SPECIFIC *HELICOBACTER PYLORI* VIRULENCE AND HOST GENETIC SUSCEPTIBILITY FACTORS: THE POTENTIAL ROLE IN GASTRODUODENAL DISEASES

HUSSEIN ALI OSMAN

UNIVERSITI SAINS MALAYSIA

2015

SPECIFIC *HELICOBACTER PYLORI* VIRULENCE AND HOST GENETIC SUSCEPTIBILITY FACTORS: THE POTENTIAL ROLE IN GASTRODUODENAL DISEASES

BY

HUSSEIN ALI OSMAN

Thesis submitted in fulfilment of the requirements

for the Degree of

Doctor of Philosophy (PhD)

June 2015

ACKNOWLEDGEMENTS

First, thanks and a special gratitude are due to “ALLAH” for His entire blessing during the pursuit of my academic career goals. Throughout the course of my PhD, I have been blessed with the support and assistance of many special people. I would like to take the opportunity to acknowledge and thank them for being a part of this memorable academic journey

I would like to express my deepest appreciation and utmost gratitude to my supervisor, **Prof. Dr. Zilfalil Bin Alwi** for his continuous advice, guidance and encouragement throughout my study. Special thanks also goes to him for introducing me to the Genome wide association study, a novel part of my study indeed. I would like to express my special thanks and gratitude to my co-supervisor, **Prof. Dr. Habsah Hasan** for her endless support and advice during challenging times of my study. My special thanks to my co-supervisor, **Dr. Rapeah Suppian** for her valuable contribution, guidance and continuous encouragement during this period.

I would like to extend my deepest gratitude to Islamic development Bank (IDB), Jeddah, Saudi Arabia for awarding me PhD scholarship. My appreciation also goes to Umma University for allowing me to pursue my study. I would like to thank USM for supporting this project by Research University grant 1001 / PPSP / 812108, short term grant 304/PPSP/61312082 and USM as anchor for the Malaysian Node of Human Variome Project number 1002/PPSP/910343 . Without this grant this project would have not been possible. My sincere appreciation goes to endoscopy staff of Hospital Universiti Sains Malaysia (HUSM) and Hospital Kuala Lumpur (HKL) for their support. Many thanks and appreciations are due to all the lecturers, staff and

postgraduate students at the department of medical microbiology and parasitology for their support.

I would like to thank all participants who took part in this study. A special thanks goes to Siti Nurain our Research assistant for her assistance and patience during my study. Sincere thanks are extended to my special friends and colleagues Dr. Omar Salad Elmi, Khalid Mohamed Ali, Abdelrahman, Nik Zuraina, Hatin, Fazreen, Sathia and Dr. Yoko for their continuous assistance during my study period.

Finally my sincere appreciations are also due to the members of my family, my sisters, brothers and my wife, Dr. Saida Ali Ibrahim, my lovely daughters Hanaan and Rayyaan for their continuous support and patience while i was away.

TABLES OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AGCC	Affymetrix genechip command console software
AlpAB	Adherence–Associated Lipoprotein
ARHGAP26	Rho GTPase activating protein 26
ATCC	American Type Culture Collection
AW1	Wash buffer 1
AW 2	Wash buffer 2
BabA	Blood Group Antigen Adhesin
BANK1	B-cell scaffold protein with ankyrin repeats 1
BHI	Brain Heart Infusion
BLAST	Basic Local Alignment Search Tool
bp	Basepair
BSA	Bovine serum albumin
^{o}C	Celsius gradient
CagA	Cytotoxin-Associated Gene A protein
*cag*PAI	*cag* Pathogenicity Island
CD-CV	Common disease-common variant
CDH13	cadherin 13
CM	CagA Multimerization
CNV	Copy number variation
CO2	Carbon dioxide
13C	Carbon-13
14C	Carbon-14
Csk	Carboxy-Terminal Src Kinase
DDBJ	DNA Database of Japan
DNA	Deoxyribose nucleic acid
dNTPs	Deoxynucleotid Triphosphate
DRAM1	DNA-damage regulated autophagy modulator 1
DSB	Double strand breaks
dsDNA	Double strand DNA
DU	Duodenal Ulcer
dupA	Duodenal Ulcer Promoting Gene A
E-CM	East Asian- CagA Multimerization
EDTA	Ethylene Diamine Tetra Acetic Acid
EMBL	European Molecular Biology Laboratory
EPIYA motif	Glu-Pro-Ile-Tyr-Ala Motif
EtOH	Ethanol
GERD	Gastro-esophageal reflux disease
GC	Gastric Cancer
GTC	Genotype console software
GU	Gastric Ulcer
HWE	Hardy Weinberg Equilibrium
GWAS	Genome wide association study
H. pylori	*Helicobacter pylori*
HopZ	*H. pylori* Outer Membrane Protein Z
IARC	International Agency for Research in Cancer
iceA	Induced by Contact to Epithelium
MAF	Minor allele frequency

MALToma	Mucosa Associated Lymphoid Tissue Lymphoma
MAST4	microtubule associated serine/threonine kinase family member 4
min	Minute
mg	miligram
ml	Mililitre
µl	Microliter
µg	Microgram
µM	Micromole
MLST	Multi Locus Sequence Typing
MMEJ	Microhomology-mediated end joining
MMR	Mismatch Repair System
MPCR	Multiplex PCR
n	Number
RUT	Rapid Urease Test

KEVIRULENAN KHUSUS *HELICOBACTER PYLORI* DAN FAKTOR KERENTANAN GENETIK PERUMAH: POTENSI PERANAN TERHADAP PENYAKIT GASTRODUODENAL

ABSTRAK

Helicobacter pylori (H. pylori) adalah salah satu patogen manusia yang paling lazim dan memberi kesan terhadap 50% daripada populasi manusia. *H. pylori* dikaitkan dengan penyakit gastrik, ulser peptik, kanser gastrik dan limfoma berkaitan dengan tisu limfoid mukosa gastrik. Tindak balas beberapa faktor seperti persekitaran, kevirulenan bakteria dan genetik hos dipercayai boleh menentukan tahap kemudaratan dan kesan selepas jangkitan *H. pylori*. Tujuan kajian ini adalah untuk menentukan distribusi gen-gen virulen *H. pylori* (*cagA, babA2*, *SabA* dan *dupA*) dan korelasinya dengan hasil klinikal. Kajian ini juga bertujuan menilai corak *H. pylori cagA* motif EPIYA; EPIYA-A, -B, -C atau -D di antara kumpulan etnik dan perkaitannya dengan penyakit gastroduodenal. Kajian ini turut mengenalpasti kehadiran SNP sebagai varian genetik di dalam genom perumah yang mungkin berkait dengan kerentanan atau pertahanan tarhadap jangkitan *H. pylori*. Ini adalah kajian keratan rentas dan kajian kes kawalan di antara Mei 2012 sehingga Jun 2014 dalam kalangan pesakit dispeptik dan berlainan kaum (Melayu, India dan Cina) di Unit Endoskopi Hospital Universiti Sains Malaysia dan Hospital Kuala Lumpur. Genotyping genom perumah dilakukan dengan mengunakan teknik PCR dan Affymetrix SNP microarray 6.0. Kajian ini merangkumi dua fasa; dalam fasa pertama, sejumlah 105 pesakit yang disahkan positif terhadap jangkitan *H. pylori* telah terlibat dalam kajian ini. Purata umur dan SD adalah 54.48+12.94 tahun dengan julat umur di antara 26 sehingga 86 tahun. Lima puluh tujuh (54.3%) pesakit yang dijangkiti adalah lelaki manakala empat puluh lapan (45.7%) adalah wanita. Berdasarkan penemuan endoskopik, 78 pesakit mengalami gastritis, sembilan gastrik ulser, lima ulser duodenal dan 13 normal. Penemuan gen-gen *H. pylori cagA, babA2*, *sabA*2 dan *dupA* dalam pesakit dispeptik *H. pylori* masing-masing adalah 69.5%, 41.0%, 43.8% dan 22.9%. Gen *cagA* dikesan dengan lebih tinggi dalam kalangan bangsa India (39.7%), *babA2* lebih lazim bagi Melayu (39.5%) manakala *dupA* adalah tertinggi bagi kalangan bangsa India dan Melayu dengan kadar yang sama (37.5%). Bangsa Cina mempunyai kelaziman paling rendah terhadap keempat-

empat gen. Majoriti pesakit Cina dijangkiti dengan *cagA* jenis A-B-D strain Asia Timur (88.9%) manakala *cagA* jenis A-B-C strain Barat (82.8%) dikesan lebih tinggi dalam kalangan bangsa India. Bangsa Melayu mempunyai strain bercampur. Terdapat perkaitan yang signifikan secara statistik (**p<0.001**) di antara etnik dan *cagA* motif EPIYA, walaupun tiada perbezaan signifikan di antara gen virulen *H. pylori* dan jenis EPIYA dengan hasil klinikal. Dalam fasa kedua, sejumlah 80 (42 *H. pylori* positif) dan 38 (*H. pylori* negatif) pesakit generasi ketiga dengan purata umur 49.87+12.335 (umur di antara 20-75 tahun). Kajian ini menunjukkan bahawa SNP rs3770521 (**P=1.33 X10^{-5}**) gen XRCC5, rs7042986 (**P=0.0001**) gen SMARCA2, dan rs10860808 (**P=0.0002**) gen DRAM1 adalah SNP yang cenderung kepada jangkitan *H. pylori* dalam kalangan pesakit gastrik berbangsa India, Melayu dan Cina. Kajian ini turut mengenalpasti dua SNP yang protektif, iaitu rs1809758 (**P=9.85X 10^{-6}**) gen BANK1 dan rs3776349 (**P=0.0001**) gen ARHGAP26 dalam kalangan pesakit gastrik India dan Melayu. Kesimpulannya, gen yang rendah dan variasi dalam kumpulan etnik yang berlainan menunjukkan strain bakteria tersebut bergantung kepada perbezaan etnik dan geografi. Kajian ini juga menunjukkan bahawa tiada perbezaan yang signifikan di antara gen virulen dengan hasil klinikal. Kajian ini turut membuktikan bahawa EPIYA A-B-D dan A-B-C adalah predominan dalam bangsa Cina dan India manakala bangsa Melayu mempunyai strain campuran. Akhir sekali, kajian GWAS terkini menunjukkan lima SNP unggul yang boleh dikaitkan dengan kecenderungan dan ketahanan terhadap *H. pylori* gastritis dalam ketiga-tiga kumpulan etnik ini.

SPECIFIC *HELICOBACTER PYLORI* VIRULENCE AND HOST GENETIC SUSCEPTIBILITY FACTORS: THE POTENTIAL ROLE IN GASTRODUODENAL DISEASES

ABSTRACT

Helicobacter pylori (H. pylori) is one of the most common human pathogens and affects over 50% of the world population. *H. pylori* is associated with gastritis, peptic ulcer, gastric cancer and gastric mucosa-associated lymphoid tissue lymphoma. The interaction of several factors like environmental, bacterial virulence and host genetic are believed to determine the severity and final outcome after *H. pylori* infection. The aim of this study was to determine the distribution of *H. pylori* virulence genes (*cagA, babA2, SabA*and *dupA*) and its correlation with clinical outcomes. This study also assessed the pattern of *H. pylori cagA* EPIYA motifs, EPIYA-A, -B, -C, or -D among different ethnic groups and its association with gastroduodenal disease. The current study also explored the presence of SNPs as genetic variants in the host genome which may be associated with susceptibility or protection to *H. pylori* infection. This was a cross-sectional and case-control study conducted between May 2012 to June 2014 among dyspeptic patients of different ethnicities (Malay, Indian and Chinese) at the Endoscopy Unit of Hospital Universiti Sains Malaysia and Hospital Kuala Lumpur. Genotyping of bacterial and host genome was performed using PCR and Affymetrix SNP 6.0 microarray. This study consists of 2 phases; in phase 1, a total of 105 patients who were confirmed positive to have *H. pylori* infection were recruited into the study. The mean age and SD were 54.48 ±12.94 years and age range of 26 to 86 years old. Fifty seven (54.3%) of the infected patients were males while forty eight (45.7%) were females. Based on the endoscopic findings, 78 patients had gastritis, nine gastric ulcer, five duodenal ulcer and 13 normal. The prevalence of *H. pylori cagA, babA2, sabA* and *dupA* genes in *H. pylori* dyspeptic patients were 69.5%, 41.0%, 43.8% and 22.9% respectively. *cagA* is more common in Indians (39.7%), *babA2* is common in Malays (39.5%) and *dupA* detection is more in Indian and Malay at the same rate (37.5%). The Chinese have

the lowest prevalence of the four genes. Majority of Chinese patients were predominantly infected with *cagA* type A-B-D East Asian strain (88.9%) while *cagA* type A-B-C Western strain (82.8%) was predominantly detected in the Indians while the Malays have mixed strain. There were statistically significant difference (***P*<0.001**) between ethnicity and *cagA* EPIYA motifs, although we could not find significant difference between *H. pylori* virulence genes and EPIYA types and clinical outcomes. In phase II, a total of 80 (42 *H. pylori* positive and 38 *H. pylori* negative) third generation patients with a mean age of 49.87 ± 12.335 years (age range 20-75 years) were recruited. The present study identified SNPs rs3770521 (***P*=1.33 x 10^{-5}**) of XRCC5 gene, rs7042986 of SMARCA2 (***P*=0.0001**) and rs10860808 (***P*=0.0002**) of DRAM1 gene as the susceptible SNPs to *H. pylori* infection among the Indian, Malay and Chinese gastritis patients respectively. This study also identified two protective SNPs rs1809578 (***P*=9.85x 10^{-6}**) of gene BANK1 and rs3776349 (***P*=0.0001**) of gene ARHGAP26 among *H. pylori* the Indian and Malay gastritis patients respectively. In conclusion, the lower prevalence of virulence genes and variations among the different ethnic groups suggest that the bacterial strains are geographically and ethnically dependent. No significant difference was observed between virulence genes and clinical outcome. This study also shows that EPIYA A-B-D and A-B-C are predominant in the Chinese and Indians respectively, while the Malays have mixed strain. Finally, the current GWAS study revealed five novel SNPs that may be associated with susceptibility and protection of *H. pylori* gastritis in the three ethnic groups.

Chapter 1: Introduction

1.1 *H. pylori*

Marshall and Warren, in 1982 were the first to isolate and culture the spiral bacterium, now known as *H. pylori*, from the gastric mucosa of humans (Marshall and Warren, 1984). In their first study of more than 100 antral biopsy samples, they observed this bacterium to be present by histology in 58 subjects undergoing endoscopic examination and were able to isolate and culture it from 11 biopsy specimens. Based on these results, Marshall and Warren suggested that there was an etiological relationship between *H. pylori* and these gastric diseases (Marshall and Warren, 1984). Therefore in 2005 as recognition of the value of their discovery to the medical world, Barry Marshall and Robin Warren were awarded the Nobel Prize in Medicine.

1.2 Historical aspects

Prior to the discovery by Marshall and Warren, other scientists had also observed this bacteria to be present in the gastric mucosa of both humans and animals. The first of such report came from Bottcher in 1874 in which he described the bacteria to be present on the ulcer floor and in the mucosal margins of ulcers. In 1938, Doenges found spiral organisms in 103 (43%) of 242 stomachs examined at autopsy. In 1954, however, Palmer reported no evidence of spiral organisms in 1180 gastric mucosal biopsies from 1000 patients; he suggested that the organism noted previously in normal stomachs represented postmortem processes and that the

bacterial source was the oral cavity. In 1975, Steer and Colin-Jones reported the presence of Gram-negative bacteria in the gastric mucosa of approximately 80% of patients with gastric ulcers (Ha, 2007).

Due to the low pH of gastric acid, the human stomach was long believed to be a sterile environment. Although some isolated reports had been made regarding the existence of bacteria in the stomach, since the 1800's, it was not until 1982 that the major breakthrough in the understanding of gastric and duodenal pathology such as gastritis, peptic ulcer and gastric cancer occurred when Robin Warren and Barry Marshall described "unidentified curved bacilli on gastric epithelium in active chronic gastritis (Robin Warren and Marshall, 1983).

1.3 Taxonomy

H. pylori was originally considered as a Campylobacter-like organism (CLO) and was named *Campylobacter pyloridis* (Marshall and Warren, 1984), based on bacterial similarity with other Campylobacter species. In 1987 this name was corrected for grammatical reasons to *Campylobacter pylori* (Marshall and Goodwin, 1987). However, based on differences in the 16S rRNA gene sequences, fatty acid profiles and flagella morphology, this bacterium was placed in new genus called *Helicobacter* with the new name for this bacterium being *H. pylori*. The name "*Helicobacter*" was based on the helical shape of the bacterium and the word "pylori" was used because the bacterium was commonly isolated from the pylorus of the stomach (Goodwin *et al.*, 1989).

1.4 Microbiological features and growth requirements

Helicobacter belongs to the family Helicobacteraceae, order Campylobacterales and the Epsilonproteobacteria class (Owen, 1998). *H. pylori* is a Gram negative spiral rod with 2 – 4 μm in length and 0.5 – 1μm wide. The bacterium can appear as rod, while coccodi shapes appear after prolonged in *vitro* culture or antibiotic treatment (Kusters *et al.*, 1997). It has 2 to 6 unipolar, sheathed flagella of nearly 3 μm in lengths, which frequently carry a unique bulb at the end. The main function of flagella is to allow quick movement in viscous solutions such as the mucus layer overlying the gastric epithelial cells (O'Toole *et al.*, 2000).

H. pylori is a slow-growing microaerophiles, fastidious microorganism and require complex growth media. It grows optimally at 37°C on a rich medium containing blood or serum. These supplements may act as additional sources of nutrients and possibly also protect against the toxic effects. Media mostly used for routine isolation and culture of *H. pylori* consist of Columbia or Brucella agar supplemented with either (lysed) horse or sheep blood or, alternatively, new-born or fetal calf serum. For isolation, selective antibiotic mixtures are available (Owen, 1998; Ndip *et al.*, 2003). Dent supplement consists of vancomycin, trimethoprim, cefsoludin, and amphotericin B, whereas Skirrow supplement consists of vancomycin, trimethoprim, polymyxin B, and amphotericin B. Liquid media usually consist of Brucella, Mueller-Hinton, or brain heart infusion broth supplemented with 2 to 10% calf serum or 0.2 to 1.0% -cyclodextrins, often together with either Dent or Skirrow's supplement (Ndip *et al.*, 2003). *H. pylori* can be identified by Gram staining, oxidase, catalase, and urease test (Maaroos *et al.*, 2004).

Currently, at least 32 *Helicobacter* species have been identified in humans and animals (http://www.bacterio.cict.fr/h/helicobacter.html). These are generally divided into two groups based on the niche they colonise: 'gastric' Helicobacters primarily colonise the stomach whilst 'enterohepatic' Helicobacters colonise the intestine and hepatobiliary system.

1.5 Epidemiology

1.5.1 Prevalence

H. pylori infects more than half of the world's population (Ryan *et al.*, 2001; Amjad *et al.*, 2010). The prevalence was around 25% in developed countries to more than 90% in developing areas, but not all infected individuals eventually developed the disease (Figure 1.1) (van Doorn *et al.*, 2000; Ribeiro *et al.*, 2003). The prevalence of *H. pylori* infection varies widely by geographic area, age, race, and socioeconomic status (Brown, 2000). The prevalence varies from one country to another for example in Australia, it is as low as 15.5% (Pandeya and Whiteman, 2011), while in India, as high as 87% (Miwa *et al.*, 2002). Among East Asian countries, the overall seroprevalence rate was 59.6% in South Korea (Yim *et al.*, 2007), 58.07% in China (Wang and Wang, 2003), 54.5% in Taiwan and 39.3% in Japan (Fock, 2014). Among Southeast Asian countries, the reported sero prevalence rate was 35.9% in Malaysia (Goh and Parasakthi, 2001), 31% in Singapore (Fock, 2014) and 57% in Thailand (Deankanob *et al.*, 2006).

H. pylori prevalence varies considerably between the three primary ethnic groups resident in Malaysia, Malays, Chinese and Indians with prevalence rates of 8-43.3%, 45-60.6% and 68.9-75% respectively (Ramelah *et al.*, 2005; Tan *et al.*, 2005). *H. pylori* acquisition is declining in developed countries at a faster rate as compared to developing countries, because of the improvement in hygiene practices in the developed world (Brown, 2000). In 1994, on the basis of various epidemiological studies, *H. pylori* was classified as a class I carcinogen in humans by a working group of the World Health Organization International Agency for Research on Cancer (Yamazaki *et al.*, 2005b).

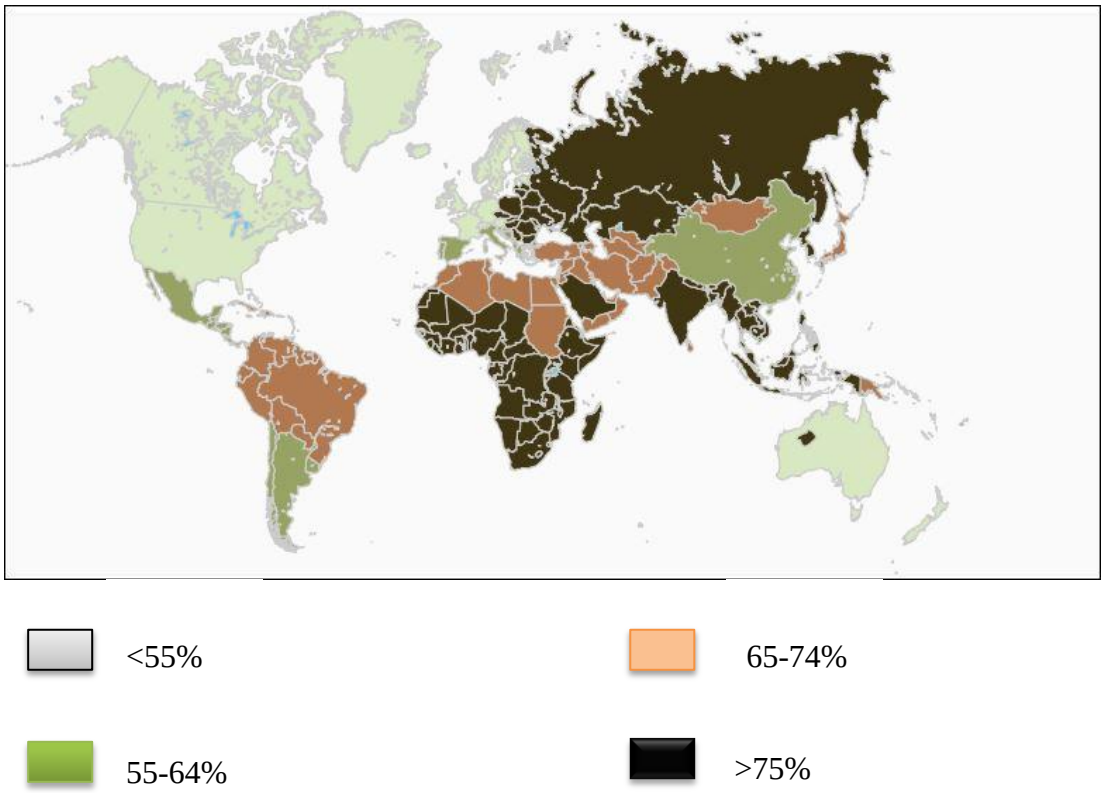

Figure 1.1 World-wide distribution of *H. pylori* infection adapted from (Bauer and Meyer, 2011)

1.5.2 Transmission and source of infection

The exact mechanisms whereby *H. pylori* is acquired are largely unknown. *H. pylori* is almost always acquired in childhood. The only known reservoir for *H. pylori* is the human stomach. Person to person spread appears to be the most likely mode of transmission (Quinn *et al.*, 2003) Transmission has been documented by vomitus, saliva, or faeces and possibly also through water source in the developing world (Brown *et al.*, 2002).

Convincing evidence of intrafamilial transmission, particularly parent to child, has been provided by various studies showing an intrafamilial clustering of infection (Kivi *et al.*, 2003; Perez-Perez *et al.*, 2004; Schwarz *et al.*, 2008; Nahar *et al.*, 2009). Studies have found that iatrogenic transmission occurs through use of a variety of inadequately disinfected gastric devices, endoscopes, and endoscopic accessories. Adequate sterilization and disinfection of endoscopes has been found to reduce the incidence of transmission (Muhammad *et al.*, 2012).

Faeco-oral transmission and oral-oral transmission of bacteria have been reported. Contaminated water supplies in developing countries may serve as an environmental source of bacteria. Children who regularly swim in rivers, streams, pools, drink stream water, or eat uncooked vegetables have greater chance of infection (Muhammad *et al.*, 2012). In the absence of treatment, infection is usually lifelong. The major risk factor for infection is poor socio-economic conditions in childhood (Malaty and Graham, 1994), overcrowding and ethnic and genetic predisposition (Das and Paul, 2007).

1.5.3 *H. pylori* as a tool for tracking human migration

H. pylori strains from different geographic areas demonstrate phylogeographical differentiation; therefore, the genotypes of *H. pylori* strains can serve as markers of the migration of human populations. Thus, the genotypes of *H. pylori* virulence factors, *cagA* and *vacA*, as well as multilocus sequence typing (MLST), are widely used markers of genomic diversity in *H. pylori* populations. The *cagA* virulence factor has two types: the East Asian and the Western types (Yamaoka *et al.*, 2000b; Yamaoka, 2009). *H. pylori* infection has rapidly declined because personal hygiene and quality of life have improved. The molecular epidemiology of *H. pylori* infection is highly informative and should be investigated before this characteristic is completely lost.

The MLST of seven housekeeping genes has been obtained from several *H. pylori* strains isolated from different geographical and ethnic origins; MLST results have shown that *H. pylori* has followed human migration from Africa (Falush *et al.*, 2003). For instance, six *H. pylori* populations, including hpAfrica1, hpAfrica2, hpNEAfrica, hpEurope, hpEastAsia, and hpAsia2, have been identified, indicating *H. pylori* has migrated with its host from Africa (Falush *et al.*, 2003). hpEastAsia comprises hspAmerInd, hspMaori, and hspAsia; hpEurope includes strains from Europe, Turkey, Bangladesh, Ladakha (India), Sudan, and Israel. Interestingly, isolates from Europe form a heterogeneous population, in which the modern hpEurope population is a combination of two ancestral European populations that likely settled in Europe in different waves (Falush *et al.*, 2003). A recent study done in Malaysian Malay, Chinese and Indian showed variations in population. hpEastAsia has been isolated from Chinese patients; furthermore, hpAsia2 isolates

have been obtained from Indians. Isolates from Malay patients comprise a mixed group; these isolates are similar to hspIndia subpopulation (Falush *et al.*, 2003; Tay *et al.*, 2009). The low prevalence of infection and the variation of Malay isolates imply that Malays were originally free of *H. pylori* but recently acquired the pathogen from other subpopulations, mainly from Indians (Tay *et al.*, 2009). It has been established that *H. pylori cagA* EPIYA patterns have a significant geographic variability and closely follow patterns of historical human migrations. EPIYA D is a characteristic Asian EPIYA pattern that virtually does not occur in the Western *H. pylori* strains (EPIYA C) (Queiroz *et al.*, 2010).

1.6 Pathogenicity and virulence factors of *H. pylori*

H. pylori is a Gram-negative spiral bacterium that colonizes and persists in human gastric mucosa. *H. pylori* infects more than half of the world's population, and it is implicated as a causative agent of gastritis, peptic ulcer disease, carcinoma, and mucosa-associated lymphoid tissue lymphoma (Ryan *et al.*, 2001; Bindayna *et al.*, 2006). It is one of the most common bacterial infections in humans (Blaser, 1997). *H. pylori* strains have been divided into two broad families, type I and type II, which are based on whether or not they possess the *vacA* and *cagA* genes. Type I strains have the ability to produce *VacA* and *CagA*, while type II strains lack that ability. Type I strains are regarded as having greater pathogenicity and potential to cause development of disease (Yamazaki *et al.*, 2005b). *H. pylori* possesses a number of virulence factors that allow it to colonize the harsh environment of the stomach, with many of these factors also implicated in its pathogenesis. Among these are urease, flagella, *cagA*, *cagA* EPIYA motifs, *vacA,dupA, babA and sabA.*

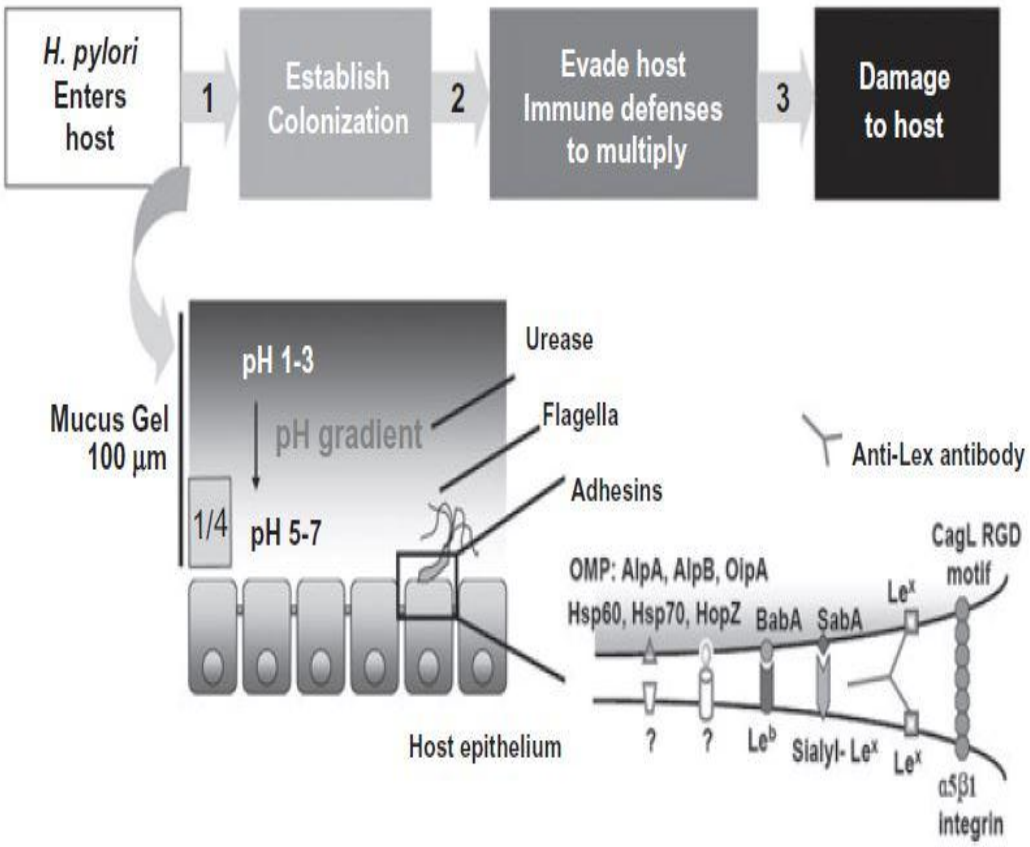

Figure 1.2 Schematic diagram of colonization as the first step in *H. pylori* infection adapted from (Sheu *et al.*, 2010).

Figure 1.2 shows the urease activity and flagella of *H. pylori* facilitate its orientation to the lower one-fourth of the mucus gel above the epithelium. Several putative adhesion molecules, including *babA*, *SabA*, Lewis antigen, and other outer membrane proteins (OMPs) are ready to adhere to known or unknown counterpart receptors on the host the epithelium. (Sheu *et al.*, 2010).

1.6.1 Urease

One of the prominent features of *H. pylori* its ability to colonize the acidic gastric environment, although the bacterium is not an acidophile it overcomes the acidic conditions by production of urease enzyme (Kusters *et al.*, 2006). *H. pylori* survives at a pH range between 4.0 and 8.0 in the absence of urea. However, in the presence of urea the organism can survive at a pH as low as 2.5 (Dunne *et al.*, 2014).

In the stomach the *H. pylori* urease enzyme converts urea into ammonia and carbon dioxide. The ammonia helps to neutralize the acidic environment around the bacterium and enable *H. pylori* to survive and multiply in the stomach (Andersen, 2007).

1.6.2 Flagella

Motility is essential for *H. pylori* colonization. Flagellar motility is thought to be required for the initial stages of infection allowing the bacterium to move from acidic environment of the gastric lumen into less acidic mucus layer (Thompson *et al.*, 2003). *H. pylori* carries 5-7 sheathed flagella that perform the motility required for colonization and infection. The flagella filament consists of two subunits. Both genes coding for these flagellins are necessary for full motility of *H. pylori*. The sheathing of the flagella is believed to protect them from the acidic environment in the stomach.(Andersen, 2007).

1.6.3 Cytotoxin-associated gene A (*cagA*)

CagA is the most widely studied *H. pylori* virulence factor and is present in many but not all *H. pylori* strain (Covacci *et al.,* 1993). The *cagA* gene is located at one end of the cag pathogenicity island (PAI) that codes a type IV secretion system (T4SS) linked with increased secretion of IL-8, a very strong pro-inflammatory chemokine that participates in the gastritis induced by *H. pylori* infection. The T4SS is also liable for the entrance of *cagA* protein into the gastric epithelial cells. After the delivery, *cagA* protein is quickly tyrosine phosphorylated on specific tyrosine residues within repeating Glu-Pro-Ile-Tyr-Ala (EPIYA) motifs and interacts with various target molecules (Naito *et al.,* 2006). Studies have shown that knockout mutants of TFSS genes could not undertake *cagA* transfer and as a result *cagA* tyrosine phosphorylation did not occur. This means that for a functional *cagA* translocation, the TFSS genes must be intact (Stein *et al.,* 2000).

The phosphorylated EPIYA motifs of *cagA* interact and activate the Src homology 2 (SH2) domain-containing tyrosine phosphatase (SHP-2) of the host cells. The *cagA*-SHP-2 complex leads to change in proliferation, morphogenesis and motility of gastric epithelial cells and hence inducing the "hummingbird" phenotype (cell elongation). As SHP-2 plays an important role in both cell growth and cell motility, deregulation of SHP-2 by *cagA* may be involved in the induction of abnormal proliferation and movement of gastric epithelial cells, a cellular condition eventually leading to gastritis and gastric carcinoma (Backert *et al.,* 2001; Higashi *et al.,* 2002a; Yamazaki *et al.,* 2003).

CagA-positive *H. pylori* strains are most virulent strains and are associated with higher risk for peptic ulcers (Oleastro *et al.*, 2003; Ribeiro *et al.*, 2003; Yamazaki *et al.*, 2005b). Several studies based on a few strains demonstrated that *cagA*-positive *H. pylori* isolates, but not *cagA*-negative isolates, were able to induce interleukin 8 (IL-8) secretion in *vitro* and in *vivo* (Huang *et al.*, 1995; Sharma *et al.*, 1995). Therefore the ability to induce IL-8 secretion is recognized as one of the major virulence factors of *H. pylori* and seems to be important in the establishment of PUD (Peek *et al.*, 1995).

It has been reported that *cagA* gene is present in approximately 60% of *H. pylori* strains from Western populations, but in contrast, it is present in over 90% of the strains from Southeast Asian populations (Chen *et al.*, 2005; Siavoshi *et al.*, 2005). *CagA* gene is considered as a marker to predict the severity of peptic ulcer disease in European and North American populations (Covacci *et al.*, 1993). The presence of *cagA* has been associated with peptic ulcer disease and gastric cancer (Miehlke *et al.*, 2001).

1.6.3.1 *cagA* EPIYA motifs

cagA is a polymorphic gene that presents different numbers of repeated sequences located in its 3' region. Each repeated region of *cagA* protein contains Glu-Pro-Ile-Tyr-Ala (EPIYA) motifs, including a tyrosine phosphorylation site (Hatakeyama, 2004). Four distinct EPIYA sites have been described, EPIYA-A, EPIYA-B, EPIYA-C and EPIYA-D each of which is conserved in sequence. In Europe, North America, and Australia, many *H. pylori* isolates *cagA* carries EPIYA-A, EPIYA-B and EPIYA-C segments in tandem and is called Western type and

shown in figure 2.3. In East Asian countries such as Japan, Korea and China *cagA* carries EPIYA-A, EPIYA-B and EPIYA-D segments in tandem and is called East Asian type and shown in figure 2.3 (Higashi *et al.*, 2002b). The EPIYA-C segment multiplies (mostly one to three times) in tandem among different Western *cagA* species. *CagA* from East Asian *H. pylori* isolates also possesses EPIYA-A and EPIYA-B segments, but not the repeatable EPIYA-C segment. Instead, it has EPIYA-D segment, which is unique to East Asian *cagA* (Hatakeyama, 2009). The EPIYA-C segment is found in Western *cagA* and it includes hpAfrica1, hpEurope, hpNEAfrica strains and some hpAsia2 strains from Southeast Asia, while the EPIYA-D segment is unique for East-Asian *cagA* in hpEastAsia strains and in some hpAsia2 strains from Ladakh in Northern India (Olbermann *et al.*, 2010).

Majority of *H. pylori* isolates in East Asian countries possess East Asian *cagA*. In contrast, nearly all *H. pylori cagA*-positive strains isolated in Western countries carry Western *cagA* (Figure 1.3). Interestingly, in Southeast Asia countries like Malaysia, Thailand, Philippines and Vietnam (Hatakeyama, 2011). East Asian (ABD) *cagA*-carrying *H. pylori* and Western (ABC) *cagA*-carrying *H. pylori* may have been introduced into Southeast Asian people through migrations of ethnic Chinese and ethnic Indian people, respectively (Graham *et al.*, 2007; Yamaoka, 2009; Sahara *et al.*, 2012).

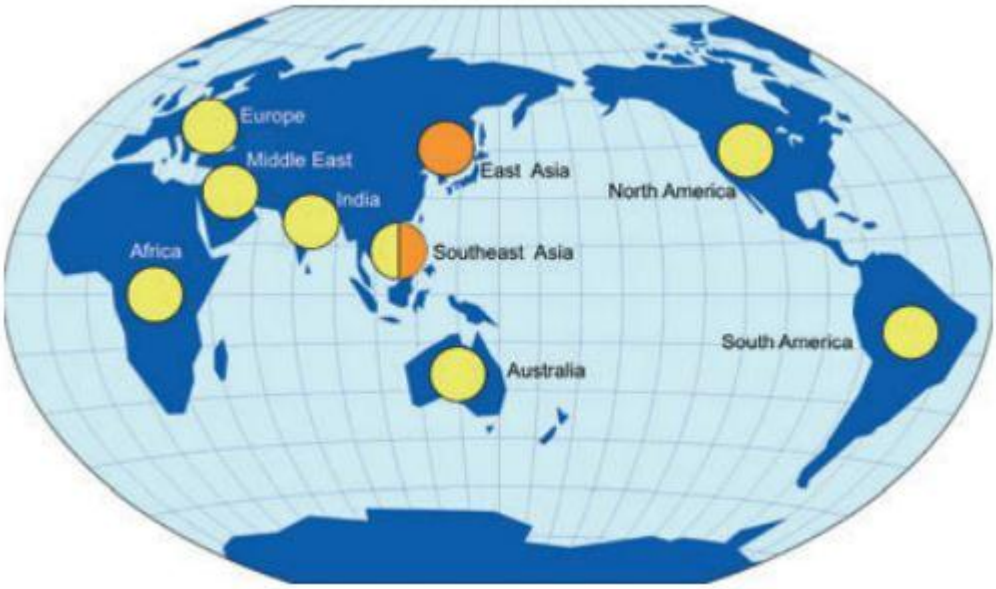

Figure 1.3 Worldwide distribution of *H. pylori* Western *cagA* and East Asian *cagA* Western *cagA* shown in yellow and East Asian *cagA* in orange colour, Adapted from (Hatakeyama, 2011).

More than 60% of Western *CagA* proteins carry the EPIYA-ABC, followed by the EPIYA-ABCC 20.3%, then the EPIYA-ABCCC 4.0% as shown in Figure 2.4 (Hatakeyama, 2011). In contrast to the EPIYA-C segment, the EPIYA-D segment seldom duplicates and thus majority of the East Asian *CagA* isolates contain a single EPIYA-D segment (98.8%), which, in many cases, is present in EPIYA-ABD (Figure 1.4) (Xia *et al.*, 2009).

EPIYA type D or multiple C repeats is associated with increased SHP-2 phosphatase activity induced by *cagA*. Strains possessing *cagA* with greater numbers of type C phosphorylation motifs predispose to precancerous lesions and gastric cancer (Azuma *et al.*, 2002; Naito *et al.*, 2006). Thus, determination of the degree of *cagA* phosphorylation or the number of phosphorylation motifs appears to be more important than detection of *cagA* alone (Argent *et al.*, 2005).

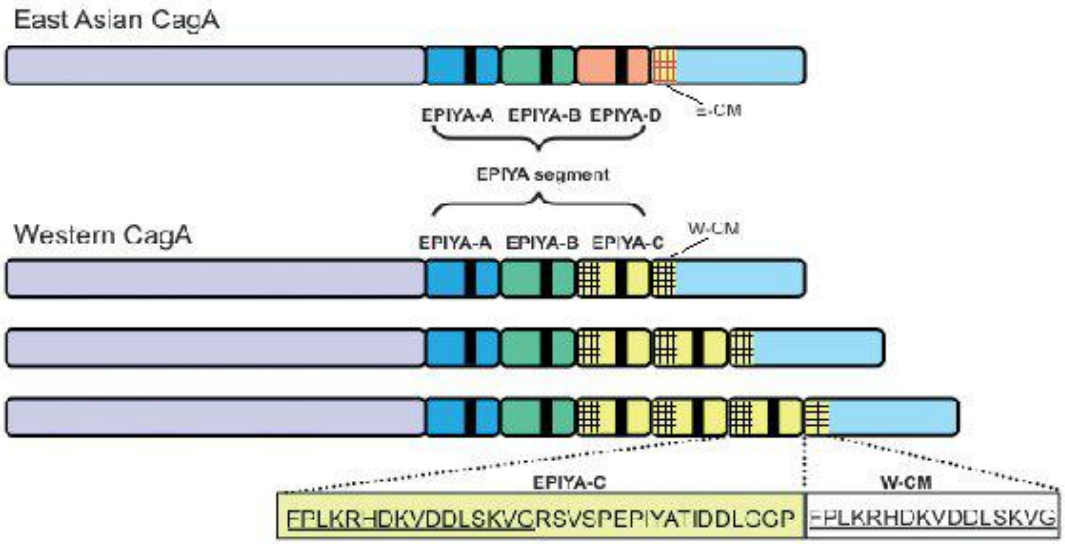

Figure 1.4 Structural polymorphisms in *cagA* and *cagA* multimerization (CM) sequence adapted from (Hatakeyama, 2011).

1.6.4 Vacuolating cytotoxin gene A (*vacA*)

The *vacA* gene is present in all *H. pylori* strains and encodes a vacuolating cytotoxin. There are at least four *vacA* signal sequence types (s1a, s1b, s1c, and s2) and two middle region types (m1, m2) (Ben Mansour *et al.*, 2010). Type s1/m1 mosaic combination strains express more cytotoxin activity than s1/m2 strains, and s2/m2 strains produce no detectable cytotoxin activity (van Doorn *et al.*, 1998; Ben Mansour *et al.*, 2010). Mature *VacA* s1-type strains has a hydrophobic N terminal region which is able to insert into the host cell membrane leading to vacuolation, whereas non-toxigenic s2- type strains have an N-terminus preceded by a hydrophilic region that blocks vacuolation (McClain *et al.*, 2001). *H. pylori vacA* type s1 strains seems to be more virulent than type s2 strains and are associated with higher risks for peptic ulcer disease, gastric atrophy, and gastric carcinoma (Figueiredo *et al.*, 2002).

There is geographical variation in the *vacA* genotypes. In Western country, including Latin America and Africa, studies have reported that individuals infected with s1 or m1 *H. pylori* strains have an increased risk of peptic ulcer or Gastric cancer compared with individuals infected with s2 or m2 strains (Sugimoto and Yamaoka, 2009). There is also a variation of m strain within East Asia; for instance, where m1 strains are common in parts of Northeast Asia, such as Japan and South Korea and m2 strains being predominant in parts of Southeast Asia, such as Taiwan and Vietnam (Yamaoka *et al.*, 2002; Uchida *et al.*, 2009).

1.6.5 Duodenal ulcer promoting gene (*dupA*)

dupA, which is located in the plasticity region of *H. pylori* genome, has been found to be a risk factor for duodenal ulcer (DU) and a protective factor against gastric cancer (GC) in Japan and Korea (Lu *et al.*, 2005). The *dupA* gene contains two continuous sequences, *jhp0917* and *jhp0918* which was initially reported to be a marker for DU development, but some studies showed that this gene can also be associated with GC (Argent *et al.*, 2007; Schmidt *et al.*, 2009a). Studies in patients from Brazil, India, China, USA, South Africa, Belgium, Iraq and Iran have confirmed that *jhp0917* and *jhp0198* are both present and form a continuous ORF, they have been unable to consistently replicate an association with gastroduodenal disease (Arachchi *et al.*, 2007; Douraghi *et al.*, 2008; Hussein *et al.*, 2008; Zhang *et al.*, 2008). A recent meta-analysis by Hussein *et al.* has concluded that the effects of *dupA* may be population specific, predisposing to DU in some populations and GC and GU in others (Hussein, 2010). Recently, *dupA* has been shown to play an important role in provoking IL-8 secretion (Queiroz *et al.*, 2011). In East Asian region this gene has been linked with the high risk of GC development. In the study that identified *cagA* gene in all samples as well suggesting that combination of these genes, may underline the high risk of GC in this area (Wang *et al.*, 2013).

1.6.6 Blood-group antigen binding adhesin (*babA*)

Attachment of *H. pylori* to the gastric epithelium is thought to be a major contributor to *H. pylori* persistence by providing access to nutrients and protection from gastric acid and mucus turnover. *H. pylori* outer membrane proteins (OMPs) have been proposed to be critical for adaptation to the host and persistent

colonisation. Indeed, the attachment is mediated by several OMPs, the best characterised of which are the Lewis(Le)b blood-group antigen binding adhesin (*babA*), the outer membrane inflammatory protein (*OipA*) and the sialic acid binding adhesin (*SabA*) (Ilver *et al.*, 1998).

BabA is a blood-group antigen-binding adhesin encoded by the *babA2* gene, which has been shown to mediate adherence of *H. pylori* to human Lewis b blood-group (Leb) antigens (Gerhard *et al.*, 1999). Although three *bab* alleles have been identified (*babA1*, *babA2* and *babB*), only the *babA2* gene product is functional for Leb binding activity (Pride *et al.*, 2001). Studies in Western countries have disclosed associations between the presence of *babA2* gene and digestive diseases such as duodenal ulcer and gastric cancer (Gerhard *et al.*, 1999). However, in Asia, most of the *H. pylori* strains are *babA2*-positive, irrespective of clinical outcome (Mizushima *et al.*, 2001; Sheu *et al.*, 2003). A study done in Iranian patients recently reported that *babA2* distribution was significantly higher in GC subjects (95%) when compared with non-ulcer dyspepsia patients (26.1%) and DU patients (Talebi Bezmin Abadi *et al.*, 2013) Thus, conclusions about the relationship between *H. pylori* genotypes and clinical outcome derived from one geographic region may not be true for other geographic regions.

1.6.7 Sialic acid-binding adhesin (*sabA*)

H. pylori OMP, the sialic acid-binding adhesin (*SabA*), binds to sialylated carbohydrate structures, which are upregulated as part of complex gangliosides in inflamed gastric tissue. The capacity to bind to the glycosylated epithelial cells is considered to be important for *H. pylori* to cause constant infection and disease

(Aspholm *et al.*, 2006; Odenbreit *et al.*, 2009). *SabA* was postulated to contribute to the chronic persistence of the infection (Mahdavi *et al.*, 2002; Aspholm *et al.*, 2006). Investigation of the role of *SabA* in disease development has gained momentum over the last few years, with evidence suggesting that *sabA* not only plays a role in disease development but that this role may be consistent across populations. For example, a study of a developing Hispanic country (Columbia) and a developed Caucasian country (USA) found a similarly increased prevalence of *sabA* in DU and GC in both populations (Yamaoka *et al.*, 2006). It has also been reported that the off status of *sabA* is associated with DU, but not GU, suggesting that *sabA* may be a reliable marker for specific disease outcome (de Jonge *et al.*, 2004). However, a more recent study in Taiwan found no association with disease (Sheu *et al.*, 2006).

1.7 *H. pylori* and clinical outcome

H. pylori infection is found to be associated with gastritis, DU, GU, non-ulcer dyspepsia, GC and gastric lymphoma of mucosa associated lymphoid tissue (MALT) (Figure 2.5). Infection usually occurs during childhood and causes symptomatic acute gastritis in most patients and persists for decades or life-long, the infection can take multiple courses. Most people infected with *H. pylori* usually show no developing symptomatic disease; however, 10-15 % will develop peptic ulcer disease, approximately 1% will develop gastric adenocarcinoma, while a small group of patients will develop gastric mucosa associated lymphoid tissue lymphoma (MALToma) (Wu *et al.*, 2008; Varbanova and Malfertheiner, 2011). A normal or high acid secretion predisposes to duodenal ulcers (DU), whereas a low acid

secretion predisposes to gastric ulcers (GU) and gastric cancer (GC) (Gerrits *et al.*, 2006) as indicated in Figure 1.5.

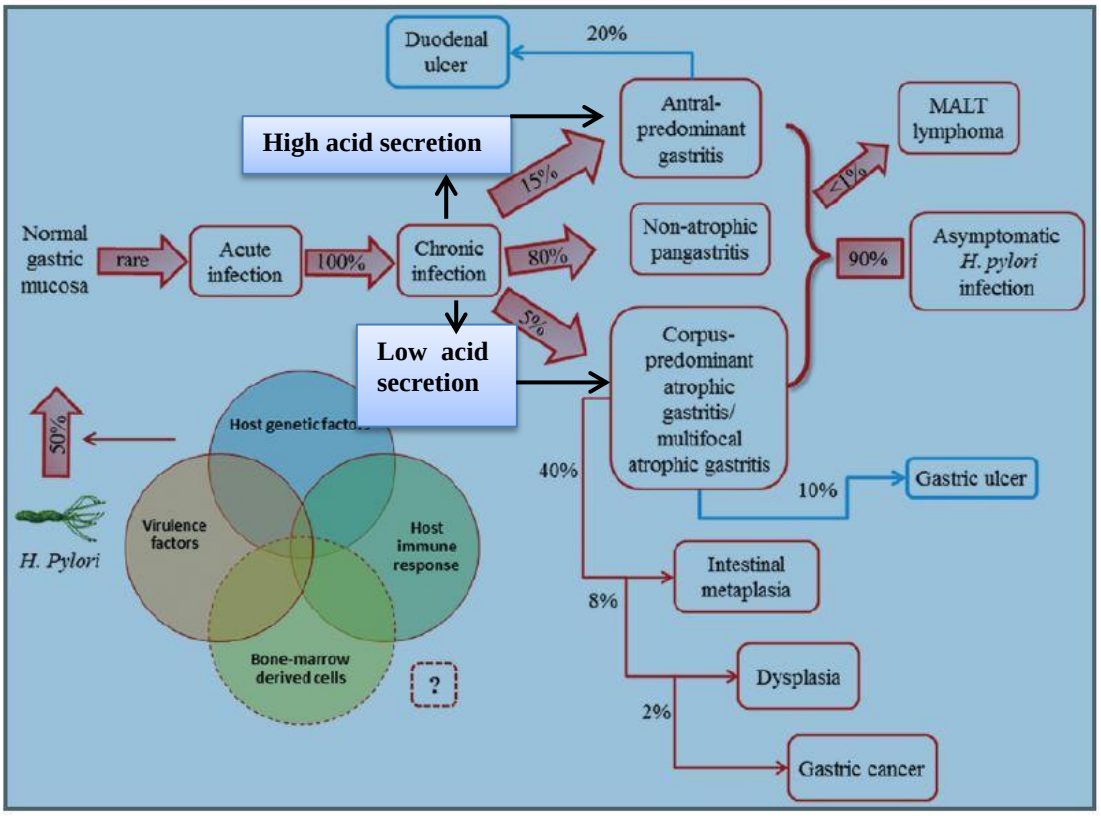

Figure 1.5 Modified natural history of *H. pylori* infection adapted from (Conteduca *et al.*, 2013).

Patients with increased acid secretion are likely to have antral-predominant gastritis, which predisposes to duodenal ulcers. Patients with low acid secretion will more likely develop gastritis in the body of the stomach and are thus more likely to develop gastric ulcer, leading to gastric atrophy, intestinal metaplasia, dysplasia and in rare cases, gastric carcinoma. This sequence of events is more frequent in people of advanced age. *H. pylori* infection induces the formation of mucosa-associated lymphoid tissue (MALT) in the gastric mucosa and MALT lymphoma is another rare complication of *H. pylori* infection (Figure 1.5).

The gastric mucosa is well protected against bacterial infections. *H. pylori* is adapted to this ecologic niche, with a array of features that permit entry into the mucus, swimming spatial orientation in the mucus, attachment to epithelial cells, evasion of the immune response and as a result persistent colonization and transmission (Suerbaum and Michetti, 2002). After being ingested, the bacteria have to evade the bactericidal activity of the gastric luminal contents and enter the mucous layer. Urease production and motility are essential for this first step of infection. Urease hydrolyzes urea into carbon dioxide and ammonia, thereby permitting *H. pylori* to survive in an acid milieu (Weeks *et al.*, 2000; Dhar *et al.*, 2003).

1.7.1 Gastritis

Gastritis refers to inflammation of the gastric mucosa. When a person is infected with *H. pylori* the initial response to infection is the development of an acute gastritis. This acute phase is characterized by the presence of fever, vomiting, nausea