutamine- binding protein
HP17_08179	Phosphate ABC transporter ATP-binding protein
HP17_08184	Glutamine ABC transporter, permease protein
HP17_08189	Glutamine ABC transporter permease
HP17_08199	Outer membrane protein HofH
HP17_08219	Thioredoxin reductase
HP17_08224	Cation transport subunit for cbb3-type oxidase
HP17_08229	Hypothetical protein
HP17_08234	Flavodoxin FldA
HP17_08239	Metal-binding heat shock protein
HP17_08249	Pyrroline-5-carboxylate reductase
HP17_08264	Outer membrane protein HopI
HP17_08269	MurG undecaprenyldiphospho-muramoylpentapeptide beta- N-acetylglucosaminyltransferase
HP17_08274	Flagellar assembly protein FliW
HP17_08279	ValS valyl-tRNA synthetase
HP17_08284	Signal recognition particle protein
HP17_08289	RpsP 30S ribosomal protein S16
HP17_08294	Hypothetical protein
HP17_08299	RimM 16S rRNA-processing protein RimM
HP17_08304	TrmD tRNA (guanine-N(1)-)-methyltransferase
HP17_08309	RplS 50S ribosomal protein L19
HP17_08334	Peptidyl-prolyl cis-trans isomerase C
HP17_08339	Fructose-bisphosphate aldolase
HP17_08344	Elongation factor P
HP17_08384	Sialic acid synthase
HP17_08389	ABC transporter
HP17_08394	Apolipoprotein N-acyltransferase
HP17_08404	Hypothetical protein
HP17_08419	Hypothetical protein
HP17_08521	Cadmium, zinc and cobalt-transporting ATPase
HP17_08536	Mg chelatase-related protein; ComM protein
HP17_08541	Def peptide deformylase
HP17_08546	ClpP ATP-dependent Clp protease proteolytic subunit
HP17_08551	Tig trigger factor
HP17_08556	Outer membrane protein HorG
HP17_08561	Neuraminyllactose-binding hemagglutinin HpaA
HP17_08566	MoaC molybdenum cofactor biosynthesis protein MoaC
HP17_08576	Molybdopterin converting factor, subunit
HP17_08586	RibA GTP cyclohydrolase II
HP17_08606	Bifunctional 3,4-dihydroxy-2-butanone 4- phosphate synthase/GTP cyclohydrolase II protein

HP17_08611	Lipooligosaccharide 5G8 epitope biosynthesis- associated protein
HP17_08621	Thioredoxin
HP17_08626	Hypothetical protein
HP17_08631	Homoserine dehydrogenase
HP17_08636	UvrC excinuclease ABC subunit C
HP17_08641	Hypothetical protein
HP17_08661	Hypothetical protein
HP17_08666	MotB flagellar motor protein MotB
HP17_08671	Flagellar motor protein MotA
HP17_08676	Thiamin biosynthesis protein (thiF)
HP17_08681	Hydrolase
HP17_08686	Hypothetical protein
HP17_08691	Hypothetical protein
HP17_08701	FliL flagellar basal body-associated protein FliL
HP17_08706	AcpS 4'-phosphopantetheinyl transferase
HP17_08716	Hydrolase
HP17_08726	Hypothetical protein
HP17_08731	rRNA large subunit methyltransferase
HP17_08736	Acetyl-CoA carboxylase subunit beta
HP17_08741	Putative recombination protein RecO
HP17_08746	Putative competence/damage-inducible protein CinA
HP17_08751	Hypothetical protein
HP17_08761	Prolipoprotein diacylglyceryl transferase
HP17_08766	Hypothetical protein
HP17_08781	Hypothetical protein
HP17_08786	GlyQ glycyl-tRNA synthetase subunit alpha

*17874 annotation.

Appendix 7 List of "disease core"-specific genes*

Locus tag	Product
HP0018	Hypothetical protein
HP0026	Type II citrate synthase
HP0048	Transcriptional regulator (hypF)
HP0069	Urease accessory protein UreF
HP0070	Urease accessory protein UreE
HP0072	Urease subunit beta
HP0092	Type II restriction enzyme M protein (hsdM)
HP0116	DNA topoisomerase I
HP0135	Hypothetical protein
HP0138	Iron-sulfur protein
HP0141	L-lactate permease (lctP)
HP0150	Hypothetical protein
HP0153	Recombinase A
HP0200	50S ribosomal protein L32
HP0213	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA
HP0215	CDP-diglyceride synthetase (cdsA)
HP0228	Hypothetical protein
HP0229	Hypothetical protein
HP0282	Hypothetical protein
HP0290	Diaminopimelate decarboxylase
HP0320	Hypothetical protein
HP0351	Flagellar MS-ring protein
HP0380	Glutamate dehydrogenase
HP0471	Glutathione-regulated potassium-efflux system protein (kefB)
HP0498	Sodium- and chloride-dependent transporter
HP0509	Glycolate oxidase subunit (glcD)
HP0555	Hypothetical protein
HP0567	Hypothetical protein
HP0582	Hypothetical protein
HP0607	Acriflavine resistance protein (acrB)
HP0610	Toxin-like outer membrane protein
HP0655	Protective surface antigen D15
HP0686	Iron(III) dicitrate transport protein (fecA)
HP0705	Excinuclease ABC subunit A
HP0717	DNA polymerase III subunits gamma and tau
HP0718	Hypothetical protein
HP0738	D-alanyl-alanine synthetase A
HP0761	Hypothetical protein
HP0779	Bifunctional aconitate hydratase 2/2-methylisocitrate dehydratase
HP0781	Hypothetical protein
HP0782	Hypothetical protein

HP0818	Osmoprotection protein (proWX)
HP0826	Lipooligosaccharide 5G8 epitope biosynthesis-associated protein (lex2B)
HP0834	GTP-binding protein EngA
HP0839	Outer membrane protein P1 (ompP1)
HP0879	Hypothetical protein
HP0887	Vacuolating cytotoxin
HP0908	Flagellar hook protein FlgE
HP0911	Rep helicase, single-stranded DNA-dependent ATPase (rep)
HP1017	Amino acid permease
HP1022	Hypothetical protein
HP1023	Hypothetical protein
HP1039	Hypothetical protein
HP1054	Hypothetical protein
HP1057	Hypothetical protein
HP1073	Copper ion binding protein (copP)
HP1083	Hypothetical protein
HP1106	Hypothetical protein
HP1157	Hypothetical protein
HP1159	Cell filamentation protein (fic)
HP1166	Glucose-6-phosphate isomerase
HP1168	Carbon starvation protein (cstA)
HP1179	Phosphopentomutase
HP1190	Histidyl-tRNA synthetase
HP1222	D-lactate dehydrogenase (dld)
HP1232	Dihydropteroate synthase (folP)
HP1272	NADH dehydrogenase subunit M
HP1275	Phosphomannomutase
HP1286	Hypothetical protein
HP1337	Hypothetical protein
HP1359	Hypothetical protein
HP1414	Hypothetical protein
HP1416	Lipopolysaccharide 1,2-glucosyltransferase (rfaJ)
HP1422	Isoleucyl-tRNA synthetase
HP1460	DNA polymerase III subunit alpha
HP1465	ABC transporter ATP-binding protein
HP1502	Hypothetical protein
HP1503	Cation-transporting ATPase, P-type (copA)
HP1584	DNA-binding/iron metalloprotein/AP endonuclease

*26695 annotation.

Appendix 8 List of "gastritis core"-specific genes*

Locus tag	Product
HP0002	6,7-dimethyl-8-ribityllumazine synthase
HP0004	Carbonic anhydrase (icfA)
HP0013	Hypothetical protein
HP0018	Hypothetical protein
HP0025	Hypothetical protein
HP0026	Type II citrate synthase
HP0028	Hypothetical protein
HP0034	Aspartate alpha-decarboxylase
HP0035	Hypothetical protein
HP0037	NADH-ubiquinone oxidoreductase subunit
HP0048	Transcriptional regulator (hypF)
HP0050	Adenine-specific DNA methyltransferase
HP0051	Cytosine specific DNA methyltransferase (DDEM)
HP0064	Hypothetical protein
HP0065	Hypothetical protein
HP0069	Urease accessory protein UreF
HP0070	Urease accessory protein UreE
HP0072	Urease subunit beta
HP0079	Hypothetical protein
HP0082	Methyl-accepting chemotaxis transducer (tlpC)
HP0092	Type II restriction enzyme M protein (hsdM)
HP0098	Threonine synthase
HP0101	Hypothetical protein
HP0103	Methyl-accepting chemotaxis protein (tlpB)
HP0107	Cysteine synthetase (cysK)
HP0111	Heat-inducible transcription repressor
HP0116	DNA topoisomerase I
HP0135	Hypothetical protein
HP0138	Iron-sulfur protein
HP0141	L-lactate permease (lctP)
HP0142	DNA glycosylase MutY
HP0150	Hypothetical protein
HP0151	Hypothetical protein
HP0153	Recombinase A
HP0154	Phosphopyruvate hydratase
HP0159	Lipopolysaccharide 1,2-glucosyltransferase (rfaJ)
HP0168	Hypothetical protein
HP0172	Molybdopterin biosynthesis protein (moeA)
HP0200	50S ribosomal protein L32
HP0201	Glycerol-3-phosphate acyltransferase PlsX
HP0213	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA

HP0215	CDP-diglyceride synthetase (cdsA)
HP0222	Hypothetical protein
HP0228	Hypothetical protein
HP0229	Hypothetical protein
HP0233	Hypothetical protein
HP0248	Hypothetical protein
HP0252	Hypothetical protein
HP0263	Adenine-specific DNA methyltransferase
HP0265	Cytochrome c biogenesis protein (ccdA)
HP0282	Hypothetical protein
HP0290	Diaminopimelate decarboxylase
HP0292	Hypothetical protein
HP0293	Para-aminobenzoate synthetase (pabB)
HP0300	Dipeptide ABC transporter permease (dppC)
HP0302	Dipeptide ABC transporter ATP-binding protein (dppF)
HP0304	Hypothetical protein
HP0320	Hypothetical protein
HP0322	Poly E-rich protein
HP0326	CMP-N-acetylneuraminic acid synthetase
HP0337	Hypothetical protein
HP0347	Hypothetical protein
HP0351	Flagellar MS-ring protein
HP0367	Hypothetical protein
HP0369	Hypothetical protein
HP0371	Biotin carboxyl carrier protein (fabE)
HP0373	Hypothetical protein
HP0376	Ferrochelatase
HP0380	Glutamate dehydrogenase
HP0389	Iron-dependent superoxide dismutase
HP0404	Protein kinase C inhibitor (SP:P16436)
HP0407	Biotin sulfoxide reductase (bisC)
HP0463	Type I restriction enzyme M protein (hsdM)
HP0471	Glutathione-regulated potassium-efflux system protein (kefB)
HP0473	Molybdenum ABC transporter periplasmic molybdate-binding protein (modA)
HP0474	Molybdenum ABC transporter permease (modB)
HP0475	Molybdenum ABC transporter ATP-binding protein (modD)
HP0482	Hypothetical protein
HP0485	Catalase-like protein
HP0491	50S ribosomal protein L28
HP0497	Sodium- and chloride-dependent transporter
HP0498	Sodium- and chloride-dependent transporter
HP0509	Glycolate oxidase subunit (glcD)
HP0516	ATP-dependent protease ATP-binding subunit HslU

HP0520	Cag pathogenicity island protein (cag1)
HP0522	Cag pathogenicity island protein (cag3)
HP0523	Cag pathogenicity island protein (cag4)
HP0525	VirB11-like protein
HP0526	Cag pathogenicity island protein (cag6)
HP0527	Cag pathogenicity island protein (cag7)
HP0528	Cag pathogenicity island protein (cag8)
HP0529	Cag pathogenicity island protein (cag9)
HP0530	Cag pathogenicity island protein (cag10)
HP0531	Cag pathogenicity island protein (cag11)
HP0532	Cag pathogenicity island protein (cag12)
HP0534	Cag pathogenicity island protein (cag13)
HP0537	Cag pathogenicity island protein (cag16)
HP0538	Cag pathogenicity island protein (cag17)
HP0539	Cag pathogenicity island protein (cag18)
HP0540	Cag pathogenicity island protein (cag19)
HP0541	Cag pathogenicity island protein (cag20)
HP0542	Cag pathogenicity island protein (cag21)
HP0543	Cag pathogenicity island protein (cag22)
HP0545	Cag pathogenicity island protein (cag24)
HP0546	Cag pathogenicity island protein (cag25)
HP0547	Cag pathogenicity island protein (cag26)
HP0552	Hypothetical protein
HP0554	Hypothetical protein
HP0555	Hypothetical protein
HP0557	Acetyl-CoA carboxylase carboxyltransferase subunit alpha
HP0567	Hypothetical protein
HP0578	Hypothetical protein
HP0582	Hypothetical protein
HP0583	Hypothetical protein
HP0589	2-oxoglutarate-acceptor oxidoreductase subunit OorA
HP0591	2-oxoglutarate-acceptor oxidoreductase subunit OorC
HP0597	Penicillin-binding protein 1A (PBP-1A)
HP0607	Acriflavine resistance protein (acrB)
HP0610	Toxin-like outer membrane protein
HP0611	Hypothetical protein
HP0613	ABC transporter ATP-binding protein
HP0621	Recombination and DNA strand exchange inhibitor protein
HP0629	Hypothetical protein
HP0630	Modulator of drug activity (mda66)
HP0636	Hypothetical protein
HP0639	Hypothetical protein
HP0644	Hypothetical protein
HP0655	Protective surface antigen D15

HP0659	Hypothetical protein
HP0660	Hypothetical protein
HP0685	Flagellar biosynthesis protein FliP
HP0686	Iron(III) dicitrate transport protein (fecA)
HP0687	Iron(II) transport protein (feoB)
HP0692	3-oxoadipate CoA-transferase subunit B
HP0693	Short-chain fatty acids transporter
HP0695	Hydantoin utilization protein A (hyuA)
HP0703	Response regulator
HP0705	Excinuclease ABC subunit A
HP0717	DNA polymerase III subunits gamma and tau
HP0718	Hypothetical protein
HP0719	Hypothetical protein
HP0723	L-asparaginase II
HP0724	Anaerobic C4-dicarboxylate transporter
HP0728	Hypothetical protein
HP0729	Hypothetical protein
HP0738	D-alanyl-alanine synthetase A
HP0749	Cell division membrane protein (ftsX)
HP0760	Phosphodiesterase
HP0761	Hypothetical protein
HP0764	Hypothetical protein
HP0769	Molybdopterin-guanine dinucleotide biosynthesis protein A
HP0771	Hypothetical protein
HP0772	N-acetylmuramoyl-L-alanine amidase
HP0779	Bifunctional aconitate hydratase 2/2-methylisocitrate dehydratase
HP0781	Hypothetical protein
HP0782	Hypothetical protein
HP0799	Molybdenum cofactor biosynthesis protein MogA
HP0801	Molybdopterin converting factor, subunit 1 (moaD)
HP0810	Hypothetical protein
HP0818	Osmoprotection protein (proWX)
HP0819	Osmoprotection protein (proV)
HP0825	Thioredoxin reductase
HP0826	Lipooligosaccharide 5G8 epitope biosynthesis-associated protein (lex2B)
HP0827	ss-DNA binding protein 12RNP2 precursor
HP0834	GTP-binding protein EngA
HP0839	Outer membrane protein P1 (ompP1)
HP0840	FlaA1 protein
HP0843	Thiamine-phosphate pyrophosphorylase
HP0844	Thiamine biosynthesis protein (thi)
HP0845	Hydroxyethylthiazole kinase
HP0846	Type I restriction enzyme R protein (hsdR)
HP0852	Hypothetical protein

HP0870	Flagellar hook protein FlgE
HP0879	Hypothetical protein
HP0887	Vacuolating cytotoxin
HP0898	Hydrogenase expression/formation protein (hypD)
HP0899	Hydrogenase expression/formation protein (hypC)
HP0900	Hydrogenase expression/formation protein (hypB)
HP0906	Hypothetical protein
HP0908	Flagellar hook protein FlgE
HP0910	Adenine-specific DNA methyltransferase
HP0911	Rep helicase, single-stranded DNA-dependent ATPase (rep)
HP0915	Iron-regulated outer membrane protein (frpB)
HP0919	Carbamoyl phosphate synthase large subunit
HP0926	tRNA pseudouridine synthase D
HP0957	3-deoxy-D-manno-octulosonic-acid transferase
HP0958	Hypothetical protein
HP0961	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase
HP0969	Cation efflux system protein (czcA)
HP0970	Nickel-cobalt-cadmium resistance protein (nccB)
HP1017	Amino acid permease
HP1022	Hypothetical protein
HP1023	Hypothetical protein
HP1024	Co-chaperone-curved DNA binding protein A (CbpA)
HP1039	Hypothetical protein
HP1054	Hypothetical protein
HP1057	Hypothetical protein
HP1067	Chemotaxis protein (cheY)
HP1072	Copper-transporting ATPase, P-type (copA)
HP1073	Copper ion binding protein (copP)
HP1082	Multidrug resistance protein (msbA)
HP1083	Hypothetical protein
HP1086	Hemolysin (tly)
HP1090	DNA translocase FtsK
HP1104	Cinnamyl-alcohol dehydrogenase ELI3-2 (cad)
HP1106	Hypothetical protein
HP1113	Hypothetical protein
HP1127	Hypothetical protein
HP1139	SpoOJ regulator (soj)
HP1157	Hypothetical protein
HP1159	Cell filamentation protein (fic)
HP1166	Glucose-6-phosphate isomerase
HP1168	Carbon starvation protein (cstA)
HP1175	Hypothetical protein
HP1177	Hypothetical protein
HP1179	Phosphopentomutase

HP1190	Histidyl-tRNA synthetase
HP1192	Secreted protein involved in flagellar motility
HP1220	ABC transporter ATP-binding protein
HP1222	D-lactate dehydrogenase (dld)
HP1231	DNA polymerase III subunit delta'
HP1232	Dihydropteroate synthase (folP)
HP1238	Formamidase
HP1244	30S ribosomal protein S18
HP1245	Single-stranded DNA-binding protein
HP1246	30S ribosomal protein S6
HP1249	Shikimate 5-dehydrogenase
HP1250	Hypothetical protein
HP1252	Oligopeptide ABC transporter periplasmic oligopeptide-binding protein (oppA)
HP1259	Hypothetical protein
HP1263	NADH dehydrogenase subunit D
HP1265	Hypothetical protein
HP1266	NADH dehydrogenase subunit G
HP1272	NADH dehydrogenase subunit M
HP1274	Paralysed flagella protein (pflA)
HP1275	Phosphomannomutase
HP1279	bifunctional indole-3-glycerol phosphate synthase/phosphoribosylanthranilate Isomerase
HP1286	Hypothetical protein
HP1291	Hypothetical protein
HP1308	50S ribosomal protein L24
HP1329	Cation efflux system protein (czcA)
HP1337	Hypothetical protein
HP1339	Biopolymer transport protein (exbB)
HP1340	Biopolymer transport protein (exbD)
HP1341	Siderophore-mediated iron transport protein (tonB)
HP1359	Hypothetical protein
HP1363	Hypothetical protein
HP1364	Histidine kinase sensor protein
HP1365	Response regulator
HP1371	Type III restriction enzyme R protein
HP1378	Competence lipoprotein (comL)
HP1382	Hypothetical protein
HP1387	DNA polymerase III subunit epsilon
HP1393	DNA repair protein (recN)
HP1400	Iron(III) dicitrate transport protein (fecA)
HP1406	Biotin synthase
HP1407	Ribonuclease N
HP1414	Hypothetical protein
HP1416	Lipopolysaccharide 1,2-glucosyltransferase (rfaJ)

HP1420	Flagellum-specific ATP synthase
HP1422	Isoleucyl-tRNA synthetase
HP1423	Hypothetical protein
HP1434	Formyltetrahydrofolate hydrolase (purU)
HP1435	Endopeptidase IV
HP1436	Hypothetical protein
HP1439	Hypothetical protein
HP1445	Biopolymer transport protein (exbB)
HP1446	Biopolymer transport protein (exbD)
HP1447	50S ribosomal protein L34
HP1448	Ribonuclease P, protein component (rnpA)
HP1451	Hypothetical protein
HP1453	Hypothetical protein
HP1460	DNA polymerase III subunit alpha
HP1462	Secreted protein involved in flagellar motility
HP1465	ABC transporter ATP-binding protein
HP1491	Phosphate permease
HP1502	Hypothetical protein
HP1503	Cation-transporting ATPase, P-type (copA)
HP1517	Type IIS restriction enzyme R and M protein (ECO57IR)
HP1529	Chromosome replication initiator DnaA
HP1533	FAD-dependent thymidylate synthase
HP1542	Hypothetical protein
HP1560	Cell division protein (ftsW)
HP1575	ABC transporter
HP1576	DL-methionine transporter ATP-binding subunit
HP1577	ABC transporter permease (yaeE)
HP1581	Methicillin resistance protein (llm)
HP1584	DNA-binding/iron metalloprotein/AP endonuclease

*26695 annotation.

Appendix 9 List of "duodenal ulcer core"-specific genes*

Locus tag	Product
jhp0004	Carbonic anhydrase
jhp0016	Hypothetical protein
jhp0021	Hypothetical protein
jhp0022	Type II citrate synthase
jhp0026	Hypothetical protein
jhp0030	Aspartate alpha-decarboxylase
jhp0031	Hypothetical protein
jhp0033	Hypothetical protein
jhp0041	Transcriptional regulator
jhp0043	Type II DNA modification enzyme
jhp0044	Type II DNA modification enzyme
jhp0064	Urease accessory protein
jhp0065	Urease accessory protein UreE
jhp0067	Urease subunit beta
jhp0085	Type II DNA modification (methyltransferase)
jhp0090	Threonine synthase
jhp0091	Methyl-accepting chemotaxis protein (MCP)
jhp0093	Hypothetical protein
jhp0095	Methyl-accepting chemotaxis protein (MCP)
jhp0099	Cysteine synthase
jhp0103	Heat-inducible transcription repressor
jhp0108	DNA topoisomerase I
jhp0112	Hypothetical protein
jhp0118	Hypothetical protein
jhp0123	Hypothetical protein
jhp0126	Iron-sulfur protein
jhp0129	L-lactate permease
jhp0130	DNA glycosylase MutY
jhp0131	Hypothetical protein
jhp0138	Hypothetical protein
jhp0139	Hypothetical protein
jhp0141	Recombinase A
jhp0142	Phosphopyruvate hydratase
jhp0147	Lipopolysaccharide biosynthesis protein
jhp0151	Histidine kinase sensor protein
jhp0154	Hypothetical protein
jhp0176	Cardiolipin synthase
jhp0186	50S ribosomal protein L32
jhp0187	Glycerol-3-phosphate acyltransferase PlsX
jhp0191	Hypothetical protein
jhp0199	TRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA

jhp0201	CDP-diacylglycerol synthase
jhp0208	Hypothetical protein
jhp0213	Hypothetical protein
jhp0214	Outer membrane protein/porin
jhp0218	Hypothetical protein
jhp0237	Hypothetical protein
jhp0241	Hypothetical protein
jhp0244	Type II DNA modification (methyltransferase)
jhp0245	Hypothetical protein
jhp0246	Hypothetical protein
jhp0248	Type II DNA modification (methyltransferase)
jhp0250	Cytochrome C-type biogenesis protein
jhp0267	Hypothetical protein
jhp0275	Diaminopimelate decarboxylase
jhp0277	Hypothetical protein
jhp0278	P-aminobenzoate synthetase
jhp0286	Peptide ABC transporter ATP-binding protein
jhp0287	Peptide ABC transporter ATP-binding protein
jhp0295	Hypothetical protein
jhp0298	Hypothetical protein
jhp0300	ABC transporter ATP-binding protein
jhp0303	Hypothetical protein
jhp0305	Poly E-rich protein
jhp0310	Flagellar biosynthesis protein
jhp0319	Hypothetical protein
jhp0325	Flagellar MS-ring protein
jhp0330	Hypothetical protein
jhp0335	Septum formation protein
jhp0336	Hypothetical protein
jhp0339	Hemolysin
jhp0342	Hypothetical protein
jhp0343	Multi-drug resistance protein
jhp0344	Hypothetical protein
jhp0350	Hypothetical protein
jhp0352	Copper-associated protein
jhp0353	Copper-transporting P-type ATPase
jhp0358	Response regulator
jhp0361	Hypothetical protein
jhp0368	Hypothetical protein
jhp0371	Hypothetical protein
jhp0376	Hypothetical protein
jhp0377	Translation initiation factor IF-2
jhp0385	Hypothetical protein
jhp0400	Co-chaperone with DnaK

jhp0401	Hypothetical protein
jhp0402	Hypothetical protein
jhp0406	Amino acid permease
jhp0415	Type I restriction enzyme modification subunit
jhp0419	Hypothetical protein
jhp0423	Glutathione-regulated potassium-efflux system protein
jhp0425	Molybdate ABC transporter periplasmic-binding protein
jhp0427	Molybdate ABC transporter ATP-binding protein
jhp0430	Type II DNA modification (methyltransferase
jhp0431	Hypothetical protein
jhp0435	Type II DNA modification (methyltransferase
jhp0437	Hypothetical protein
jhp0443	50S ribosomal protein L28
jhp0446	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase
jhp0447	Hypothetical protein
jhp0449	Transporter
jhp0450	Transporter
jhp0458	Hypothetical protein
jhp0459	Glycolate oxidase
jhp0462	Hypothetical protein
jhp0465	ATP-dependent protease ATP-binding subunit HslU
jhp0468	Hypothetical protein
jhp0469	Cag island protein
jhp0471	Cag island protein
jhp0472	Cag island protein
jhp0473	Cag island protein, DNA transfer protein
jhp0474	Cag island protein, DNA transfer protein
jhp0475	Cag island protein
jhp0476	Cag island protein
jhp0477	Cag island protein
jhp0478	Cag island protein
jhp0479	Cag island protein
jhp0480	Cag island protein
jhp0481	Cag island protein
jhp0482	Cag island protein
jhp0483	Cag island protein
jhp0485	Cag island protein
jhp0486	Cag island protein
jhp0487	Cag island protein
jhp0488	Cag island protein
jhp0489	Cag island protein
jhp0490	Cag island protein
jhp0491	Cag island protein
jhp0492	DNA transfer protein

jhp0493	Cag island protein
jhp0494	Cag island protein
jhp0495	Cag island protein, cytotoxicity associated immunodominant antigen
jhp0499	Hypothetical protein
jhp0502	Hypothetical protein
jhp0503	Hypothetical protein
jhp0504	Acetyl-CoA carboxylase carboxyltransferase subunit alpha
jhp0514	Hypothetical protein
jhp0525	Hypothetical protein
jhp0529	Siderophore-mediated IRON transport protein
jhp0530	Hypothetical protein
jhp0533	Hypothetical protein
jhp0537	2-oxoglutarate-acceptor oxidoreductase subunit OorA
jhp0539	2-oxoglutarate-acceptor oxidoreductase subunit OorC
jhp0544	Penicillin-binding protein
jhp0547	Secretion/efflux ABC transporter ATP-binding protein
jhp0550	Hypothetical protein
jhp0554	Efflux transporter
jhp0556	Vacuolating cytotoxin (VacA) paralog
jhp0563	Lipopolysaccharide biosynthesis protein
jhp0565	Recombination and DNA strand exchange inhibitor protein
jhp0572	Hypothetical protein
jhp0573	Hypothetical protein
jhp0579	Hypothetical protein
jhp0581	Hypothetical protein
jhp0582	Hypothetical protein
jhp0589	Hypothetical protein
jhp0596	Alpha (1,3)-fucosyltransferase
jhp0600	Protective surface antigen D15
jhp0604	Hypothetical protein
jhp0605	Hypothetical protein
jhp0619	Hypothetical protein
jhp0626	Iron(III) dicitrate transport protein
jhp0627	Ferrous iron transport protein B
jhp0635	Short-chain fatty acids transporter
jhp0636	3-oxoacid CoA-transferase subunit B
jhp0643	Transcriptional regulator
jhp0644	Excinuclease ABC subunit A
jhp0654	Hypothetical protein
jhp0655	DNA polymerase III subunits gamma/tau
jhp0656	Hypothetical protein
jhp0657	Hypothetical protein
jhp0665	Hypothetical protein
jhp0666	Hypothetical protein

jhp0675	D-alanyl-alanine synthetase A
jhp0681	Hypothetical protein
jhp0686	Hypothetical protein
jhp0697	Phosphodiesterase
jhp0698	Hypothetical protein
jhp0706	Molybdopterin-guanine dinucleotide biosynthesis protein A
jhp0707	Flagellar biosynthesis protein FlhB
jhp0708	Hypothetical protein
jhp0709	N-acetylmuramoyl-L-alanine amidase
jhp0716	Bifunctional aconitate hydratase 2/2-methylisocitrate dehydratase
jhp0717	Hypothetical protein
jhp0718	Hypothetical protein
jhp0719	Hypothetical protein
jhp0739	Hypothetical protein
jhp0743	Iron(III) dicitrate transport protein
jhp0746	Hypothetical protein
jhp0757	Osmoprotection binding protein
jhp0758	Osmoprotection ATP-binding protein
jhp0764	Thioredoxin reductase
jhp0765	Lipopolysaccharide biosynthesis protein
jhp0766	Hypothetical protein
jhp0773	GTP-binding protein EngA
jhp0777	Hypothetical protein
jhp0778	Sugar nucleotide biosynthesis protein
jhp0786	Type I restriction enzyme modification subunit
jhp0788	Hypothetical protein
jhp0790	Guanosine 5'-monophosphate oxidoreductase
jhp0797	Hypothetical protein
jhp0804	Flagellar hook protein FlgE
jhp0812	Hypothetical protein
jhp0817	Hypothetical protein
jhp0819	Vacuolating cytotoxin
jhp0835	Hydrogenase expression/formation protein
jhp0836	Hydrogenase expression/formation protein
jhp0837	Hydrogenase expression/formation protein
jhp0842	Hypothetical protein
jhp0844	Flagellar hook protein FlgE
jhp0845	Hypothetical protein
jhp0846	Type II DNA modification (methyltransferase
jhp0847	ATP-dependent helicase
jhp0850	Hypothetical protein
jhp0851	IRON-regulated outer membrane protein
jhp0853	Carbamoyl phosphate synthase large subunit
jhp0856	Vacuolating cytotoxin (VacA) paralog

jhp0857	Hypothetical protein
jhp0860	TRNA pseudouridine synthase D
jhp0871	Proline/betaine transporter
jhp0880	Hypothetical protein
jhp0881	Hypothetical protein
jhp0891	3-deoxy-D-manno-octulosonic-acid transferase
jhp0892	Hypothetical protein
jhp0895	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase
jhp0902	Hypothetical protein
jhp0903	Cation efflux system protein
jhp0904	Cation efflux system protein
jhp0907	Hypothetical protein
jhp0915	Hypothetical protein
jhp0954	Hypothetical protein
jhp0965	Hypothetical protein
jhp0967	Methionyl-tRNA synthetase
jhp0974	S/N-oxide reductase
jhp0977	HIT family protein
jhp0987	Hypothetical protein
jhp0992	Iron-dependent superoxide dismutase
jhp0994	Primosome assembly protein PriA
jhp0999	Zinc-metallo protease
jhp1001	Glutamate dehydrogenase
jhp1003	Cytochrome C-type biogenesis protein
jhp1005	Ferrochelatase
jhp1008	Hypothetical protein
jhp1010	Biotin carboxyl carrier protein
jhp1012	Type II DNA modification enzyme
jhp1013	Hypothetical protein
jhp1014	Hypothetical protein
jhp1015	Spore coat polysaccharide biosynthesis protein C
jhp1023	Short chain alcohol dehydrogenase
jhp1030	Zinc-dependent alcohol dehydrogenase
jhp1032	Lipopolysaccharide biosynthesis protein
jhp1033	Hypothetical protein
jhp1056	Hypothetical protein
jhp1067	Hypothetical protein
jhp1069	Methionyl-tRNA formyltransferase
jhp1084	Hypothetical protein
jhp1086	CAMP-induced cell filamentation protein
jhp1092	Hypothetical protein
jhp1093	Glucose-6-phosphate isomerase
jhp1095	Hypothetical protein
jhp1102	Hypothetical protein

jhp1103	Outer membrane function
jhp1105	Phosphopentomutase
jhp1115	Histidyl-tRNA synthetase
jhp1117	Motility protein
jhp1121	DNA-directed RNA polymerase subunit beta/beta'
jhp1141	ABC transporter ATP-binding protein
jhp1143	D-lactate dehydrogenase
jhp1152	DNA polymerase III subunit delta'
jhp1153	Dihydropteroate synthase
jhp1162	Alanyl-tRNA synthetase
jhp1170	Shikimate 5-dehydrogenase
jhp1172	Peptide ABC transporter ATP-binding protein
jhp1173	Hypothetical protein
jhp1180	Hypothetical protein
jhp1186	NADH oxidoreductase I
jhp1187	NADH dehydrogenase subunit G
jhp1193	NADH dehydrogenase subunit M
jhp1196	Phosphomannomutase
jhp1200	Bifunctional indole-3-glycerol phosphate synthase/phosphoribosylanthranilate isomerase
jhp1206	Hypothetical protein
jhp1211	Hypothetical protein
jhp1228	50S ribosomal protein L24
jhp1241	Hypothetical protein
jhp1244	Hypothetical protein
jhp1249	Cation efflux system protein
jhp1256	Hypothetical protein
jhp1258	Biopolymer transport protein
jhp1259	Biopolymer transport EXBD protein
jhp1277	Hypothetical protein
jhp1279	DNA transfer protein
jhp1281	Hypothetical protein
jhp1284	Type II DNA modification (methyltransferase
jhp1285	Type III restriction enzyme R protein
jhp1292	Hypothetical protein
jhp1295	Endonuclease
jhp1296	Type III DNA modification enzyme
jhp1298	Biotin synthase
jhp1299	Ribonuclease N
jhp1309	Hypothetical protein
jhp1310	TRNA delta(2)-isopentenylpyrophosphate transferase
jhp1311	Lipopolysaccharide biosynthesis protein
jhp1315	Flagellum-specific ATP synthase
jhp1317	Isoleucyl-tRNA synthetase

jhp1318	Hypothetical protein
jhp1338	Biopolymer transport protein
jhp1339	Biopolymer transport protein
jhp1341	Ribonuclease P protein component
jhp1344	Hypothetical protein
jhp1353	DNA polymerase III subunit alpha
jhp1355	Motility protein
jhp1358	ABC transporter ATP-binding protein
jhp1359	Hypothetical protein
jhp1365	Type II DNA modification enzyme
jhp1384	Phosphate permease
jhp1394	Hypothetical protein
jhp1395	Hypothetical protein
jhp1396	Component of cation transport for cbb3-type oxidase
jhp1401	Ferrodoxin
jhp1405	Iron-regulated outer membrane protein
jhp1411	Type III DNA modification enzyme
jhp1417	Chromosome replication initiator DnaA
jhp1418	Hypothetical protein
jhp1421	FAD-dependent thymidylate synthase
jhp1422	Type I restriction enzyme (specificity subunit)
jhp1423	Type I restriction enzyme modification subunit
jhp1433	Hypothetical protein
jhp1438	DNA polymerase III subunit epsilon
jhp1439	Ribulose-phosphate 3-epimerase
jhp1456	Hypothetical protein
jhp1457	Hypothetical protein
jhp1468	Rod shape-determining protein
jhp1483	Flagellar biosynthesis protein
jhp1485	ABC transporter permease
jhp1487	Hypothetical protein
jhp1488	Undecaprenyl-phosphate-alpha-N- acetylglucosaminyltransferase
jhp1491	DNA-binding/iron metalloprotein/AP endonuclease

*J99 annotation.

Appendix 10 List of "gastric cancer core"-specific genes

Locus tag	Product
HPF32_0011	Argininosuccinate synthase
HPF32_0016	Hypothetical protein
HPF32_0036	Putative transcriptional regulator
HPF32_0037	Signal-transducing protein, histidine kinase
HPF32_0040	Competence locus E
HPF32_0042	Hypothetical protein
HPF32_0049	Adenine specific DNA methyltransferase
HPF32_0050	Adenine specific DNA methyltransferase
HPF32_0067	Hypothetical protein
HPF32_0071	Hypothetical protein
HPF32_0079	Urease accessory protein
HPF32_0080	Urease accessory protein UreE
HPF32_0082	Urease subunit alpha
HPF32_0104	Type II restriction enzyme M protein
HPF32_0110	Threonine synthase
HPF32_0115	Methyl-accepting chemotaxis protein
HPF32_0127	DNA topoisomerase I
HPF32_0137	Hypothetical protein
HPF32_0143	Hypothetical protein
HPF32_0146	Hypothetical protein
HPF32_0149	L-lactate permease
HPF32_0151	C(4)-dicarboxylates and tricarboxylates/succinate antiporter
HPF32_0159	Hypothetical protein
HPF32_0162	Recombinase A
HPF32_0172	Putative histidine kinase sensor protein
HPF32_0198	Hypothetical protein
HPF32_0208	50S ribosomal protein L32
HPF32_0216	Lipopolysaccharide 1,2-glycosyltransferase
HPF32_0222	TRNA uridine 5-carboxymethylaminomethyl modification protein GidA
HPF32_0224	CDP-diglyceride synthetase
HPF32_0232	Hypothetical protein
HPF32_0237	Putative sulfate permease
HPF32_0238	Outer membrane protein HopA
HPF32_0258	Hypothetical protein
HPF32_0261	Oligopeptide permease integral membrane protein
HPF32_0266	Hypothetical protein
HPF32_0291	Hypothetical protein
HPF32_0300	Diaminopimelate decarboxylase
HPF32_0314	Hypothetical protein
HPF32_0320	Hypothetical protein
HPF32_0323	Nitrite extrusion protein

HPF32_0328	Cytochrome c-type biogenesis protein
HPF32_0330	Glutamate dehydrogenase
HPF32_0332	Putative zinc-metallo protease
HPF32_0337	Primosome assembly protein PriA
HPF32_0367	Methionyl-tRNA synthetase
HPF32_0380	Hypothetical protein
HPF32_0393	Hypothetical protein
HPF32_0412	Hypothetical protein
HPF32_0434	Putative vacuolating cytotoxin VacA
HPF32_0439	Iron-regulated outer membrane protein
HPF32_0440	Putative outer membrane protein
HPF32_0443	Rep helicase, single-stranded DNA-dependent ATPase
HPF32_0445	Type II restriction enzyme
HPF32_0446	Flagellar hook protein FlgE
HPF32_0462	Vacuolating cytotoxin A
HPF32_0464	Virulence factor MviN
HPF32_0470	Hypothetical protein
HPF32_0475	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase
HPF32_0476	Hypothetical protein
HPF32_0479	Sodium- and chloride-dependent transporter
HPF32_0480	Phospholipase A1
HPF32_0488	Plasminogen binding protein
HPF32_0489	Putative Glycolate oxidase
HPF32_0491	Urease-enhancing factor
HPF32_0499	Hypothetical protein
HPF32_0500	Cag pathogenicity island protein
HPF32_0501	Cag pathogenicity island protein
HPF32_0502	Cag island protein
HPF32_0504	Cag pathogenicity island protein
HPF32_0505	Cag pathogenicity island protein
HPF32_0507	Cag pathogenicity island protein
HPF32_0508	Cag pathogenicity island protein
HPF32_0509	Cag pathogenicity island protein
HPF32_0510	Cag pathogenicity island protein
HPF32_0511	Cag pathogenicity island protein
HPF32_0512	Cag pathogenicity island protein
HPF32_0513	Cag pathogenicity island protein
HPF32_0514	Cag pathogenicity island protein
HPF32_0515	Cag pathogenicity island protein
HPF32_0516	Cag pathogenicity island protein
HPF32_0517	Cag island protein
HPF32_0518	Cag pathogenicity island protein
HPF32_0519	Cag pathogenicity island protein
HPF32_0520	DNA transfer protein

HPF32_0521	Cag pathogenicity island protein
HPF32_0522	Cag pathogenicity island protein
HPF32_0523	Cag pathogenicity island protein
HPF32_0531	Hypothetical protein
HPF32_0542	Membrane protein
HPF32_0557	Hypothetical protein
HPF32_0561	Hypothetical protein
HPF32_0578	Hypothetical protein
HPF32_0582	Cytoplasmic pump protein of the hefABC efflux system HefC
HPF32_0584	Putative vacuolating cytotoxin (VacA)-like protein
HPF32_0595	Putative lipopolysaccharide biosynthesis protein
HPF32_0607	Modulator of drug activity
HPF32_0615	Outer membrane protein OipA1/A2
HPF32_0621	Excinuclease ABC subunit A
HPF32_0629	Hydantoin utilization protein A
HPF32_0636	Iron(II) transport protein
HPF32_0637	Iron(III) dicitrate transport protein FecA1
HPF32_0664	Putative outer membrane protein
HPF32_0684	DNA polymerase III subunits gamma and tau
HPF32_0685	Hypothetical protein
HPF32_0688	L-asparaginase II
HPF32_0689	Anaerobic C4-dicarboxylate transporter
HPF32_0704	D-alanyl-alanine synthetase A
HPF32_0729	Hypothetical protein
HPF32_0740	Hypothetical protein
HPF32_0748	Bifunctional aconitate hydratase 2/2-methylisocitrate dehydratase
HPF32_0749	Hypothetical protein
HPF32_0750	Hypothetical protein
HPF32_0751	Outer membrane protein HofE
HPF32_0766	Molybdenum cofactor biosynthesis protein
HPF32_0785	Osmoprotection protein
HPF32_0794	Lipooligosaccharide 5G8 epitope biosynthesis-associated protein
HPF32_0802	GTP-binding protein EngA
HPF32_0806	Outer membrane protein P1
HPF32_0811	Thiamine biosynthesis protein
HPF32_0812	Thiamine phosphate pyrophosphorylase
HPF32_0813	Type I restriction enzyme R protein
HPF32_0822	Guanosine 5'-monophosphate oxidoreductase
HPF32_0829	Hypothetical protein
HPF32_0844	Hypothetical protein
HPF32_0854	Outer membrane protein HopK
HPF32_0856	Molybdenum ABC transporter ModD
HPF32_0861	Glutathione-regulated potassium-efflux system protein
HPF32_0865	Integral membrane protein

HPF32_0879	Amino acid permease
HPF32_0883	DNA-polymerase I-like 5'-3' exonuclease
HPF32_0884	Hypothetical protein
HPF32_0900	Hypothetical protein
HPF32_0909	Translation initiation factor IF-2
HPF32_0910	Hypothetical protein
HPF32_0915	Hypothetical protein
HPF32_0918	Hypothetical protein
HPF32_0925	Hypothetical protein
HPF32_0933	Copper ion binding protein
HPF32_0935	Hypothetical protein
HPF32_0942	Outer membrane protein HofB
HPF32_0958	Flagellar MS-ring protein
HPF32_0962	Hypothetical protein
HPF32_0978	CMP-N-acetylneuraminic acid synthetase
HPF32_0979	CMP-N-acetylneuraminic acid synthetase
HPF32_0983	Poly E-rich protein
HPF32_0985	Sec-independent protein translocase protein
HPF32_1019	Outer membrane protein HomC
HPF32_1026	Spore coat polysaccharide biosynthesis protein C
HPF32_1033	Short chain alcohol dehydrogenase
HPF32_1042	Putative lipopolysaccharide biosynthesis protein
HPF32_1043	Hypothetical protein
HPF32_1050	Outer membrane protein HorI
HPF32_1059	Hypothetical protein
HPF32_1077	Methionyl-tRNA formyltransferase
HPF32_1092	Outer membrane protein HopL
HPF32_1094	Cell filamentation protein
HPF32_1100	Hypothetical protein
HPF32_1101	Glucose-6-phosphate isomerase
HPF32_1103	Carbon starvation protein
HPF32_1111	Outer membrane protein
HPF32_1113	Phosphopentomutase
HPF32_1123	Histidyl-tRNA synthetase
HPF32_1125	Hypothetical protein
HPF32_1129	DNA-directed RNA polymerase subunit beta/beta'
HPF32_1151	D-lactate dehydrogenase
HPF32_1160	DNA polymerase III subunit delta'
HPF32_1161	Dihydropteroate synthase
HPF32_1167	Formamidase
HPF32_1168	Hypothetical protein
HPF32_1170	Alanyl-tRNA synthetase
HPF32_1174	30S ribosomal protein S18
HPF32_1176	30S ribosomal protein S6

HPF32_1180	Hypothetical protein
HPF32_1181	Putative peptide ABC transporter ATP-binding protein
HPF32_1193	NADH dehydrogenase subunit D
HPF32_1202	NADH dehydrogenase subunit M
HPF32_1204	Paralysed flagella protein
HPF32_1205	Phosphomannomutase
HPF32_1208	Bifunctional indole-3-glycerol phosphate synthase/ phosphoribosylanthranilate isomerase
HPF32_1214	Hypothetical protein
HPF32_1247	ATP-binding protein
HPF32_1263	Hypothetical protein
HPF32_1270	Transcriptional regulator
HPF32_1281	Aspartate alpha-decarboxylase
HPF32_1287	Hypothetical protein
HPF32_1289	Type II citrate synthase
HPF32_1291	Putative type III restriction enzyme M protein
HPF32_1292	Putative type III restriction enzyme
HPF32_1296	Hypothetical protein
HPF32_1300	Hypothetical protein
HPF32_1302	Lipopolysaccharide 1,2-glucosyltransferase
HPF32_1309	Isoleucyl-tRNA synthetase
HPF32_1319	Formyltetrahydrofolate hydrolase
HPF32_1320	Protease IV
HPF32_1321	Hypothetical protein
HPF32_1334	50S ribosomal protein L34
HPF32_1340	Outer membrane protein HomD
HPF32_1347	DNA polymerase III subunit alpha
HPF32_1352	ABC transporter ATP-binding protein
HPF32_1353	Hypothetical protein
HPF32_1378	Phosphate permease
HPF32_1390	Putative outer membrane protein
HPF32_1391	Hypothetical protein
HPF32_1392	Putative cation transporting P-type ATPase
HPF32_1397	Ferrodoxin-like protein
HPF32_1401	Putative IRON-regulated outer membrane protein
HPF32_1406	Type III restriction enzyme
HPF32_1415	Purine nucleoside phosphorylase
HPF32_1423	Iron(III) dicitrate transport protein
HPF32_1430	DNA repair protein
HPF32_1437	Ribulose-phosphate 3-epimerase
HPF32_1468	Hypothetical protein
HPF32_1482	ABC transporter ATP-binding protein
HPF32_1485	Hypothetical protein
HPF32_1489	O-sialoglycoprotein endopeptidase

Appendix 11 List of "MALT lymphoma core"-specific genes*

Locus tag	Product
HELPY_0002	6,7-dimethyl-8-ribityllumazine synthase
HELPY_0004	Beta-carbonic anhydrase
HELPY_0007	Outer membrane protein HopZ
HELPY_0011	Hypothetical protein
HELPY_0016	Hypothetical protein
HELPY_0021	Hypothetical protein
HELPY_0022	Outer membrane protein HopD
HELPY_0024	Type II citrate synthase
HELPY_0026	Hypothetical protein
HELPY_0028	Hypothetical protein
HELPY_0032	Aspartate alpha-decarboxylase
HELPY_0033	Hypothetical protein
HELPY_0035	Hypothetical protein
HELPY_0043	Hydrogenase maturation protein HypF
HELPY_0045	Type II adenine methyltransferase
HELPY_0051	Hypothetical protein
HELPY_0052	Hypothetical protein
HELPY_0055	Hypothetical protein
HELPY_0056	Hypothetical protein
HELPY_0057	Hypothetical protein
HELPY_0058	Hypothetical protein
HELPY_0059	Hypothetical protein
HELPY_0065	Urease accessory protein UreF
HELPY_0066	Urease accessory protein UreE
HELPY_0068	Urease subunit beta
HELPY_0078	Methyl-accepting chemotaxis protein
HELPY_0090	Type II adenine methyltransferase
HELPY_0091	Alpha-1,2-fucosyltransferase
HELPY_0095	Threonine synthase
HELPY_0097	Methyl-accepting chemotaxis protein
HELPY_0099	Outer membrane protein
HELPY_0101	Methyl-accepting chemotaxis protein tlpB
HELPY_0106	Cysteine synthase
HELPY_0111	Heat-inducible transcription repressor
HELPY_0116	DNA topoisomerase I
HELPY_0121	Hypothetical protein
HELPY_0128	Hypothetical protein
HELPY_0136	Hypothetical protein
HELPY_0139	Iron-sulfur cluster binding protein
HELPY_0143	L-lactate permease, LctP family; membrane protein
HELPY_0144	DNA glycosylase MutY

HELPY_0145	Sodium/sulfate symporter
HELPY_0154	Hypothetical protein
HELPY_0155	Hypothetical protein
HELPY_0157	Recombinase A
HELPY_0158	Phosphopyruvate hydratase
HELPY_0163	LPS 1,2-glycosyltransferase
HELPY_0168	Histidine kinase sensor protein
HELPY_0171	Hypothetical protein
HELPY_0175	Molybdopterin biosynthesis protein MoeA
HELPY_0193	Phospholipase D
HELPY_0203	50S ribosomal protein L32
HELPY_0204	Glycerol-3-phosphate acyltransferase PlsX
HELPY_0211	LPS 1,2-glycosyltransferase
HELPY_0216	TRNA uridine 5-carboxymethylaminomethyl modification protein GidA
HELPY_0218	Phosphatidate cytidylyltransferase
HELPY_0232	Sulfate permease
HELPY_0233	Outer membrane protein HopA
HELPY_0237	Glutathionylspermidine synthase
HELPY_0253	Hypothetical protein
HELPY_0258	Outer membrane protein HopF
HELPY_0262	Hypothetical protein
HELPY_0265	TCGA site-specific m6A methyltransferase
HELPY_0266	Hypothetical protein
HELPY_0269	Non-functional cytosine methyltransferase
HELPY_0271	Cytochrome C biogenesis protein; membrane protein
HELPY_0288	Hypothetical protein
HELPY_0295	Toxin-like outer membrane protein/vacuolating cytotoxin VacA
HELPY_0296	Diaminopimelate decarboxylas (DAP decarboxylase)
HELPY_0298	Hypothetical protein
HELPY_0299	Para-aminobenzoate (PABA)-synthetase
HELPY_0306	ABC transporter permease; dipeptide transporter protein 3; membrane protein
HELPY_0308	ABC transporter ATP-binding protein; dipeptide transporter protein 4
HELPY_0309	ABC transporter ATP-binding protein; dipeptide transporter protein 5
HELPY_0311	Hypothetical protein
HELPY_0317	NodB-like polysaccharide deacetylase
HELPY_0320	Nitrite extrusion protein, major facilitator family protein; membrane protein
HELPY_0323	Sec-independent protein translocase protein tatA/E-like protein
HELPY_0325	Hypothetical protein
HELPY_0331	Flagellar biosynthesis protein
HELPY_0340	Hypothetical protein
HELPY_0350	Pseudouridine synthase (RNA-uridine isomerase) (RNA pseudouridylate synthase)
HELPY_0354	Flagellar MS-ring protein

HELPY_0360	Hypothetical protein
HELPY_0363	DNA translocase FtsK; membrane protein
HELPY_0364	Hypothetical protein
HELPY_0367	Pore-forming cytolysin
HELPY_0370	Outer membrane protein HofB
HELPY_0371	ABC transporter ATP-binding protein; lipid A and glycerophospholipid transporter; membrane protein
HELPY_0372	Hypothetical protein
HELPY_0376	Hypothetical protein
HELPY_0378	Copper-associated protein
HELPY_0379	Copper-transporting ATPase/P-type transporting ATPase; membrane protein
HELPY_0383	Chemotactic response regulator, two-component system
HELPY_0386	Hypothetical protein
HELPY_0393	Hypothetical protein
HELPY_0396	Hypothetical protein
HELPY_0401	Hypothetical protein
HELPY_0402	Translation initiation factor IF-2
HELPY_0405	Acetyl-CoA synthetase
HELPY_0409	Hypothetical protein
HELPY_0410	Hypothetical protein
HELPY_0414	Hypothetical protein
HELPY_0420	Hypothetical protein
HELPY_0429	Chaperone protein DnaJ
HELPY_0430	Hypothetical protein
HELPY_0431	Hypothetical protein
HELPY_0436	Transporter; amino-acid transporter, AAT family; membrane protein
HELPY_0444	Type I restriction/modification specificity protein
HELPY_0445	Type I restriction enzyme M protein
HELPY_0450	Hypothetical protein
HELPY_0455	Sodium/hydrogen exchanger
HELPY_0457	ABC transporter substrate-binding protein
HELPY_0458	ABC transporter permease; molybdate transporter; membrane protein
HELPY_0459	ABC transporter ATP-binding protein; molybdate transporter
HELPY_0462	Type II adenine methyltransferase
HELPY_0463	Non-functional type II restriction endonuclease
HELPY_0465	Type II cytosine specific DNA methyltransferase
HELPY_0466	Non-functional type II restriction endonuclease
HELPY_0468	Catalase
HELPY_0472	Hypothetical protein
HELPY_0474	Hypothetical protein
HELPY_0482	Flagellar hook protein FlgE
HELPY_0489	Plasminogen-binding protein PgbB
HELPY_0500	Guanosine 5'-monophosphate oxidoreductase
HELPY_0507	Type I R-M system specificity subunit

HELPY_0510	Hydroxyethylthiazole kinase
HELPY_0511	Phosphomethylpyrimidine kinase (HMP-phosphate kinase) (HMP-P kinase)
HELPY_0512	Thiamine-phosphate pyrophosphorylase
HELPY_0515	UDP-GlcNAc C6 dehydratase/C5 epimerase
HELPY_0516	Outer membrane protein P1
HELPY_0520	GTP-binding protein EngA
HELPY_0528	Nucleotide binding protein
HELPY_0529	Glycosyltransferase, family 25
HELPY_0531	Thioredoxin reductase (TRXR) (TR)
HELPY_0538	ABC transporter ATP-binding protein; osmoprotection ABC transporter involved in glycine/betaine/L-proline transport
HELPY_0539	ABC transporter permease; betaine/proline/choline transporter; membrane protein
HELPY_0547	N-6 adenine methyltransferase
HELPY_0557	Hypothetical protein
HELPY_0559	Molybdopterin-converting factor subunit 1
HELPY_0561	Molybdenum cofactor biosynthesis protein MogA
HELPY_0582	Outermembrane protein HofE
HELPY_0583	Hypothetical protein
HELPY_0584	Hypothetical protein
HELPY_0585	Bifunctional aconitate hydratase 2/2-methylisocitrate dehydratase
HELPY_0592	N-acetylmuramoyl-L-alanine amidase
HELPY_0593	Hypothetical protein
HELPY_0594	Flagellar biosynthesis protein FlhB
HELPY_0595	Molybdopterin-guanine dinucleotide biosynthesis protein A
HELPY_0596	Molybdenum cofactor biosynthesis protein A
HELPY_0599	Hypothetical protein
HELPY_0604	Hypothetical protein
HELPY_0605	Phosphodiesterase
HELPY_0616	ABC transporter permease
HELPY_0624	Rod shape-determining protein RodA; membrane protein
HELPY_0629	D-alanyl-alanine synthetase A
HELPY_0636	LeoA protein
HELPY_0638	Hypothetical protein
HELPY_0639	TRNA(Ile)-lysidine synthase
HELPY_0642	Outer membrane protein HopP
HELPY_0643	Anaerobic C4-dicarboxylate transporter
HELPY_0644	L-asparaginase
HELPY_0647	Hypothetical protein
HELPY_0648	L-lysine exporter; membrane protein
HELPY_0649	DNA polymerase III subunits gamma and tau
HELPY_0650	ATPase
HELPY_0656	Outer membrane protein HomA
HELPY_0661	Excinuclease ABC subunit A

HELPY_0662	Hypothetical protein
HELPY_0666	Transcriptional activator
HELPY_0671	Hypothetical protein
HELPY_0679	Short-chain fatty acid transport protein, scFAT family; membrane protein
HELPY_0681	Succinyl-CoA-transferase subunit B
HELPY_0685	Ferrous iron transport protein B; membrane protein
HELPY_0686	Iron(III) dicitrate transport protein FecA
HELPY_0687	Flagellar biosynthesis protein FliP
HELPY_0690	Hypothetical protein
HELPY_0693	Hypothetical protein
HELPY_0711	Hypothetical protein
HELPY_0712	Hypothetical protein
HELPY_0716	Surface antigen protein
HELPY_0727	Hypothetical protein
HELPY_0729	NAD(P)H-flavin oxidoreductase
HELPY_0730	3-hydroxyacid dehydrogenase
HELPY_0732	Hypothetical protein
HELPY_0733	Outer membrane protein HopH
HELPY_0735	Hypothetical protein
HELPY_0741	NAD(P)H oxidoreductase (NADPH quinone reductase)
HELPY_0750	Recombination and DNA strand exchange inhibitor protein
HELPY_0763	Vacuolating cytotoxin VacA-like
HELPY_0765	Cytoplasmic pump proteins of the hefABC efflux system
HELPY_0769	Hypothetical protein
HELPY_0772	ABC transporter permease and ATP-binding protein; membrane protein
HELPY_0775	Penicillin-binding protein 1 (peptidoglycan glycosyltransferase)
HELPY_0783	2-oxoglutarate-acceptor oxidoreductase subunit OorC
HELPY_0785	2-oxoglutarate-acceptor oxidoreductase subunit OorA
HELPY_0789	Hypothetical protein
HELPY_0792	Hypothetical protein
HELPY_0793	Siderophore-mediated iron transport protein
HELPY_0798	Hypothetical protein
HELPY_0810	Hypothetical protein
HELPY_0819	Acetyl-CoA carboxylase carboxyltransferase subunit alpha
HELPY_0820	Hypothetical protein
HELPY_0821	Hypothetical protein
HELPY_0822	Hypothetical protein
HELPY_0824	Methylase
HELPY_0828	Hypothetical protein
HELPY_0831	ATP-dependent protease ATP-binding subunit HslU
HELPY_0842	Urease-enhancing factor Lpp
HELPY_0844	Glycolate oxidase subunit
HELPY_0845	Hypothetical protein

HELPY_0850	Hypothetical protein
HELPY_0855	Sodium-and chloride-dependent transporter; membrane protein
HELPY_0856	Sodium-and chloride-dependent transporter; membrane protein
HELPY_0858	Hypothetical protein
HELPY_0859	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase
HELPY_0862	50S ribosomal protein L28
HELPY_0865	Hypothetical protein
HELPY_0870	Virulence factor MviN protein; membrane protein
HELPY_0872	Vacuolating cytotoxin
HELPY_0878	Hypothetical protein
HELPY_0879	Hypothetical protein
HELPY_0882	Hydrogenase expression/formation protein HypD
HELPY_0883	Hydrogenase expression/formation protein (HUPF/HYPC)
HELPY_0884	Hydrogenase and urease maturation protein
HELPY_0885	Hypothetical protein
HELPY_0889	Phosphotransacetylase
HELPY_0890	Flagellar control protein
HELPY_0892	Flagellar hook protein FlgE
HELPY_0893	Type II restriction endonuclease Hpy8I
HELPY_0894	Type II m6A methylase
HELPY_0895	ATP-dependent single-stranded DNA helicase
HELPY_0900	Outer membrane protein
HELPY_0901	Iron-regulated outer membrane protein
HELPY_0903	Carbamoyl phosphate synthase large subunit
HELPY_0906	Toxin-like outer membrane protein/vacuolating cytotoxin VacA
HELPY_0907	Outer membrane protein HopK
HELPY_0910	TRNA pseudouridine synthase D
HELPY_0924	Hypothetical protein
HELPY_0931	Cation antiporter; membrane protein
HELPY_0932	Hypothetical protein
HELPY_0933	Hypothetical protein
HELPY_0940	NAD(P)H-dependent nitroreductase
HELPY_0943	3-deoxy-D-manno-octulosonic-acid transferase
HELPY_0944	Hypothetical protein
HELPY_0947	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase
HELPY_0949	ATPase
HELPY_0955	Hypothetical protein
HELPY_0956	Cobalt-zinc-cadmium resistance protein, CzcA family; membrane protein
HELPY_0957	Cobalt-zinc-cadmium resistance protein, CzcB family
HELPY_0960	Hypothetical protein
HELPY_0976	Small-conductance mechanosensitive channel; membrane protein
HELPY_1001	Hypothetical protein
HELPY_1002	Hypothetical protein

HELPY_1009	Hypothetical protein
HELPY_1011	Methionyl-tRNA synthetase
HELPY_1021	Hydrolase
HELPY_1031	Hypothetical protein
HELPY_1036	Superoxide dismutase
HELPY_1038	Primosome assembly protein PriA
HELPY_1043	Metalloprotease; membrane protein
HELPY_1045	Glutamate dehydrogenase
HELPY_1046	Alpha1,3-fucosyltransferase
HELPY_1047	Cytochrome c biogenesis protein; membrane protein
HELPY_1049	Ferrochelatase
HELPY_1054	Biotin carboxyl carrier protein of acetyl-CoA carboxylase
HELPY_1056	Adenine methyltransferase
HELPY_1057	Type II restriction endonuclease
HELPY_1058	Hypothetical protein
HELPY_1059	UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase
HELPY_1066	Short chain dehydrogenase
HELPY_1073	Zinc-containing alcohol dehydrogenase
HELPY_1074	Lipopolysaccharide 1,3-galactosyltransferase
HELPY_1075	Hypothetical protein
HELPY_1083	Outer membrane protein
HELPY_1095	Hypothetical protein
HELPY_1100	Hypothetical protein
HELPY_1111	SpoOJ regulator
HELPY_1113	Methionyl-tRNA formyltransferase
HELPY_1114	Hypothetical protein
HELPY_1119	Hypothetical protein
HELPY_1131	Outer membrane protein HopL
HELPY_1133	CAMP-induced cell filamentation protein
HELPY_1139	Permease, MFS superfamily; membrane protein
HELPY_1140	Glucose-6-phosphate isomerase
HELPY_1142	Carbon starvation protein A; membrane protein
HELPY_1149	Permease; membrane protein
HELPY_1150	Outer membrane protein HopQ
HELPY_1152	Phosphopentomutase
HELPY_1160	Hypothetical protein
HELPY_1162	Histidyl-tRNA synthetase
HELPY_1164	Hypothetical protein
HELPY_1166	Aldo-keto reductase
HELPY_1170	DNA-directed RNA polymerase subunit beta/beta'
HELPY_1178	50S ribosomal protein L33
HELPY_1195	ABC transporter ATP-binding protein
HELPY_1197	D-lactate dehydrogenase
HELPY_1206	DNA polymerase III subunit delta'

HELPY_1207	Dihydropteroate synthase (DHPS) (Dihydropteroate pyrophosphorylase)
HELPY_1213	Formamidase
HELPY_1214	Hypothetical protein
HELPY_1216	Alanyl-tRNA synthetase
HELPY_1220	30S ribosomal protein S18
HELPY_1221	Single-stranded DNA-binding protein
HELPY_1222	30S ribosomal protein S6
HELPY_1225	Shikimate 5-dehydrogenase
HELPY_1226	Hypothetical protein
HELPY_1227	ABC transporter permease
HELPY_1228	ABC transporter substrate-binding protein
HELPY_1235	NAD+-dependent deacetylase, Sir2 family
HELPY_1239	NADH dehydrogenase subunit D
HELPY_1241	NADH-ubiquinone oxidoreductase subunit F
HELPY_1242	NADH dehydrogenase subunit G
HELPY_1248	NADH dehydrogenase subunit M
HELPY_1250	Paralysed flagella protein PflA
HELPY_1252	Phosphomannomutase (PMM)
HELPY_1253	Hypothetical protein
HELPY_1256	Bifunctional indole-3-glycerol phosphate synthase/phosphoribosylanthranilate isomerase
HELPY_1262	Hypothetical protein
HELPY_1264	Pantothenate kinase (type III)
HELPY_1265	Hypothetical protein
HELPY_1267	Thiamine pyrophosphokinase
HELPY_1283	50S ribosomal protein L24
HELPY_1297	Hypothetical protein
HELPY_1305	Heavy metal efflux pump CzcA
HELPY_1312	Nicotinate-nucleotide adenyltransferase
HELPY_1314	Biopolymer transport accessory protein; membrane protein
HELPY_1315	Biopolymer transport protein ExbD
HELPY_1316	Siderophore-mediated iron transport protein
HELPY_1317	Outer membrane protein HopM
HELPY_1340	Type II restriction endonuclease
HELPY_1341	Type II m6A methylase
HELPY_1349	Hypothetical protein
HELPY_1351	DNA competence protein; membrane protein
HELPY_1353	Hypothetical protein
HELPY_1354	Hypothetical protein
HELPY_1355	Transcriptional regulatory protein
HELPY_1358	Type III restriction enzyme R protein
HELPY_1365	Hypothetical protein
HELPY_1369	DNA/RNA endonuclease G (nucG)
HELPY_1370	Type III restriction enzyme M protein

HELPY_1371	Type III restriction enzyme R protein
HELPY_1372	Biotin synthase
HELPY_1373	Ribonuclease N
HELPY_1380	Hypothetical protein
HELPY_1382	Hypothetical protein
HELPY_1383	TRNA delta(2)-isopentenylpyrophosphate transferase
HELPY_1384	Lipopolysaccharide biosynthesis protein; LPS glycosyltransferase
HELPY_1389	Flagellum-specific ATP synthase
HELPY_1391	Isoleucyl-tRNA synthetase
HELPY_1392	RNA binding protein
HELPY_1401	Hypothetical protein
HELPY_1402	Type I restriction enzyme specificity protein
HELPY_1403	Formyltetrahydrofolate deformylase
HELPY_1404	Signal peptide protease IV (Protease IV) (Endopeptidase IV)
HELPY_1405	Hypothetical protein
HELPY_1406	Hypothetical protein
HELPY_1407	Hypothetical protein
HELPY_1408	Hypothetical protein
HELPY_1416	Biopolymer transport protein ExbD/TolR; membrane protein
HELPY_1417	Biopolymer transport accessory protein ExbD/TolR
HELPY_1418	Ribonuclease P
HELPY_1421	Hypothetical protein
HELPY_1423	Outer membrane protein HomD
HELPY_1431	DNA polymerase III subunit alpha
HELPY_1434	Hypothetical protein
HELPY_1437	ABC transporter ATP-binding protein
HELPY_1438	ABC transporter permease
HELPY_1464	Transporter; phosphate transporter; membrane protein
HELPY_1473	Hypothetical protein
HELPY_1474	Outer membrane protein HorK
HELPY_1475	Hypothetical protein
HELPY_1476	ATPase, P-type copper-transporter; membrane protein
HELPY_1481	Ferredoxin-like protein
HELPY_1485	Iron-regulated outer membrane protein
HELPY_1491	Type IIS restriction-modification protein
HELPY_1492	Type III R-M system restriction enzyme
HELPY_1500	Chromosomal replication initiation protein
HELPY_1501	Purine nucleoside phosphorylase PunB
HELPY_1504	FAD-dependent thymidylate synthase
HELPY_1505	Hypothetical protein
HELPY_1507	Type I restriction-modification enzyme subunit M
HELPY_1508	Type I restriction-modification enzyme subunit R
HELPY_1510	Iron(III) dicitrate transport protein FecA
HELPY_1514	Hypothetical protein

HELPY_1516	Inorganic polyphosphate/ATP-NAD kinase (Poly(P)/ATP NAD kinase)
HELPY_1517	DNA repair protein
HELPY_1520	Hac prophage II protein
HELPY_1521	Hac prophage II integrase
HELPY_1522	Hac prophage II protein
HELPY_1523	Hac prophage II protein
HELPY_1525	Hac prophage II protein
HELPY_1527	Hac prophage II protein
HELPY_1531	Hypothetical protein
HELPY_1532	Hypothetical protein
HELPY_1534	Hypothetical protein
HELPY_1535	DNA polymerase III subunit epsilon
HELPY_1536	Ribulose-phosphate 3-epimerase
HELPY_1539	N-6 Adenine-specific DNA methylase
HELPY_1540	Hypothetical protein
HELPY_1545	Hypothetical protein
HELPY_1546	Zn-metallopeptidase, M23 family
HELPY_1563	Cell division protein FtsW
HELPY_1564	ABC transporter substrate-binding protein
HELPY_1578	Hypothetical protein
HELPY_1579	DL-methionine transporter ATP-binding subunit
HELPY_1580	ABC transporter permease; D-and L-methionine transport protein; membrane protein
HELPY_1584	Hypothetical protein
HELPY_1585	UDP-phosphate N-acetylgalactosaminyl-1-phosphate transferase
HELPY_1588	DNA-binding/iron metalloprotein/AP endonuclease
HELPY_1590	Hypothetical protein
HELPY_1591	Hypothetical protein
HELPY_CDS1274156R	50S ribosomal protein L36

*HELPY annotation.

Appendix 12 List of unique genes present in the genome of *H. pylori* CCUG 17874 when compared to a number of *H. pylori* strains

Unique 17874	Size (aa)	26695	J99	P79	Product
HP17_00065	34	Absent	Absent	Absent	Hypothetical protein
HP17_00165	35	Absent	Absent	HP79_07905	Hypothetical protein
HP17_00230	93	Absent	Absent	HP79_07835	Hypothetical protein
HP17_00245	>26	Absent	Absent	Absent	Hypothetical protein
HP17_00536	32	Absent	Absent	Absent	Hypothetical protein
HP17_01178	40	Absent	Absent	Absent	Hypothetical protein
HP17_01508	44	HP_0091	Absent	HP79_04682	Type II restriction enzyme R protein (hsdR)
HP17_01533	41	Absent	Absent	Absent	Hypothetical protein
HP17_01783	31	Absent	jhp0915	Absent	Hypothetical protein
HP17_01823	41	Absent	Absent	Absent	Hypothetical protein
HP17_01893	>37	Absent	Absent	Absent	Hypothetical protein
HP17_01925	31	HP_0341	Absent	HP79_00662	Hypothetical protein
HP17_01930	42	Absent	Absent	Absent	Hypothetical protein
HP17_01945	74	HP_0345	Absent	HP79_00682	Hypothetical protein
HP17_02539	52	Absent	Absent	Absent	Hypothetical protein
HP17_02624	>46	Absent	Absent	Absent	Hypothetical protein
HP17_03092	30	HP_0024	Absent	HP79_06174	Hypothetical protein
HP17_03459	66	Absent	Absent	Absent	Hypothetical protein
HP17_03594	30	HP_0502	Absent	HP79_01809	Hypothetical protein
HP17_03699	38	Absent	jhp0884	HP79_06976	Hypothetical protein
HP17_03704	33	Absent	Absent	Absent	Hypothetical protein
HP17_03749	35	Absent	jhp0274	Absent	Hypothetical protein

HP17_03779	42	Absent	Absent	Absent	Hypothetical protein
HP17_04044	31	Absent	Absent	HP79_01040	Hypothetical protein
HP17_04074	39	Absent	Absent	Absent	Hypothetical protein
HP17_04164	30	HP_0024	Absent	HP79_06174	Hypothetical protein
HP17_04194	47	Absent	Absent	HP79_06059	Hypothetical protein
HP17_04626	46	HP_1366	Absent	HP79_05849	Hypothetical protein
HP17_04948	30	Absent	Absent	HP79_08555	Hypothetical protein
HP17_05630	39	Absent	Absent	HP79_02099	Hypothetical protein
HP17_05825	35	HP_0081	Absent	Absent	Hypothetical protein
HP17_06092	102	Absent	Absent	Absent	Hypothetical protein
HP17_06122	42	Absent	Absent	Absent	Hypothetical protein
HP17_06777	37	Absent	Absent	Absent	Hypothetical protein
HP17_06822	36	HP_0225	Absent	HP79_07001	Hypothetical protein
HP17_06837	>38	Absent	Absent	Absent	Hypothetical protein
HP17_07124	>44	HP_1208	Absent	HP79_08967	Hypothetical protein
HP17_07269	36	Absent	Absent	HP79_05686	Hypothetical protein
HP17_07512	71	HP_0502	jhp0454	Absent	Hypothetical protein
HP17_07662	42	Absent	Absent	Absent	Hypothetical protein
HP17_07712	30	HP_0024	Absent	HP79_06174	Hypothetical protein
HP17_07717	90	Absent	Absent	HP79_00130	Hypothetical protein
HP17_07742	73	Absent	jhp0110	HP79_00100	Hypothetical protein
HP17_07957	48	Absent	Absent	Absent	Hypothetical protein
HP17_08134	40	HP_0761	Absent	HP79_03311	Hypothetical protein
HP17_08429	>63	Absent	Absent	HP79_09132	Hypothetical protein

HP17_08454	64	HP_0453	Absent	HP79_04457	Hypothetical protein
HP17_08504	38	HP_0063	Absent	HP79_04844	Hypothetical protein
HP17_08596	43	Absent	Absent	HP79_06391	Hypothetical protein
HP17_08601	45	Absent	Absent	Absent	Hypothetical protein

Appendix 13 List of unique genes present in the genome of *H. pylori* 26695 when compared to *H. pylori* CCUG 17874 and P79

Unique 26695	Size (aa)	17874	P79	Product
HP_0081	40	HP17_05825	Absent	Hypothetical protein
HP_0151	270	Absent	HP79_07890	Hypothetical protein
HP_0161	36	Absent	Absent	Hypothetical protein
HP_0314	39	Absent	Absent	Hypothetical protein
HP_0412	32	Absent	Absent	Hypothetical protein
HP_0450	44	Absent	HP79_04472	Hypothetical protein
HP_0504	49	Absent	Absent	Hypothetical protein
HP_0756	48	Absent	Absent	Hypothetical protein
HP_0881	31	Absent	HP79_02604	Hypothetical protein
HP_0945	98	Absent	HP79_07570	Hypothetical protein
HP_1033	131	Absent	HP79_05066	Hypothetical protein
HP_1176	34	Absent	Absent	Hypothetical protein
HP_1194	28	Absent	Absent	Hypothetical protein
HP_1239	29	Absent	Absent	Hypothetical protein
HP_1381	77	Absent	HP79_05939	Hypothetical protein
HP_1405	34	HP17_08079	Absent	Hypothetical protein
HP_1427	60	Absent	Absent	Histidine-rich metal-binding polypeptide
HP_1432	72	Absent	Absent	Histidine and glutamine-rich protein
HP_0007	23	Absent	Absent	Hypothetical protein
HP_0008	27	Absent	HP79_05151	Hypothetical protein
HP_0225	22	HP17_06822	Absent	Hypothetical protein
HP_0359	21	Absent	Absent	Hypothetical protein
HP_0429	12	HP17_08444	Absent	Hypothetical protein
HP_0533	29	Absent	Absent	Hypothetical protein
HP_0560	26	Absent	Absent	Hypothetical protein
HP_0767	24	Absent	Absent	Hypothetical protein
HP_1093	28	Absent	Absent	Hypothetical protein
HP_1500	23	Absent	Absent	Hypothetical protein
HP_1536	18	HP17_07512	Absent	Hypothetical protein

Appendix 14 List of unique genes present in the genome of *H. pylori* J99 when compared to a number of *H. pylori* strains

Unique J99	Size (aa)	17874	P79	Product
jhp0693	48	Absent	Absent	Hypothetical protein
jhp0916	64	HP17_03604	Absent	Hypothetical protein, periplasmic competence-like protein
jhp0952	63	Absent	Absent	Hypothetical protein
jhp0970	32	Absent	Absent	Hypothetical protein
jhp1049	250	Absent	HP79_02664	Hypothetical protein
jhp1320	60	Absent	Absent	Hypothetical protein
jhp1321	77	Absent	Absent	Hypothetical metal-binding protein
jhp1393	26	Absent	HP79_05661	Hypothetical protein

Appendix 15 List of unique genes present in the genome of *H. pylori* P79 when compared to a number of *H. pylori* strains

Unique P79	Size (aa)	26695	J99	17874	Product
HP79_00100	>58	Absent	Absent	HP17_07742	Hypothetical protein
HP79_00125	45	Absent	Absent	Absent	Hypothetical protein
HP79_00130	84	Absent	Absent	HP17_07717	Hypothetical protein
HP79_00662	31	HP_0341	Absent	HP17_01925	Hypothetical protein
HP79_00667	130	HP_0342	Absent	HP17_01935	Hypothetical protein
HP79_00682	111	HP_0345	Absent	HP17_01945	Hypothetical protein
HP79_00913	30	HP_1366	jhp1442	Absent	Hypothetical protein
HP79_01025	179	HP_1516	Absent	HP17_07159	Hypothetical protein
HP79_01040	31	Absent	Absent	HP17_04044	Hypothetical protein
HP79_01110	44	Absent	Absent	Absent	Hypothetical protein
HP79_01550	31	HP_1097	Absent	HP17_02633	Hypothetical protein
HP79_01734	40	Absent	jhp0659	Absent	Hypothetical protein
HP79_02099	39	Absent	Absent	HP17_05630	Hypothetical protein
HP79_02389	75	Absent	Absent	HP17_05900	Cag pathogenicity island protein B
HP79_02594	42	Absent	Absent	Absent	Hypothetical protein
HP79_02604	31	HP_0881	Absent	Absent	Hypothetical protein
HP79_02779	>31	Absent	Absent	Absent	Hypothetical protein
HP79_02931	31	HP_1328	Absent	Absent	Hypothetical protein
HP79_03341	48	Absent	Absent	Absent	Hypothetical protein
HP79_03523	>40	Absent	Absent	HP17_01025	Hypothetical protein
HP79_04072	31	HP_1017	Absent	Absent	Hypothetical protein
HP79_04147	59	Absent	jhp0415	HP17_02469	Hypothetical protein

HP79_04162	44	Absent	Absent	Absent	Hypothetical protein
HP79_04472	44	HP_0450	Absent	Absent	Hypothetical protein
HP79_04602	39	Absent	Absent	Absent	Hypothetical protein
HP79_04607	44	Absent	Absent	Absent	Hypothetical protein
HP79_04617	>36	Absent	jhp0698	Absent	Hypothetical protein
HP79_04737	34	HP_0174	Absent	Absent	Hypothetical protein
HP79_04941	45	HP_0881	Absent	HP17_04471	Hypothetical protein
HP79_05151	27	HP_0008	Absent	Absent	Hypothetical protein
HP79_05221	46	Absent	Absent	HP17_04576	Hypothetical protein
HP79_05251	45	Absent	Absent	Absent	Hypothetical protein
HP79_05516	>38	Absent	Absent	Absent	Hypothetical protein
HP79_05646	39	Absent	Absent	Absent	Hypothetical protein
HP79_05671	37	Absent	Absent	Absent	Hypothetical protein
HP79_05686	33	Absent	Absent	HP17_07269	Hypothetical protein
HP79_05704	>31	Absent	Absent	Absent	Hypothetical protein
HP79_05939	77	HP_1381	Absent	Absent	Hypothetical protein
HP79_06114	36	Absent	Absent	Absent	Hypothetical protein
HP79_06154	40	Absent	Absent	Absent	Hypothetical protein
HP79_06174	30	HP_0024	Absent	HP17_03092	Hypothetical protein
HP79_06236	36	HP_1424	Absent	HP17_03884	Hypothetical protein
HP79_06371	39	Absent	Absent	Absent	Hypothetical protein
HP79_06376	31	Absent	Absent	Absent	Hypothetical protein
HP79_06406	34	Absent	jhp0953	Absent	Hypothetical protein
HP79_06671	37	Absent	Absent	Absent	Hypothetical protein

HP79_06701	44	HP_0237	Absent	HP17_04866	Hypothetical protein
HP79_07373	53	Absent	Absent	Absent	Hypothetical protein
HP79_07418	35	Absent	Absent	Absent	Hypothetical protein
HP79_07453	>27	HP_0461	Absent	Absent	Hypothetical protein
HP79_07770	72	Absent	Absent	HP17_00305	Hypothetical protein
HP79_07835	36	Absent	Absent	HP17_00230	Hypothetical protein
HP79_07890	255	HP_0151	jhp0139	Absent	Hypothetical membrane protein
HP79_07905	35	Absent	Absent	HP17_00165	Hypothetical protein
HP79_08218	31	Absent	jhp0931	Absent	Hypothetical protein
HP79_08228	40	HP_0592	Absent	Absent	Hypothetical protein
HP79_08383	32	Absent	Absent	Absent	Hypothetical protein
HP79_08555	30	Absent	Absent	HP17_04948	Hypothetical protein
HP79_08967	>40	Absent	Absent	HP17_03449	Hypothetical protein
HP79_08977	29	Absent	Absent	Absent	Hypothetical protein
HP79_09132	>73	Absent	Absent	Absent	Hypothetical protein

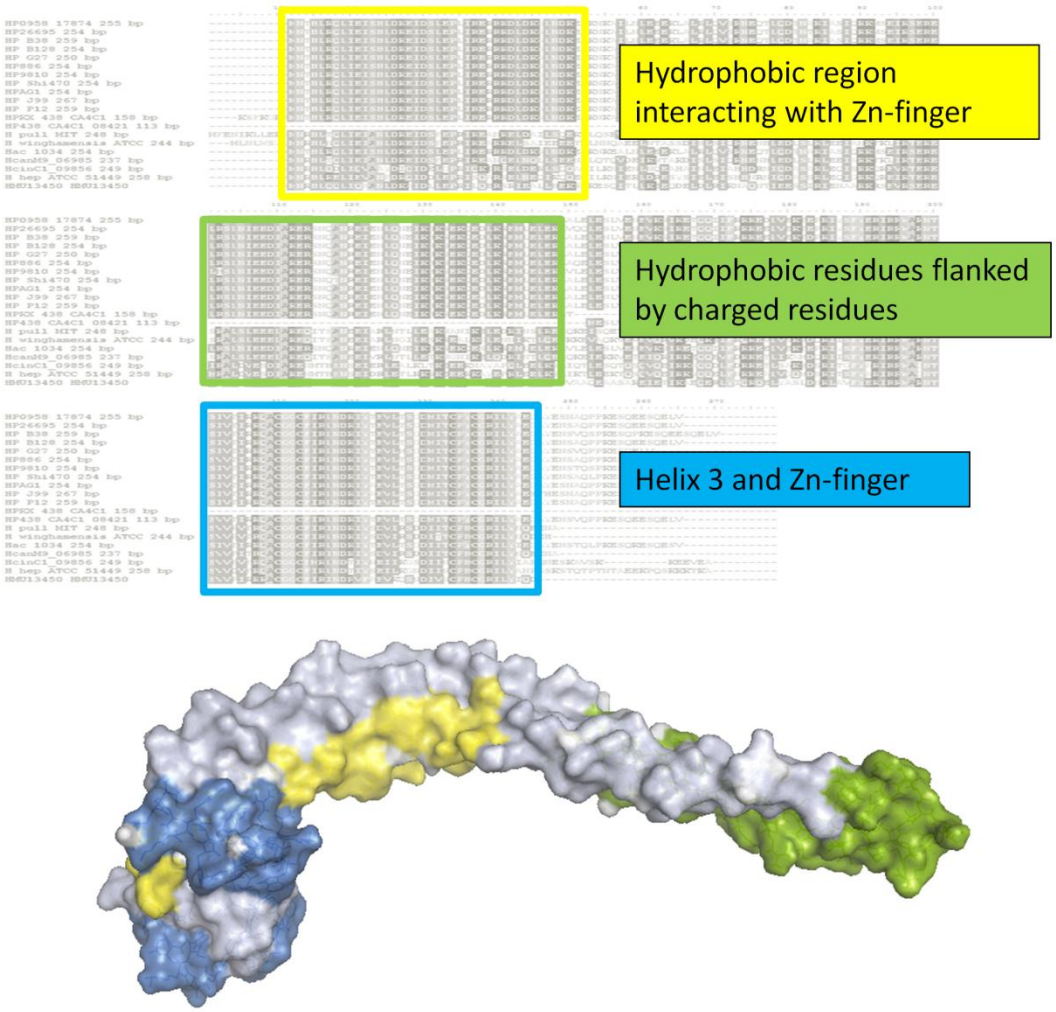
Appendix 16 List of oligonucleotide primers used for cloning and expression

Primer	Sequence (5' - 3')	Source
HP0958-F	CGCCGCGGATCCAACACCCACCTCAAACAATTG	(Caly *et al.*, 2010)
HP0958-R	CCGCCGGAATTCTTACTAAACTAATTCTTGGCTTTCTTCTTG	(Caly *et al.*, 2010)
FLAG_FliH-F	CGCCGCGGATCCGATTATAAAGATGATGATGATAAATCATTGAATAGCCGCAAAAAT	This study
FLAG_FliH 89-F	CGCCGCGGATCCGATTATAAAGATGATGATGATAAAAGCAAAGCTTTGATTGAAAAC	This study
FliH258-R	CCGCCGGAATTCTTACTACACCTTAAAATTTTCCAACAC	This study
pGBKT7 MCS-F	TTCATCGGAAGAGAGTAGTAAC	This study
pGBKT7 MCS-R	AAGAGTCACTTTAAAATTTGTATACAC	This study
pGADT7 MCS-F	CTATTCGATGATGAAGATACCCCACC	This study
pGADT7 MCS-R	AGATGGTGCACGATGCACAGTTG	This study
HP0958_Eco-F	CGCCGCGAATTCAACACCCACCTCAAACAATTG	This study
HP0958_Bam-R	CCGCCGGGATCCTCAAACTAATTCTTGGCTTTCTTC	This study
Mut2_Eco-F	CGCCGCGAATTCAACGCCCACCTCAAACAATTG	This study
Mut3_Eco-F	CGCCGCGAATTCAACACCGCCCTCAAACAATTG	This study
FliH 89-258_Eco-F	CGCCGCGAATTCAGCAAAGCTTTGATTGAAAACGC	This study
FliH 89-258_Bam-R	CCGCCGGGATCCTCACACCTTAAAATTTTCCAACAC	This study
RpoN 74-F	CGCCGCGAATTCATCGCATCTAAAAGCCTTTTTG	This study
RpoN 284-R	CCGCCGGGATCCAAGCATCAGACCGATTTTATAAATC	This study
RpoN 210-R	CCGCCGGGATCCCTCAATGGCTGGGGGGTTTTTAAAGG	This study

Appendix 17 List of primers used for site-directed mutagenesis of *hp0958*

Primer	Sequence (5' - 3')	Source
Mut2-F	CCCTGGGATCCAACGCCCACCTCAAACAATTG	This study
Mut2-R	CAATTGTTTGAGGTGGGCGTTGGATCCCAGGG	This study
Mut3-F	CCCCTGGGATCCAACACCGCCCTCAAACAATTG	This study
Mut3-R	CAATTGTTTGAGGGCGGTGTTGGATCCCAGGGG	This study
Mut4-F	GATTGAAATTTCGGCTTTGGATAAAGAAATTGACTCCTTAGAGCCG	This study
Mut4-R	CGGCTCTAAGGAGTCAATTTCTTTATCCAAAGCCGAAATTTCAATC	This study
Mut6-F	GGAAAAATTAGCCCTAAAAGCACAGGTTTCTAAAAACGAGCAAACCC	This study
Mut6-R	GGGTTTGCTCGTTTTTAGAAACCTGTGCTTTTAGGGCTAATTTTTCC	This study
Mut9-F	CGAGCCTAAAATCTATAGCTTTTATGAAGAGATCAGAAGATGGGCG	This study
Mut9-R	CGCCCATCTTCTGATCTCTTCATAAAAGCTATAGATTTTAGGCTCG	This study
Mut10-F	GCTTTTATGAAAGGATCAGAGAATGGGCGAAAAACACGAGC	This study
Mut10-R	GCTCGTGTTTTTCGCCCATTCTCTGATCCTTTCATAAAAGC	This study
Mut11-F	GGATCAGAAGATGGGCGGCAAACACGAGCATTGTAACG	This study
Mut11-R	CGTTACAATGCTCGTGTTTGCCGCCCATCTTCTGATCC	This study
Mut12-F	CGAGCATTGTAACGATCGCAAAACAGGCTTGTGGGGG	This study
Mut12-R	CCCCCACAAGCCTGTTTTGCGATCGTTACAATGCTCG	This study
Mut13-F	CAGGCTTGTGGGGGTTGTGTTATTAGACTAAATGATAAG	This study
Mut13-R	CTTATCATTTAGTCTAATAACACAACCCCCACAAGCCTG	This study
Mut17-F	GGTTGTTTTATTAGACTAAATGATGAGATTTATACTGAAGTGCTAACG	This study
Mut17-R	CGTTAGCACTTCAGTATAAATCTCATCATTTAGTCTAATAAAACAACC	This study
Mut18-F	GGGGATATGATCGCGTGCCCGTATTGCGGGCG	This study

Mut18-R	CGCCCGCAATACGGGCACGCGATCATATCCCC	This study
Mut19-F	GGGCGTATTTTAGCCGCTGAGGGCGCGTATGAAAGTAACGC	This study
Mut19-R	GCGTTACTTTCATACGCGCCCTCAGCGGCTAAAATACGCCC	This study



Appendix 18 Selection of targets for site-directed mutagenesis of HP0958.

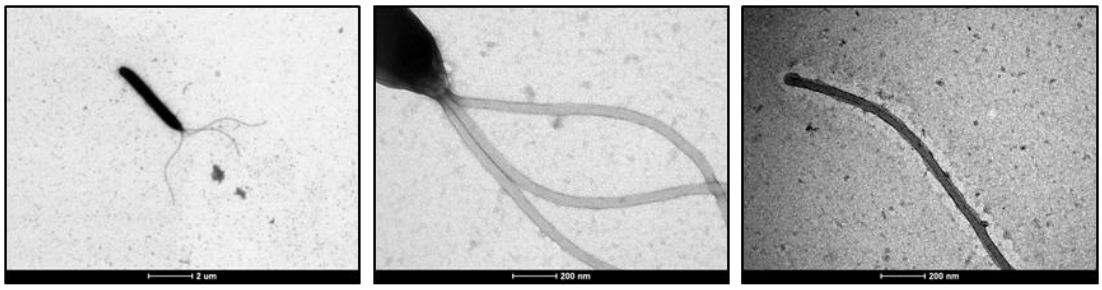
Rationale behind selection of residues for site-directed mutagenesis. Multiple alignment of amino acid sequences of HP0958 homologues identified 3 highly conserved regions: an N-terminal hydrophobic region which forms part of the coiled-coil in close contact with the Zn-ribbon (coloured yellow); hydrophobic residues in the coiled-coil flanked by charged residues (coloured green) and the Zn-ribbon which is associated with nucleic acid interactions (coloured blue). Residues which were conserved, solvent exposed and not critical in maintaining the structure of HP0958 were selected for mutagenesis. Image of HP0958 structure was generated using Pymol (DeLano Scientific, CA).

Appendix 19 List of oligonucleotide primers used for qPCR

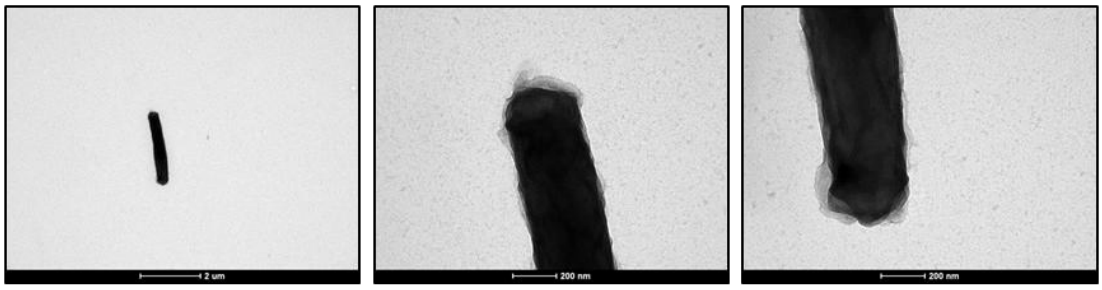
Primer	Sequence (5' - 3')	Comments	Source
qflaA-F	CCGATAGTGTCAGTAATGGGC	Forward primer for real time PCR of *flaA*	This study
qflaA-R	GATTCCCAAAACCAATCGCTGTG	Reverse primer for real time PCR of *flaA*	This study
qflgE-F	GGCTAACGAGCGTGGATAAG	Forward primer for real time PCR of *flgE*	(Douillard *et al.*, 2008)
qflgE-R	GAGCGAGCGCTAAAGTCCTA	Reverse primer for real time PCR of *flgE*	(Douillard *et al.*, 2008)
qflaB-F	ACCAGAACCGACGCTAGAGA	Forward primer for real time PCR of *flaB*	(Douillard *et al.*, 2008)
qflaB-R	CCACATTCGCATCAAAAATG	Reverse primer for real time PCR of *flaB*	(Douillard *et al.*, 2008)
qropN-F	AGCACGATTTCAAGGGCCAT	Forward primer for real time PCR of *rpoN*	This study
qropN-R	CACAGCGTTTGAAGTCTCGC	Reverse primer for real time PCR of *rpoN*	This study
qfliA-F	GAATGCCCAAAGGAATTCAA	Forward primer for real time PCR of *fliA*	(Douillard *et al.*, 2008)
qfliA-R	AGCGAGATCGTCTTGATGGT	Reverse primer for real time PCR of *fliA*	(Douillard *et al.*, 2008)
qera-F	AAGGCTAATGCGACCAGAAA	Forward primer for real time PCR of *era*	(Douillard *et al.*, 2008)
qera-R	GGAGCCCTGGTGTGTCTAAA	Reverse primer for real time PCR of *era*	(Douillard *et al.*, 2008)

Appendix 20 List of *flaA* oligonucleotide primers

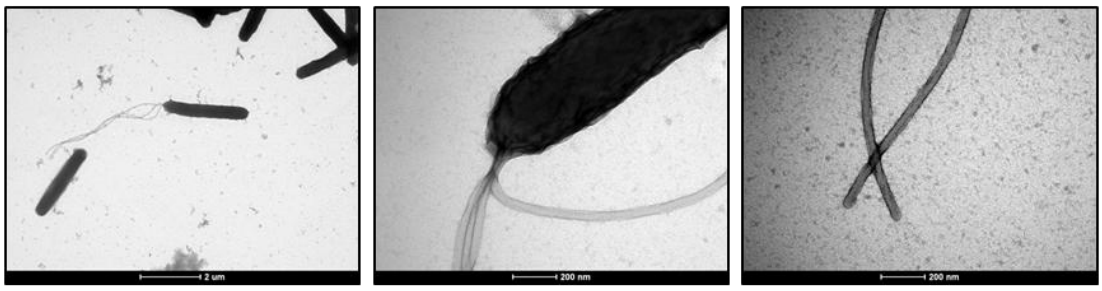
Primer	Sequence (5' - 3')	Source
FlaFL-F	TGTAATACGACTCACTATAGGTCCAACCAAAAGCAAGGATG	(Douillard *et al.*, 2008)
FlaFL-R	AGCCCCATACAAACACCTTTCTTAAAA	(Douillard *et al.*, 2008)
Reg1.1-F	TGTAATACGACTCACTATAGGTTCAACCAAAAGCAAGGATGCC	This study
Reg1.1-R	GACATTAGCGTTAAAATTCCCATAAGATTTTCAACTGCTCATCCATAGC	This study
Reg1.2-F	TATGGATGAGCAGTTGAAATCTTATGGGAATTTTAACGCTAATGTC	This study
Reg1.2-R	GCCAACGCTTAAAGCGTTAGCC	This study
Reg2-F	TGTAATACGACTCACTATAGGTAAGGTTAAAGCGACTCAAGC	This study
Reg2-R	CCATTTAAGGTTAAATTACTCAAACTTCC	This study
Reg3-F	TGTAATACGACTCACTATAGGATTGGTTGCAGCGATCAATGCG	This study
Reg3-R	ACATCGCGCAAATTCACCGTG	This study



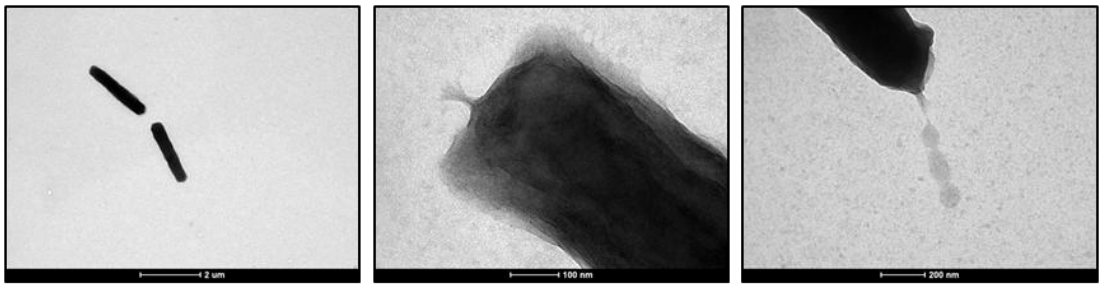
P79 WT



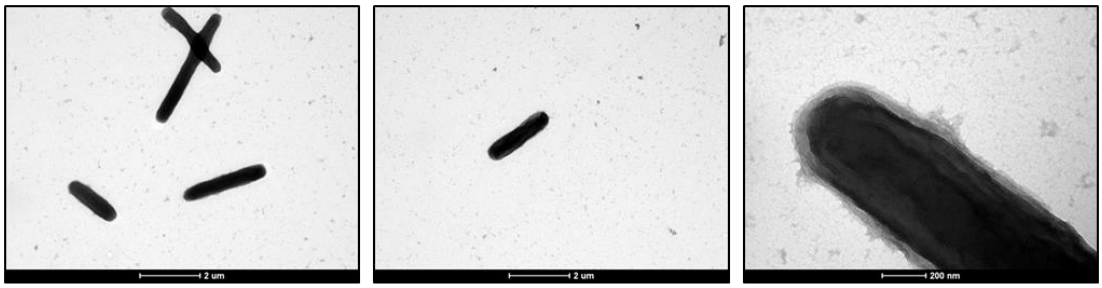
P79-0958KO



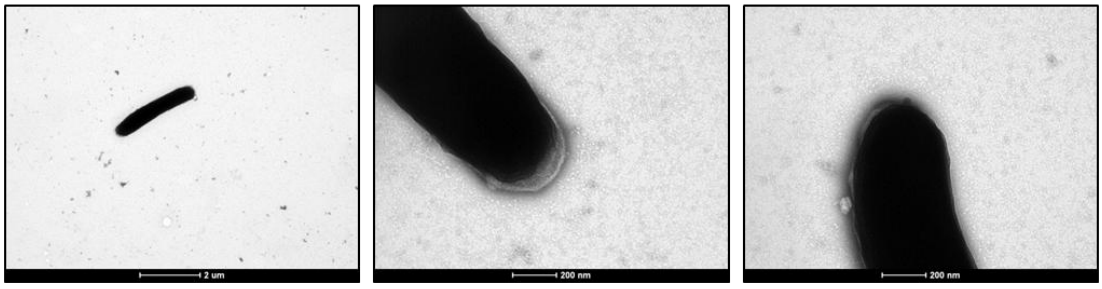
P79-0958KO_WT



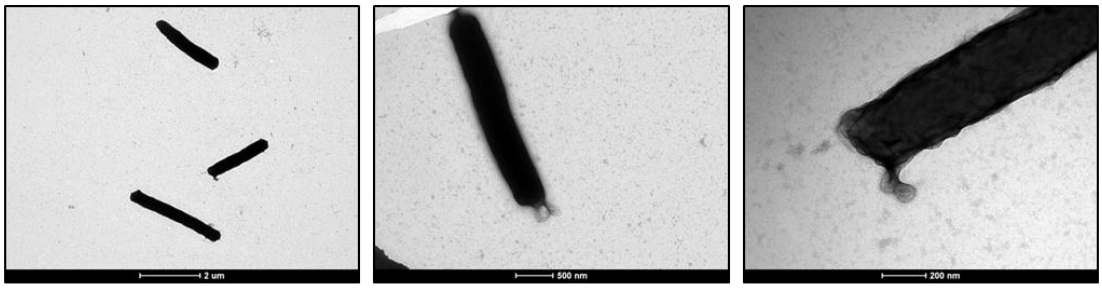
P79-0958KO_T3A



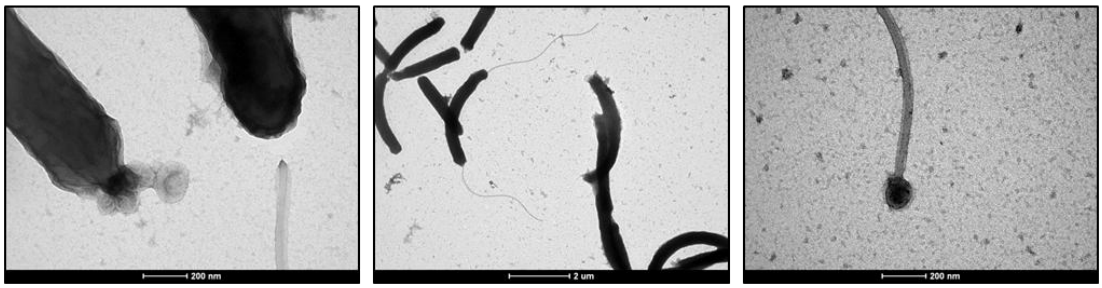
P79-0958KO_H4A



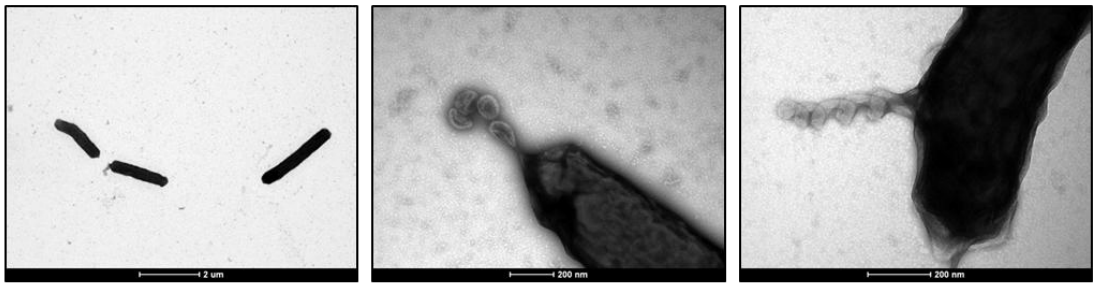
P79-0958KO_H13A



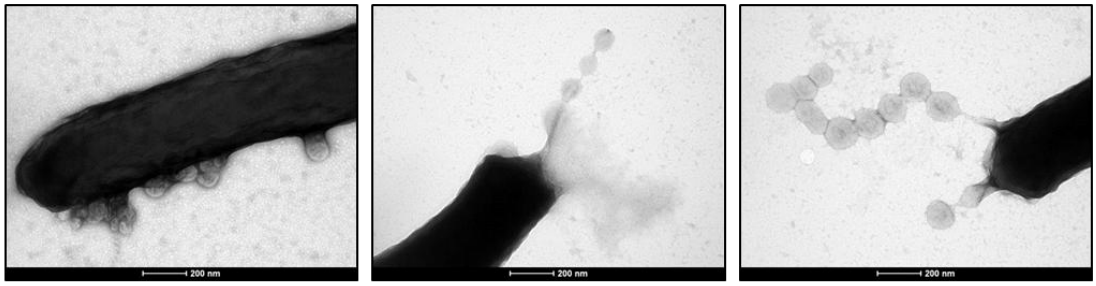
P79-0958KO_L47A



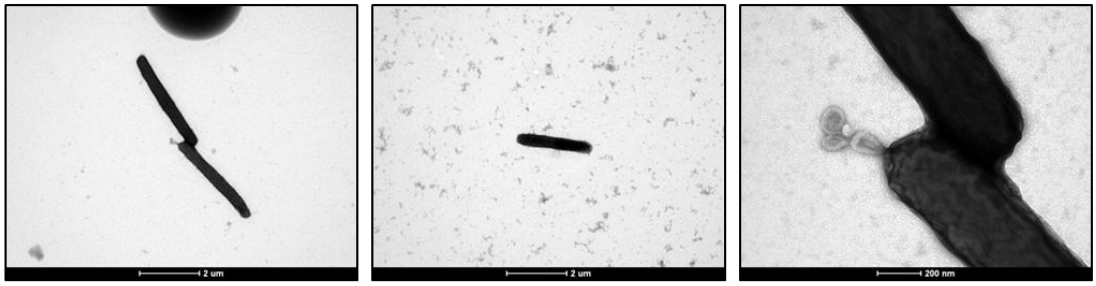
P79-0958KO_L58A



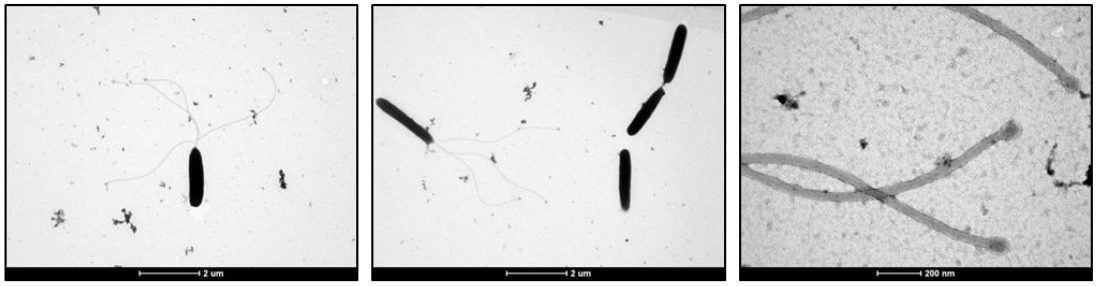
P79-0958KO_I99A



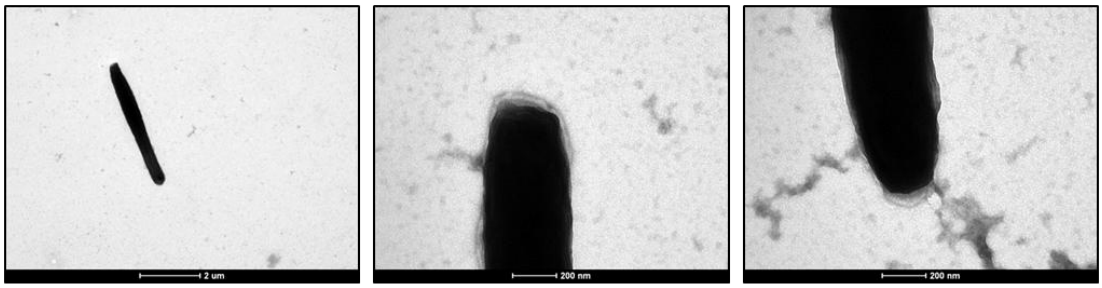
P79-0958KO_F161A



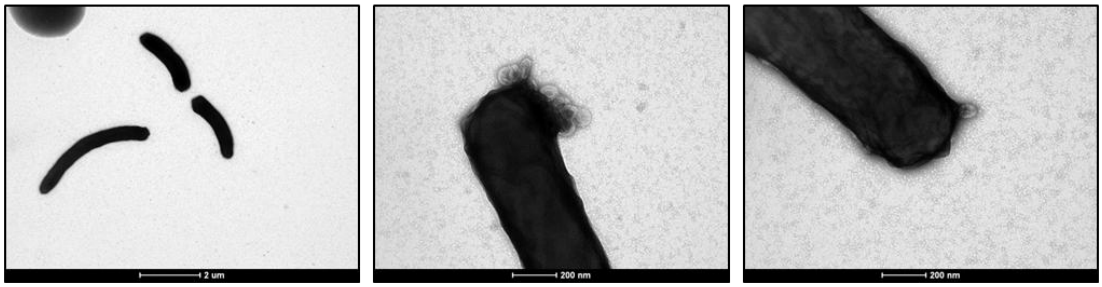
P79-0958KO_R181E



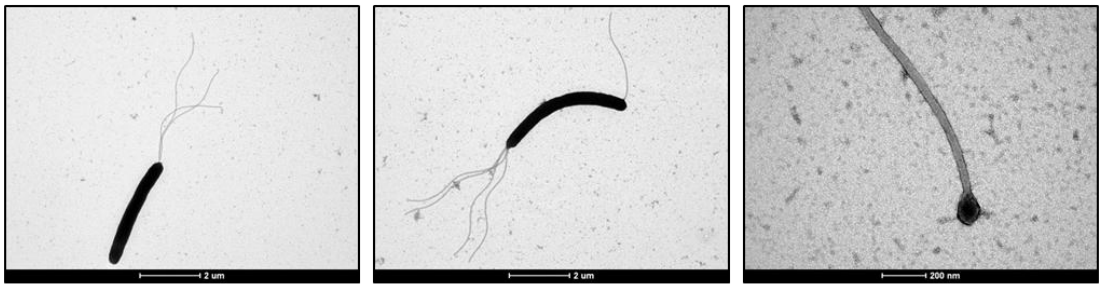
P79-0958KO_R184E



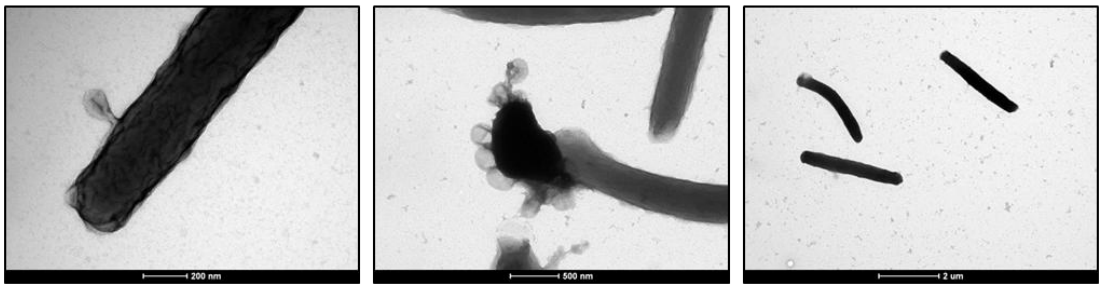
P79-0958KO_K187A



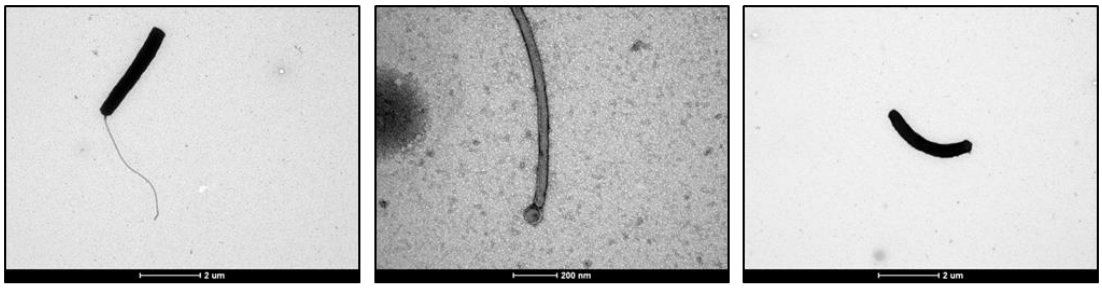
P79-0958KO_K195A



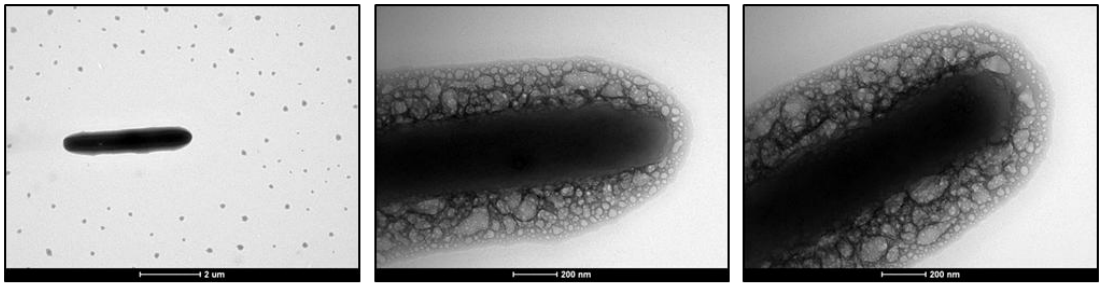
P79-0958KO_F203V



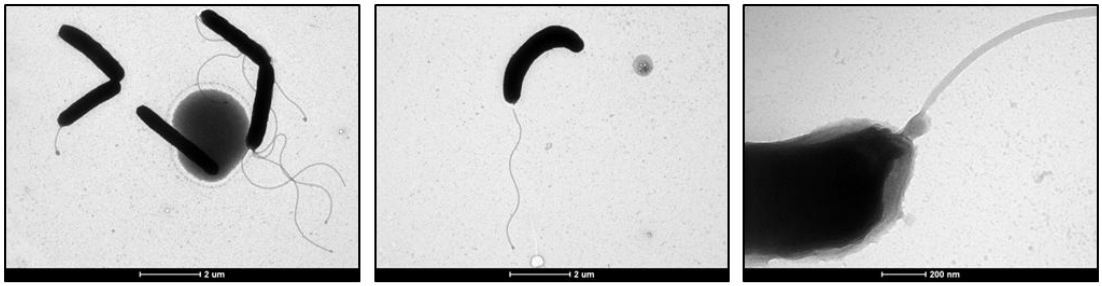
P79-0958KO_I204A



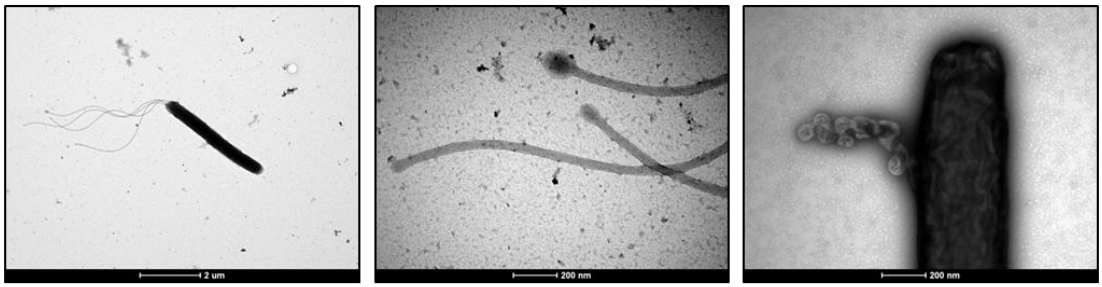
P79-0958KO_R205A



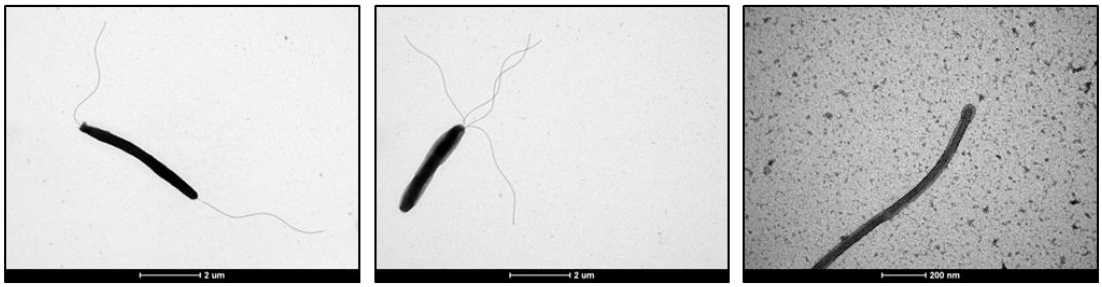
P79-0958KO_R205V



P79-0958KO_K209E



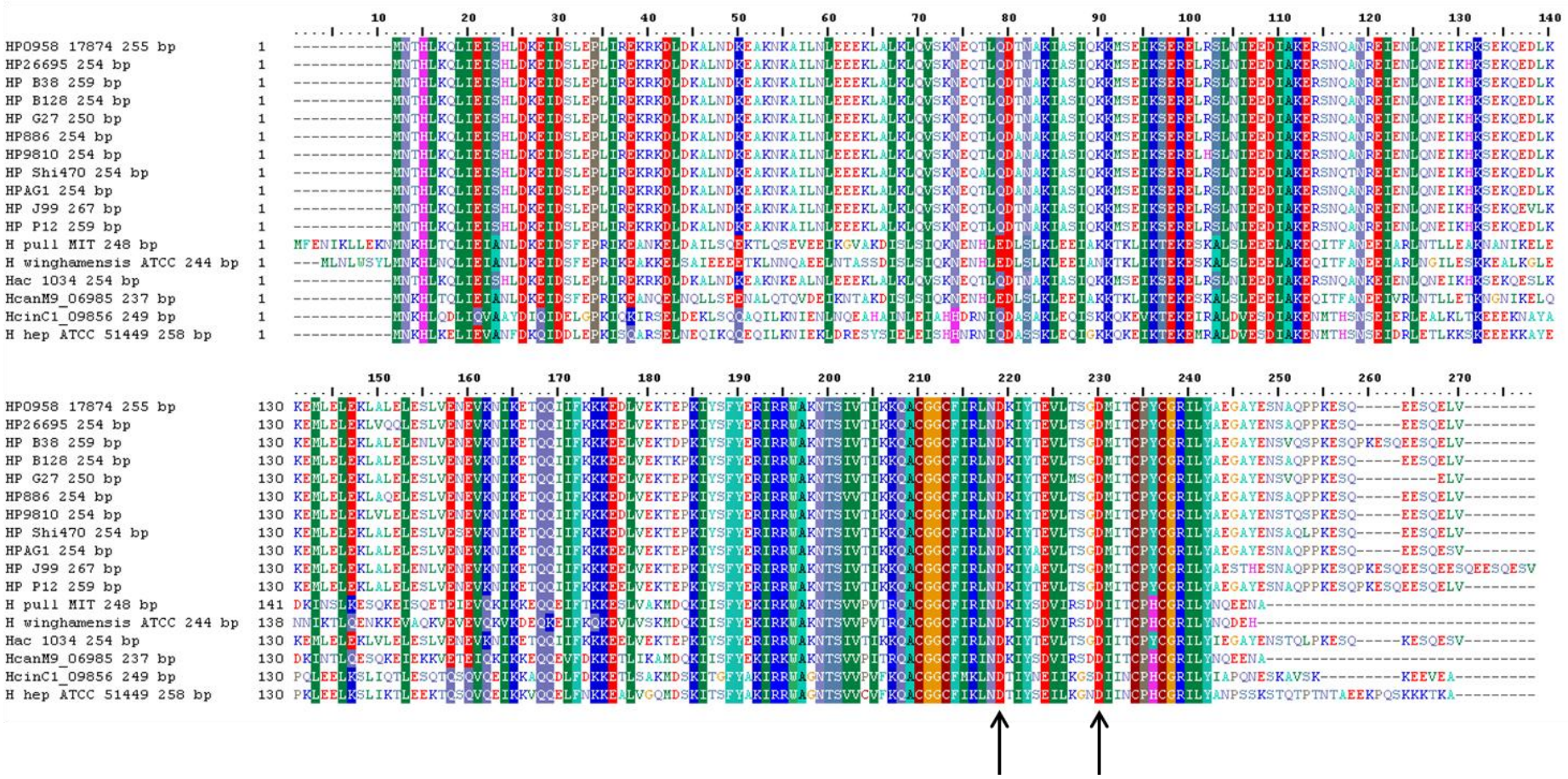
P79-0958KO_T222A



P79-0958KO_Y231F

Appendix 21 EM analysis of flagellum production by *H. pylori* P79 and derivatives.

Transmission electron micrographs of *H. pylori* cells stained with uranyl acetate; 3 images per strain.



Appendix 22 ClustalW multiple sequence alignment of HP0958 and orthologues in the genus *Helicobacter*.

Multiple alignment of the amino acid sequences of HP0958 homologues in the genus *Helicobacter* where arrows indicate conserved aspartic acid residues (D208 and D219 of strain 17874) in the zinc-finger.

JB Journals.ASM.org

GENOME ANNOUNCEMENT

Draft Genome Sequences of *Helicobacter pylori* Strains 17874 and P79

Ceara D. Clancy,[a] Brian M. Forde,[a] Stanley A. Moore,[b] and Paul W. O'Toole[a,c]

Department of Microbiology, University College Cork, Cork, Ireland[a]; Department of Biochemistry, University of Saskatchewan, Saskatoon, Saskatchewan, Canada[b]; and Alimentary Pharmabiotic Centre, University College Cork, Cork, Ireland[c]

***Helicobacter pylori* is a human pathogen that colonizes the human gastric mucosa, causing gastritis, duodenal and gastric ulcers, and gastric carcinoma. Here we announce the draft genomes of *H. pylori* strain 17874, commonly used for studying motility, and P79, a strain for which plasmid vectors have been developed.**

Helicobacter pylori genomes sequenced to date exhibit significant variation. *H. pylori* CCUG 17874 was originally isolated from the gastric antrum of a patient in Perth, Australia, and is the type strain for the species (4) and is often used for flagellum biogenesis studies. P79 is a derivative of strain P1, transformed with 17874 chromosomal DNA to generate a streptomycin-resistant mutant (3). This readily transformable strain facilitates *in vivo* studies of *H. pylori*. The genomes of these strains were sequenced to provide a clearer genomic platform for investigation of *H. pylori* motility.

The *H. pylori* 17874 and P79 genomes were sequenced at the Beijing Genomics Institute (BGI) on the Illumina HiSeq platform, generating a paired-end library containing 20,154,284 and 13,298,804 reads of 90 bp, respectively. In a reference-guided assembly strategy using MIRA (version 3.2.1), reads for both genomes were mapped to the genomes of *H. pylori* 26695 (GenBank accession no. NC_000915) (5) and J99 (NC_000921.1) (1). A *de novo* assembly using Velvet was also performed and aligned to the MIRA assembly to close gaps. Strain 17874 and P79 contigs were assembled into 80 and 48 scaffolds, respectively. Protein coding regions were identified using the NCBI Prokaryotic Genome Automated Annotation Pipeline (PGAAP) and manually curated, with particular interest in flagellum-related genes. Predicted coding regions were identified with a minimum cutoff size of 30 amino acids.

H. pylori 17874 and P79 have genome sizes of 1,615,763 bp and 1,641,495 bp, respectively, and GC contents of 38.97% and 38.86%, respectively. Both strains are positive for *cagA* and *vacA*, well-described virulence factors (2). Strain-unique genes were identified using a pairwise bidirectional BLASTP comparison, where the query sequence has no detectable homologues. The 17874 genome contains 1,639 open reading frames, with 35, 45, and 24 unique genes that are absent in 26695, J99, and P79, respectively. Sixteen genes from 26695 and 6 genes from J99 are absent in strain 17874. *H. pylori* P79 contains 1,699 open reading frames, with 40, 52, and 36 unique genes that are absent in 26695, J99, and 17874, respectively. Twelve genes from 26695 and 6 genes from J99 are absent in P79. Twenty-one genes are unique to the 17874 and P79 genomes compared across these four strains.

The majority of strain-unique genes identified encode hypothetical protein products. Of note, strain 17874 possesses a unique type II restriction enzyme, and P79 possesses a unique hypothetical membrane protein that is absent in 26695 and J99. Strains 17874 and P79 lack metal-binding proteins present in both 26695 and J99 but possess Cag island protein B. All major flagellar and outer membrane proteins are present and intact in both 17874 and P79 compared to 26695 and J99. A hypothetical protein with predicted involvement in ATPase activity during flagellum biogenesis is absent in P79.

Nucleotide sequence accession numbers. The draft genome sequence of *H. pylori* 17874 has been deposited in GenBank, available through BioProject accession no. PRJNA76569 and project identification (ID) no. 76569. Similarly, the draft sequence of P79 is available in GenBank through BioProject accession no. PRJNA76567 and project ID no. 76567.

ACKNOWLEDGMENTS

This project was supported by a Research Frontiers Programme award (09_RFP_GEN2443) from Science Foundation Ireland to P.W.O. and by an Embark scholarship from IRCSET to C.D.C.

REFERENCES

1. **Alm RA, et al.** 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. Nature **397**:176–180.
2. **Basso D, Plebani M, Kusters JG.** 2010. Pathogenesis of *Helicobacter pylori* infection. Helicobacter **15**:14–20.
3. **Heuermann D, Haas R.** 1998. A stable shuttle vector system for efficient genetic complementation of *Helicobacter pylori* strains by transformation and conjugation. Mol. Gen. Genet. **257**:519–528.
4. **Marshall BJ, Royce H, Annear D.** 1984. Original isolation of *Campylobacter pyloris* from human gastric mucosa. Microbios Lett. **25**:83–88.
5. **Tomb J-F, et al.** 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature **388**:539–547.

Received 14 February 2012 Accepted 20 February 2012
Address correspondence to Paul W. O'Toole, pwotoole@ucc.ie.

doi:10.1128/JB.00230-12

Proceedings of the 10th International Workshop on Pathogenesis and Host Response in Helicobacter Infections
Helsingör, Denmark, July 4 – 7th, 2012.

Investigation of the interaction node of the *Helicobacter pylori* flagellum biogenesis protein HP0958

C. D. Clancy[1], S. A. Moore[2] and P.W. O'Toole[1]

[1]Department of Microbiology, University College Cork , Ireland.
[2]Department of Biochemistry, University of Saskatchewan, Saskatoon SK, CanadaS7N 5E5.
This project was supported by a Research Frontiers Programme award (09_RFP_GEN2443) from Science Foundation Ireland to P.W.O. and by an Embark scholarship from IRCSET to C.D.C.

Motility is an essential feature of *Helicobacter pylori* infection. *H. pylori* flagellum biogenesis differs from the well-characterised model organism *E. coli* by lacking the master regulator, $FlhD_4C_2$. A yeast two hybrid study investigating the proteome of *H. pylori* 26695 previously identified that HP0958 interacts with the flagellar proteins FliH and RpoN (σ^{54}). We hypothesised that HP0958 may have a regulatory role in flagellum construction, possibly in the switch between expression of Class II and Class III flagellar genes. The nature of the interaction between HP0958 and FliH (full length and 89-258) was investigated to expand upon the yeast two hybrid data. However, pull-down assay failed to identify an interaction between HP0958 and FliH. Yeast two hybrid was performed with HP0958 and FliH 89-258, but this also indicated that there is no detectable interaction between these proteins. Additionally, the HP0958/RpoN interaction was confirmed as a relatively weak interaction by yeast two hybrid analysis. The HP0958/FliH interaction appears to be a false positive within a large scale analysis identifying over 1,200 interactions.

Proceedings of the 10th International Workshop on Pathogenesis and Host Response in Helicobacter Infections
Helsingör, Denmark, July 4 – 7th, 2012.

Comparative genomic analysis of the *Helicobacter pylori* strains 17874 and P79.

C. D. Clancy[1], B. M. Forde[1], S. A. Moore[2] and P.W. O'Toole[1]

[1]Department of Microbiology, University College Cork , Ireland.
[2]Department of Biochemistry, University of Saskatchewan, Saskatoon SK, CanadaS7N 5E5.
This project was supported by a Research Frontiers Programme award (09_RFP_GEN2443) from Science Foundation Ireland to P.W.O. and by an Embark scholarship from IRCSET to C.D.C.

Helicobacter pylori was the first species to have more than one genome sequenced. To date, there are 46 genome sequences of *H. pylori* strains available. Earlier this year, we published the draft genome sequences of CCUG 17874 and P79. *H. pylori* 17874 is the type strain for the species, often used in motility studies. P79 is a readily transformable derivative of the strain P1 and hence is useful for bacterial motility studies. Initial inspection of the sequences of 17874 and P79 revealed that the major flagellar and outer membrane proteins are conserved when compared to 26695 and J99. Core genome analysis of 43 sequenced genomes allowed a more conservative estimation of the core genome of the species than previously estimated. Phylogenetic analysis of these 43 strains revealed the evolutionary relationship of 17874 and P79 to the other sequenced strains.

Proceedings of The Society for General Microbiology Irish Division Autumn Meeting: Gut Microbes
University College Dublin, Ireland, 21 – 22nd March, 2013.

Protein-protein interaction analysis of the *Helicobacter pylori* flagellum biogenesis protein HP0958

C. D. Clancy[1], S. A. Moore[2] and P.W. O'Toole[1]

[1]Department of Microbiology, University College Cork , Ireland.
[2]Department of Biochemistry, University of Saskatchewan, Saskatoon SK, CanadaS7N 5E5.
This project was supported by a Research Frontiers Programme award (09_RFP_GEN2443) from Science Foundation Ireland to P.W.O. and by an Embark scholarship from IRCSET to C.D.C.

Helicobacter pylori is a gastric pathogen which currently infects approximately 50% of the global population. Infection can lead to the development of gastric and duodenal ulcers, gastric cancer and MALT lymphoma. Motility is an essential feature of *Helicobacter pylori* infection. *H. pylori* flagellum biogenesis differs from the well-characterised model organism *E. coli* by lacking the master regulator, $FlhD_4C_2$. A yeast two-hybrid study investigating the proteome of *H. pylori* 26695 previously identified that HP0958 interacts with the flagellar proteins FliH and RpoN (σ^{54}). We hypothesised that HP0958 may have a regulatory role in flagellum construction, possibly in the switch between expression of Class II and Class III flagellar genes.
The nature of the interaction between HP0958 and FliH (full length and residues 89-258) was investigated to expand upon the yeast two hybrid data. However, pull-down assay failed to identify an interaction between HP0958 and FliH. Yeast two-hybrid was performed with HP0958 and FliH 89-258, but this also failed to detect interaction between these proteins. The HP0958/RpoN interaction was confirmed as a relatively weak interaction by yeast two-hybrid analysis and PXG assay. The HP0958/FliH interaction appears to be a false positive within a large scale analysis identifying over 1,200 interactions. The C-terminus of RpoN is essential for the interaction with HP0958. A panel of site-directed HP0958 mutants were generated to further investigate the nature of the HP0958/RpoN interaction.

Proceedings of the 17th International Workshop on Campylobacter, Helicobacter and Related Organisms
University of Aberdeen, Scotland, 15 – 19th September, 2013.

Investigation of the interaction node of *Helicobacter pylori* flagellar biogenesis protein HP0958.

C. D. Clancy[1], S. A. Moore[2] and P.W. O'Toole[1]

[1]Department of Microbiology, University College Cork , Ireland.
[2]Department of Biochemistry, University of Saskatchewan, Saskatoon SK, CanadaS7N 5E5.
This project was supported by a Research Frontiers Programme award (09_RFP_GEN2443) from Science Foundation Ireland to P.W.O. and by an Embark scholarship from IRCSET to C.D.C.

Background: Motility is an important feature of *Helicobacter pylori* infection. HP0958 is a flagellar biosynthesis protein which is essential for motility. HP0958 stabilises RpoN (σ^{54}), the sigma factor controlling expression of the Class II flagellar genes. HP0958 also interacts with the *flaA* mRNA transcript, encoding the major flagellin protein, FlaA. The crystal structure of HP0958 revealed two structural domains: an N-terminal anti-parallel, α-helical coiled-coil and a C-terminal Zn-finger domain. This structural data has provided information that has informed our design of mutations to test interactions with protein and mRNA.
Materials and Methods: Site-directed mutagenesis of HP0958 was performed to identify potential residues involved in interactions with the *flaA* mRNA transcript and RpoN. The HP0958/*flaA* mRNA interaction was investigated by electrophoretic mobility shift assay (EMSAs). The HP0958/RpoN interaction was investigated by yeast two-hybrid assay followed by enzyme assay. A panel of *hp0958* mutants were re-introduced into a *hp0958-null* mutant strain of P79 by homologous recombination and effects on expression of Class II and Class III flagellar genes were monitored by western blot.
Results: A panel of HP0958 mutants were generated based on their potential role in protein-protein/protein-RNA interactions. A number of candidates have been identified as involved in the interaction of HP0958 with RpoN and the *flaA* mRNA transcript.
Impact of research: Construction of the bacterial flagellum is a complex, hierarchical process involving over 40 proteins; regulation in *Helicobacter* differs from the well-studied model organisms, *E. coli* and *S. enterica* serovar Typhimurium. Understanding the mechanism by which HP0958 contributes to this complex process will improve our understanding of these differences.

Acknowledgements

There are many people I would like to thank for their help and support throughout the undertaking of this PhD. First and foremost, I want to thank my supervisor, Prof. Paul O'Toole. Thank you for giving me the opportunity to do my PhD in your lab, my second home, for these past 4 years. I really appreciate all your help and advice over the years, and especially the inspirational one-liners that come to you so easily; these really helped motivate me to rise above the difficulties I encountered during this process.

A special thanks to Prof. Stanley Moore for all your help throughout the PhD. Your words of encouragement came at a time when I needed them the most and so for this, I thank you also. To my predecessors on the *"Hp*" team, Dr. Delphine Caly and Dr. Francois Douillard, I am so grateful for all the hours spent discussing my project and the intricacies of this bug. Most of all I am thankful for your availability no matter the hour (or the distance!) and your friendship. To Dr. Brian Forde, thank you for your help with the genome assembly, you truely have unbreakable patience! Thanks to Dr. Ian Jeffery and Hugh Harris for your bioinformatics-related help.

Thank you to the technicians and staff who ensured the smooth running of the Department for the past 4 years with a special mention to Margaret, Liam, Paddy, Máire and Dan who never saw me stuck for anything.

A massive thank you to all the guys in the Microbiology Dept. (especially new doctors Stephen and Stuart) with whom I shared this experience, it would have been a much harder slog without the company. Also to my science family (past and present) in labs 339/438 and BSI 4.21... I will miss the craic, the Friday TB conventions and Daybreak brunches! A special thanks to Anne, Rachel, James, Eliette, Emma and Marcus for taking me in as the "newbie", teaching me loads and creating such a friendly atmosphere to work in. Thank you Jen for your friendship and unfaltering optimism, I'm so glad I had you for a right arm! Thank you (and congratulations!) to Dr. Susan Power, my left arm, for sharing the highs and lows of the daily PhD grind. You were always there at the drop of a hat to help in my hour(s) of need or supply welcome distractions to prevent meltdown! To Ludovica, thanks for your friendship and help; I wish you the best of luck for your viva and beyond. Thank you to Fabien for being so generous with your time and helpful in all things science. To Huizi, Cian and Agnieszka, thanks for the laughs! Denise, thank you for always making the tea! ;) To Angela, Fergal, Adam, Guy, Alex, Paul, Guillaume,

Burkhardt and Jillian, cheers for the past year and best of luck with all your future endeavours.

A special thank you goes to Dr. Jakki Cooney, my original mentor. Thank you for your time and support both during my time at UL and throughout my PhD. Most of all, thank you for telling me that I could do this long before I believed you.

Turning to my much neglected family and friends... this thesis is dedicated to my parents, who have always gone above and beyond to support and encourage me in anything I've ever wanted. Thank you Marie and Jer for everything, I could never have completed this if it wasn't for you both. Thanks for taking care of all the furry friends at home who cheered me up when I needed it! Thanks to Glic for hugs like no other. Thanks to Brían and Cathal for never asking about my PhD come to think about it! Special thanks to Neil for your words of encouragement and insight; also, thanks for your help in sorting out the microscope camera. Thanks Caitríona for your interest and advice!

To my friends and most of all Tomás, thanks for waiting for me all those times when I was glued to my pipette! Tomás, if it wasn't for you I would never have taken any breaks or had a square meal in Cork. Thanks for making me laugh and being my concert buddy! Thanks to Roseanne, Muireann, Ashling, Niamh, Clare, Charlie and Claire for being true friends and listening to me talk non-stop about my PhD, I don't know what I'll talk about now...

Finally, thank you Pearl Jam and Sigur Rós for keeping me company on those late nights and weekends in the lab, and for being the soundtrack to my thesis writing!