Detection of *Helicobacter pylori* Microevolution and Multiple Infection, and Genomic Analysis of *Helicobacter pylori* Strains

Montserrat Palau de Miguel

Aquesta tesi doctoral està subjecta a la llicència ***Reconeixement 4.0. Espanya de Creative Commons***.

Esta tesis doctoral está sujeta a la licencia ***Reconocimiento 4.0. España de Creative Commons.***

This doctoral thesis is licensed under the ***Creative Commons Attribution 4.0. Spain License.***

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I
CIÈNCIES DE L'ALIMENTACIÓ

DEPARTAMENT DE BIOLOGIA, SANITAT
I MEDI AMBIENT

SECCIÓ DE MICROBIOLOGIA

Detection of *Helicobacter pylori* Microevolution and Multiple Infection, and Genomic Analysis of *Helicobacter pylori* Strains

MONTSERRAT PALAU DE MIGUEL

Setembre 2019

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

PROGRAMA DE DOCTORAT DE BIOTECNOLOGIA

Detection of *Helicobacter pylori* Microevolution and Multiple Infection, and Genomic Analysis of *Helicobacter pylori* Strains

Memòria presentada per Montserrat Palau de Miguel per optar al títol de doctor per la Universitat de Barcelona

Director i Tutor | Doctoranda

David Miñana i Galbis | Montserrat Palau de Miguel

MONTSERRAT PALAU DE MIGUEL
Setembre 2019

“If you go immunosuppressed for a little bit, they’ll kill you.

When you die, they’ll eat you. They don’t care. It’s not a relationship.

It’s just biology.”

Ed Yong. I contain multitudes

"Science is not only a discipline of reason

but also one of romance and passion."

Stephen Hawking

ACKNOWLEDGMENTS

Necessitaria moltes pàgines per poder agrair a tothom aquests 5 anys de doctorat... Primer que res al David, gràcies per creure en mi des del primer moment. Amb tu vaig començar aquesta aventura quan només feia segon de carrera i vaig atendre les teves classes de Microbiologia I. Es notava la dedicació i l'entusiasme que posaves en ensenyar-nos-ho tot, moltíssimes gràcies per això. De veritat dic que no puc imaginar-me un millor director de tesi, si he arribat fins aquí és gràcies a tu.

Gràcies Vicen i gràcies Merche, amb vosaltres he passat moltes hores i sempre m'heu donat un cop de mà quan l'he necessitat. Pels incomptables cafès que he fet amb tu, Vicen; i les converses infinites per esvair-nos de la feina amb tu, Merche. Sempre és un plaer treballar i compartir el temps amb vosaltres. Espero que encara que passin els anys us pugui tenir al meu costat.

A tota la gent amb la que he compartit departament, la Lucía, Maria Jesús, Carme, Carla, Shirin, Nico, Guille, Anderson, Carme F, Ana, Àngels, Mile, Maribel i Núria (P. i R.), m'heu vist créixer durant aquests anys i sempre ha sigut un plaer treballar amb vosaltres. També a la Pepita per resoldre sempre tots els dubtes que tenia sobre el programa de doctorat, i una menció especial per l'Amparo Cortés, gràcies a ella vaig tenir l'oportunitat de fer una estada d'11 mesos al Canadà en el 3r any de tesi, i la relació amb ella sempre ha sigut molt bona. Gràcies també a la Fundació de la Marató de TV3, per la seva feina i per concedir-nos el projecte amb el que aquesta tesi va començar a agafar forma.

Thanks to all my labmates in Halifax, my stay with you changed my life professionally and personally. When I met you I didn't know anything about bioinformatics, your patience and your excellent work atmosphere surpassed all my expectations. Thanks to you I discovered a whole new world (although the computer has driven me crazy a lot of times, I would learn bioinformatics with you a thousand times more). Thank you for always helping me, for repeating your words twice or more because I didn't catch them

at first. Thanks for letting me know the magic of Canada and the kindness of you, Canadians. Morgan, André, Gavin, Akhilesh, Emily, Casey and Karl, you will always be in my heart.

A tots els meus amics, els del Morell (especial menció a l'artista Dani Rull, gràcies per la portada), les del màster, les del col·legi (MiG), que sempre que em preguntàveu com ho portava, em donàveu ànims i força per no defallir. També a la meva família del pis Arizala, els 'Ballestas y fregasuelos', tornar al pis i estar amb valtrus era el millor del dia. A tota la gent que m'he creuat pel camí mentre estava fent la tesi, a totes les catalanes de Halifax, que us vau convertir en la meva família durant un temps, i ara sempre us tindré presents. També a les amigues de Dresden, que m'han conegut en les últimes etapes del doctorat. A la intel·ligenta i repelenta Ga, que ens vam conèixer al màster i per mi és una persona indispensable. I sobretot a la Carlins, amb la que he conviscut quasi 10 anys de la meva vida i estimo com una germana més.

També li dedico a tota la meva família, però sobretot a mon pare, ma mare, i ma germanes Maria Teresa i Anna. Gràcies per recolzar-me en tot moment, en els bons i en els mals moments. Per alegrar-vos quan tenia bones notícies i alegrar-me el dia quan he estat lluny i tenia dies negres. Papa i mama, vosaltres mai heu deixat de creure en mi i sempre heu tingut bons consells per donar-me; Tresi tu sempre m'has dit que sóc una crack i això em feia (i em fa) pujar els ànims i tenir més energia per treballar; i Anna, la teva experiència m'ha servit sempre per no perdre'm en aquest camí, ets el meu referent. I gràcies àvia, sé que em cuides sempre.

Gràcies Jordi, amb tu em vaig creuar quan començava aquest camí, i sempre, en tot moment t'he tingut a prop. Gràcies per donar-me consells, gràcies per no deixar-me enfonsar, gràcies per ser com ets, gràcies per estar al meu costat.

ABSTRACT

In the past decades, *Helicobacter pylori* has received the attention of many researchers because of its known relation with gastric cancer. Although many studies have tried to decipher the exact relation between the bacteria and cancer state, and several virulence factors have been discovered, an exact answer has not been found yet. Further work should be made in order to study more accurately the genome of this bacterium and to understand its precise involvement. The bacterium is characterised for a highly genetic diversity, meaning it is continuously changing in order to adapt itself to its hostile niche, the human stomach.

Infection by *H. pylori* is estimated to affect half of the world's population, being more extended in developing countries than in developed ones, possibly due to the high consumption of antibiotics and the increased level of sanitation in the latest. It has been demonstrated that the gastric lumen can be colonized by more than just one strain of the bacterium, sometimes these strains could have evolved from the same 'mother' strain, or they could come from unrelated strains. The study of these situations is important in order to elucidate if there is just one strain who is responsible for starting the pathogenic cascade, and what are the specific differences between the different strains that inhabit the human stomach.

On the first work of this thesis, our group studied the usefulness of six housekeeping genes for the detection of *H. pylori* infection and the characterization of various strains isolated from gastric isolates, studying as well their phylogeny. In some cases, the distance value between the strains was high, indicating and event of multiple infection. In other cases, small differences were found between clones, suggesting events of microevolution rather than multiple infection.

This work was further extended with the study of the usefulness of amplicon sequencing of these housekeeping genes in the detection of microevolution and mixed infections from gastric biopsies of patients with dyspeptic symptoms and different

histopathological findings (from atrophy to adenocarcinoma). Five gastric biopsies from four patients infected by *H. pylori* were involved in this study. We detected in all the analyzed gastric biopsies multiple *H. pylori* infections with a predominant strain. These results suggest that *H. pylori* colonizes the human stomach through diverse infection circumstances that lead to a gastric multi-infection with a predominant strain together alongside other minority strains. Furthermore, it was shown that mixed infections are the main status in the colonization of the human gastric mucosa.

The last part of this thesis started with a preliminary study of 51 complete sequenced *H. pylori* genomes and further focused on three genomes obtained from the same patient in order to analyse and compare them. Particularly, these isolates were sampled at the same time from a stomach with adenocarcinoma, one strain was from the non-tumoral tissue, and the other two were isolated from the tumoral tissue. They all lacked from the most noticeable virulence factor, the *cag* pathogenicity island; one of the most studied and the main factor related to the malignancy of the bacterium. On the other hand, we found differences in the genotype of the vacuolating cytotoxin gene (*vacA*) and in genes related with urease, the outer membrane and flagella.

Despite the contributions made in this thesis, further studies are needed to find better genetic markers of *H. pylori* related to virulence and progression to gastric cancer.

RESUM

En les últimes dècades, *Helicobacter pylori* ha estat el focus d'atenció de molts estudis a causa de la seva coneguda relació amb el càncer gàstric. Encara que molts estudis han tractat de desxifrar la relació exacta entre el bacteri i la progressió del càncer, i tot i que s'han descobert diversos factors de virulència, encara no s'ha trobat una resposta exacta. Es requereixen majors esforços per conèixer millor el genoma d'aquest bacteri i comprendre la seva implicació d'una manera més precisa. El bacteri es caracteritza per tenir una gran diversitat genètica, que li permet estar canviant contínuament per adaptar-se al seu nínxol hostil, l'estómac humà.

S'estima que la infecció per *H. pylori* afecta a la meitat de la població mundial, essent més estesa en els països en desenvolupament que en els desenvolupats, possiblement a causa de l'alt consum d'antibiòtics i el major nivell de sanejament en aquests últims. S'ha demostrat que el lumen gàstric pot estar colonitzat per més d'una soca del bacteri; en alguns casos aquestes soques poden haver evolucionat a partir de la mateixa soca "mare", i en d'altres poden provenir de soques no relacionades. L'estudi d'aquestes situacions és important per tal de dilucidar si només hi ha una soca responsable d'iniciar la cascada patogènica, i quines són les diferències específiques entre les diferents soques que habiten l'estómac humà.

En la primera part d'aquesta tesi, el nostre grup va estudiar la utilitat de sis gens conservats per la detecció de la infecció per *H. pylori* i la caracterizació de diverses soques aïllades de biòpsies gàstriques, estudiant-se també la seva filogènia. En alguns casos, el valor de la distància entre les soques era alt, fet indicatiu d'un esdeveniment d'infecció múltiple. En altres casos, es van trobar petites diferències entre els clons, el que suggereix esdeveniments de microevolució en lloc d'infecció múltiple.

Aquesta tesi es va ampliar amb l'estudi de la utilitat de la seqüenciació d'amplicons d'aquests gens conservats en la detecció de microevolució i infeccions múltiples en biòpsies gàstriques de pacients amb símptomes dispèptics i amb diferents diagnòstics histopatològics (des d'atròfia fins adenocarcinoma). En aquest estudi es van analitzar

cinc biòpsies gàstriques de quatre pacients infectats per *H. pylori*. En totes les biòpsies gàstriques analitzades es van detectar infeccions múltiples per *H. pylori* amb una soca predominant. Aquests resultats suggereixen que *H. pylori* colonitza l'estómac humà mitjançant diverses circumstàncies d'infecció que condueixen a una infecció múltiple gàstrica amb una soca predominant juntament amb altres soques minoritàries. A més, es va demostrar que les infeccions mixtes són l'estat principal de la colonització de la mucosa gàstrica humana.

L'última part d'aquesta tesi va començar amb l'estudi preliminar de 51 genomes complets d'*H. pylori* i es va centrar després en l'estudi i comparació de tres genomes obtinguts del mateix pacient. Específicament, aquestes soques van ser aïllades alhora d'un estómac amb adenocarcinoma, on una soca provenia del teixit no tumoral i les altres dues del teixit tumoral. Totes mancaven del factor de virulència més destacat, l'illa de patogenicitat *cag*; un dels factors més estudiats i el més relacionat amb la malignitat del bacteri. Per una altra banda, vam trobar diferències en el genotip del gen *vacA* i en gens relacionats amb la ureasa, la membrana externa i els flagels.

Malgrat les contribucions fetes en aquesta tesi, encara són necessaris més estudis per trobar millors marcadors genètics d'*H. pylori* relacionats amb la virulència i la progressió a càncer gàstric.

INDEX

ACKNOWLEDGMENTS VII

ABSTRACT IX

RESUM XI

CHAPTER 1 INTRODUCTION 1

1.1 History, discovery and Nobel laureates. 3

1.2 Phenotypic and cultural characteristics 4

1.3 Cellular structure 6

1.3.1 Flagella 7

1.3.2 Lipopolysaccharide (LPS) 9

1.4 Epidemiology 13

1.4.1 Transmission 14

1.4.2 Mixed infections 15

1.4.3 *Helicobacter pylori* as a marker for human migrations 16

1.5 Diagnosis, treatment, and antibiotic resistances 21

1.5.1 Diagnosis 21

1.5.2 Treatment 23

1.5.3 Antibiotic resistances 25

1.6 Pathogenicity 29

1.6.1 Virulence factors 31

1.6.2 Correa pathway 37

1.7 Is *Helicobacter pylori* beneficial or detrimental? 41

CHAPTER 2 OBJECTIVES 43

CHAPTER 3 Usefulness of housekeeping genes for the diagnosis of *Helicobacter pylori* infection, strain discrimination and detection of multiple infection 45

3.1 Introduction 47

3.2 Material and Methods 48

3.2.1 *Helicobacter pylori* clones and DNA extraction 48

3.2.2 DNA extraction, PCR amplification and sequencing 50

3.2.3 Data analyses 52

3.3 Results 52

3.4 Discussion 58

CHAPTER 4 Detection of *Helicobacter pylori* microevolution and multiple infection from gastric biopsies by housekeeping gene amplicon sequencing 61

4.1 Introduction 63

4.2 Material and Methods 64

4.2.1 Gastric biopsies 64

4.2.2 DNA extraction and PCR amplification 65

4.2.3 Preparation of libraries 66

4.2.4 Data processing 69

4.3 Results 70

4.3.1 Samples and amplicon sequences 70

4.3.2 OTU assignment and distribution 73

4.3.3 Amplicon sequence identification 73

4.4 Discussion 79

CHAPTER 5 Whole-genome sequencing and comparative genomics of three *Helicobacter pylori* strains from a stomach with adenocarcinoma 81

5.1 Introduction 83

5.2 Materials and Methods 85

5.2.1 *Helicobacter pylori* strains and DNA sequencing 85

5.2.2 Bioinformatic analysis for genomic comparison of strains B508A-S1, B508A-T2A, and B508A-T4. 87

5.3 Results 89

5.3.1 Basic comparison of all the genomes 89

5.3.2 B508A-S1, B508A-T2A, and B508A-T4 genome comparison 93

5.4 Discussion 101

5.4.1 Basic comparison of all the genomes 101

5.4.2 B508A-S1, B508A-T2A, and B508A-T4 genome comparison 103

CHAPTER 6 CONCLUSIONS 109

REFERENCES 115

ANNEX I. Supplementary material to Chapter 3 143

ANNEX II. Supplementary material to Chapter 4 145

ANNEX III. Supplementary material to Chapter 5 149

ANNEX IV. Publications during the PhD period 159

ANNEX V. Participation in projects, Consolidated Research Groups and research stays 161

CHAPTER 1 **INTRODUCTION**

1.1 History, discovery and Nobel laureates.

“3 October 2005.

The Nobel Assembly at Karolinska Institutet has today decided to award The Nobel Prize in Physiology or Medicine for 2005 jointly to **Barry J. Marshall** and **J. Robin Warren** for their discovery of ‘**the bacterium *Helicobacter pylori* and its role in gastritis and peptic ulcer disease’**.”

These words culminated the long trajectory accomplished since the 80's by Dr. Marshall and Dr. Warren about the study and discovery of the gastric bacterium *Helicobacter pylori*. Since then, its study has grown exponentially, drawing the attention of many researchers.

It was in 1981 when Marshall and Warren met each other. Together they became an unstoppable tandem. They had to refute a fact which was really internalised in the scientific community: nothing could ever grow in the stomach due to its low pH. Robin Warren had been observing for years curved bacteria colonizing the antrum of the stomach in about half of the biopsies he obtained from his patients. He realised that nearby where these bacteria were seen, there were always signs of inflammation. Barry Marshall became interested in Warren’s work and together they undertook a more complete clinico-pathological study with 100 patients. Their first publication described curved bacilli related to the genus *Campylobacter* that were isolated from stomach biopsies of patients with gastritis and peptic ulceration (Marshall and Warren, 1984). Marshall managed to cultivate this by then unknown bacterium and together they discovered that this organism was present in almost all patients with gastric inflammation, duodenal ulcer or gastric cancer. Based on their results, they concluded that *Helicobacter pylori* was involved in the aetiology of these diseases (NobelPrize.org, 2005a), while refusing the fact that emotional stress and spicy food were the major causes of gastritis and ulcers, as it was thought to be at that time (Kyle *et al.*, 2016). Robin Warren and Barry Marshall proved in 1982 that patients could only be cured if the

bacterium was eliminated. This is now achieved by treatment with antibiotics, and gastric ulcers are no longer a chronic illness (NobelPrize.org, 2005b).

It is also famously known how Marshall experimented with his own body to prove Koch's postulates (Koch, 1882). After several attempts to develop an animal model and several unfortunate endings, he finally drank a culture of the microorganism. Three days later he manifested nausea and achlorhydria. New endoscopy and biopsies showed marked gastritis and a positive *Helicobacter pylori* culture. In this way he successfully proved Koch's postulates. Marshall finally recovered entirely when he was treated with antibiotics and bismuth.

Helicobacter pylori was initially named "*Campylobacter*-like organism", "gastric *Campylobacter*-like organism", "*Campylobacter pyloridis*" and "*Campylobacter pylori*", but is now named *Helicobacter pylori*, which is usually shortened to *H. pylori* in recognition for the fact that this organism is distinct from members of the *Campylobacter* genus, and described as a novel type species, constituting the novel genus *Helicobacter* (Goodwin *et al.*, 1989).

1.2 Phenotypic and cultural characteristics

Helicobacter pylori is a Gram-negative bacterium with the following phenotypic characteristics (Gu, 2017; Kusters *et al.*, 2006; Reshetnyak and Reshetnyak, 2017):

- It has helical or spiral shape and measures 2 to 4 μm in length and 0.5 to 1 μm in width (**Figure 1.1**).
- The organism has 4 to 8 unipolar, sheathed flagella of approximately 3 μm in length.

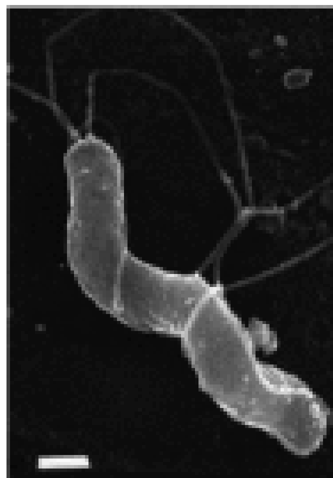


Figure 1.1 *H. pylori* microscopic image (Reshetnyak and Resyetnyak, 2017).

1.2 Phenotypic and cultural characteristics

- It is urease, catalase and oxidase-positive.
- It belongs to the Epsilon subdivision of the *Proteobacteria*, order *Campylobacterales*, family *Helicobacteraceae*.
- It is microaerophilic, i.e. it needs a low concentration of oxygen (2–5%) and capnophylic, i.e. needs elevated concentrations of CO_2 (5–10%) to grow.
- Its optimal growth temperature is 37ºC, but it can grow within a range of 34ºC to 40ºC.
- Despite its natural habitat being the acidic gastric mucosa, it is considered to be a neutrophile. Growth occurs only at pH range of 5.5 to 8.0, with optimal growth at neutral pH.

Initial colonization depends on bacterial urease activity and cell-shape modulation to penetrate the gastric mucus. Constitutive DNA and protein repair pathways, combined with bacterial genome diversification and attenuation of host cell chemical radical production are now recognized as essential for persistence of the bacterium in this niche (Salama *et al.*, 2013).

H. pylori disperses within the gastric mucus layer and remains in direct contact with the gastric epithelial cells. It is hypothesized that the bacteria localize these areas to escape the low pH of the stomach lumen, where they can survive only for a matter of minutes, and to avoid elimination by peristalsis. *H. pylori* requires flagella-mediated and chemosensory-directed motility to access and maintain itself in the mucus layer. *H. pylori*'s helical cell shape may contribute to this process by enabling the bacteria to bore into the mucus layer via a cork-screwing mechanism (Sycuro *et al.*, 2012). The microaerophilic nature of the bacterium may be due in part to the involvement of oxygen-sensitive enzymes in central metabolic pathways (Kelly, 1998).

In vitro, *H. pylori* needs complex culture media to grow, among them, rich mediums containing blood or serum are the most utilized. The most common used media are Columbia blood agar, Brucella agar, brain heart infusion agar or trypticase soy agar as the base. The last three must be supplemented with sheep blood or horse blood (Ndip

et al., 2003). When growth of *H. pylori* in liquid media is required, Brucella broth supplemented with fetal bovine serum (5–10%) is frequently used (Buck and Smith, 1987). One hypothesis regarding fetal bovine serum usage is that serum may contain growth stimulating factors (Cover, 2012).

For the culture it is also necessary a microaerophilic atmosphere (2–5% O_2), elevated concentrations of CO_2 (5–10%), a temperature of 37ºC and incubation periods extending from 4 to 10 days (Cover, 2012). Incubation periods of up to 10 days are usually recommended to optimize the culture isolation rates.

H. pylori uses glucose as its unique source of energy and to start substrate-level phosphorylation (Ndip *et al.*, 2003). Practically all *H. pylori* strains have an absolute need for arginine, histidine, leucine, methionine, phenylalanine and valine aminoacids for their growth (Cover, 2012; Reynolds and Penn, 1994). Interestingly, these amino acids are similar to the ones essential for humans (Cover, 2012). Other aminoacids, like alanine, serine, cysteine, proline and isoleucine, have variable requirements among strains. Supplementary studies show *H. pylori* preferential usage of amino acids over sugars as carbon source (Mendz and Hazell, 1995). Additional factors needed for *H. pylori* growth are pyruvate, thiamine and hypoxanthine, and the metals iron, zinc and magnesium which are critical for growth (Testerman *et al.*, 2006, 2001).

1.3 Cellular structure

H. pylori is a Gram-negative bacterium and consequently, its cellular structure consists of (i) an outer-membrane, where outer-membrane proteins (OMP), porins and virulence components, i.e. the lipopolysaccharide (LPS), reside; (ii) a thin peptidoglycan layer, which maintains the structure of the cell; (iii) the periplasm which separates the outer membrane from the (iv) inner membrane, which is in direct contact with the cytoplasmic space. Regarding its morphology, a typical *H. pylori* cell has a helical shape, which is essential for retaining its pathogenic potential. Additionally, it is a flagellated bacterium with lophotrichous flagella. The flagella play a significant role in the mobility

of *H. pylori* and in reaching the mucus layer. For this reason, it is of interest to further delve into its characteristics. Additionally, due to the special type of lipopolysaccharide found in *H. pylori*, further details are provided about it thereafter.

1.3.1 Flagella

1.3.1.1 Function and structure

Flagellar motility of *H. pylori* influences host colonization by promoting migration through viscous milieus such as gastrointestinal mucus. *H. pylori* has between four and eight unipolar flagella. The bacterial flagellum is a complex motility organ built with multiple types of protein subunits. These protein subunits conform then two main substructures of the flagellum: the hook-basal body complex and the extracellular filament (Gu, 2017; Lertsethtakarn *et al.*, 2011).

Helicobacter species manifest unusually high velocities and movements in viscous substances compared to many motile bacteria, which may probably be due to evolutionary pressure from residing in viscous mucus environments. In the typically studied peritrichous, rod-shaped bacteria (such as *Salmonella* species and *Escherichia coli*), highest swimming velocities occur in low to slightly viscous substances, with severe velocity reductions as viscosity increases. In contrast, *H. pylori* is a helical bacterium with exclusively polar flagella. Many of these *H. pylori* bacterial strains demonstrate roughly twice the velocity of rod-shaped bacteria in substances of low viscosity (ranging from 22 to 40 µm s^{-1}) and can retain these velocities as the viscosity increases between 40 to 80 fold (Worku *et al.*, 1999). The flagella provide both propulsive torque and rotation to the cell body. The helical shape of the cell produces a corkscrew-shaped rotation that allows bacteria to press against a rigid substance to promote motility in the viscous mucus (Lertsethtakarn *et al.*, 2011).

1.3.1.2 Flagellar biosynthesis

As mentioned above, flagella consist of a hook-basal body complex and an extracellular filament. The hook-basal body complex is divided into three structures, which are depicted in **Figure 1.2**: (i) the base located in the cytoplasm and inner

membrane, (ii) the periplasmic rod and associated ring structures, and (iii) the surface-localized hook (Lertsethtakarn *et al.*, 2011). These structures are synthesized from inside out and they are composed of multiple types of protein subunits.

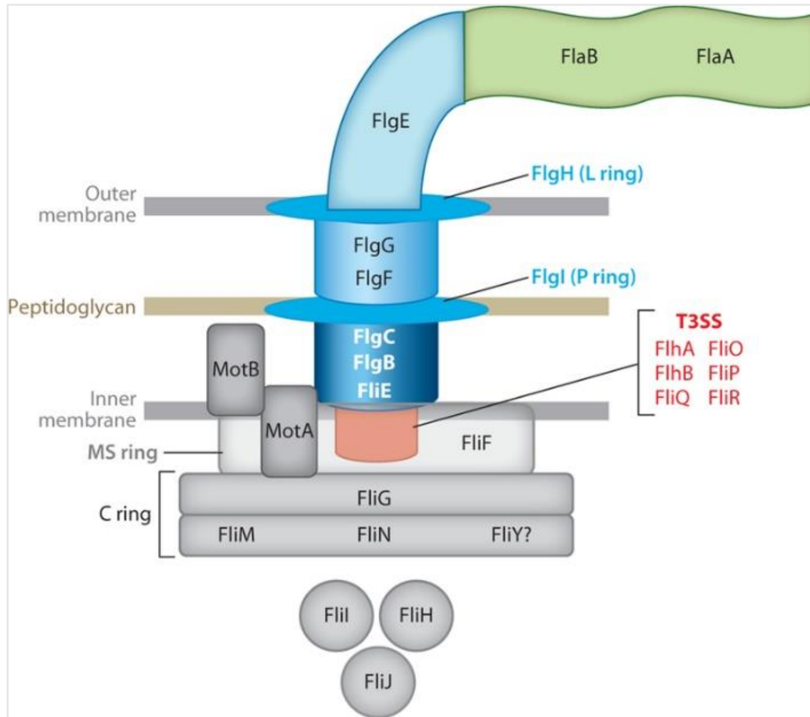


Figure 1.2 Flagellum structure (Lertsethtakarn *et al.*, 2011).

The flagellar base consists of the MS ring, the flagellar type III secretion system (T3SS), the cytoplasmic C ring (or switch complex) and the motor. The MS ring is a homomultimer of FliF and it is involved in the synthesis of the proteins FlaA, FlaB and FlgE (Allan *et al.*, 2000). The functional flagellar T3SS (composed of FlhA, FlhB, FliO, FliP, FliQ and FliR) transfers the majority of the flagellar proteins that are part of the flagellum beyond the inner membrane. The C ring, located at the cytoplasmic site of the MS ring, transports proteins, coordinates motor rotation and facilitates protein secretion. This complex is composed of the proteins FliG, FliM, FliN and FliY, which share homology with FliN and may have a partially redundant function in *H. pylori*. The motor components of the base include MotA and MotB, and its function is to fix and to generate the flagellar rotation (Gu, 2017; Lertsethtakarn *et al.*, 2011).

After the MS ring and C ring are constructed, different proteins, which are in charge of the correct flagellar biosynthesis, pass through T3SS. FliE, FlgB and FlgC are the first to be secreted, which polymerize on the periplasmic surface of the MS ring to form a channel. Afterwards, the FlgF and FlgG proteins are secreted into this tube and polymerize on the apex of the growing flagellar structure. This pattern of secretion and polymerization on the tip extends throughout the biosynthesis of the hood and filament. FlgI protein forms the P ring in the peptidoglycan and FlgH the L ring in the outer membrane. These rings are important elements to provide support. After completion of the inner structures, the flagellar hook protein –FlgE– and minor hook proteins are secreted. FlgE connects the basal body to the flagellar filament.

Flagellar biosynthesis terminates with filament biosynthesis. The filament is composed primarily of the major flagellin, FlaA (which plays an important role in bacterial motility), and the minor flagellin, FlaB, less abundant in the structure.

Flagella of *H. pylori* are covered by a membranous sheath, a relatively atypical feature of bacterial flagella. Its role is hitherto unknown, but it's suggested that the sheath prevents recognition by host antibodies or innate immune receptors, a hypothesis sustained by the observation that flagellin-specific antibodies do not bind to *H. pylori*. The sheath is composed by the proteins HpaA and FaaA, which protect the flagellin subunits against depolymerisation at low pH (Geis *et al.*, 1993; Goodwin *et al.*, 1985; Gu, 2017; Kostrzynska *et al.*, 1991).

1.3.2 Lipopolysaccharide (LPS)

1.3.2.1 Function and structure

The lipopolysaccharide (LPS) is an important factor to promote immune escape and chronic infection in *H. pylori*, due to both a decoration of the O-antigen with Lewis antigens which develop host mimicry to facilitate immune escape, and the singular structure of the lipid A-core which provides resistance to host cationic antimicrobial peptides (CAMPs) (Li *et al.*, 2018, 2016).

LPS is located in the external membrane of most Gram-negative bacteria and is typically composed of three domains (**Figure 1.3**): (i) a hydrophobic domain termed lipid A, which is fixed in the outer membrane; (ii) a relatively conserved non-repeating core oligosaccharide (core OS) which links the O-antigen to the lipid A; and (iii) a variable outermost polysaccharide known as O-antigen (Oliveira and Reygaert, 2019).

The singular structure of lipid A enables *H. pylori* to evade the host innate immune system and its intrinsic resistance to antimicrobial peptides, facilitating the establishment of a perpetual colonisation. This structure of lipid A is a result of enzymatic dephosphorylation (removal of phosphate groups from the 1- and 4'- positions of the lipid A backbone) and deacylation. As a result, the *de novo* synthesised lipid A in *H. pylori* is integrally modified into an unusual lipid A with the loss of negative charges, which confers its above mentioned properties (Cullen *et al.*, 2011; Li *et al.*, 2018).

In addition, the O-antigen of *H. pylori* LPS contains fucosylated oligosaccharides that mimic human Lewis antigens. *H. pylori* is known to extensively vary its Lewis antigen expression pattern *in vivo*, which also contributes to its ability to evade host immune detection and to adapt to the host environment during persistent infection (Li *et al.*, 2017).

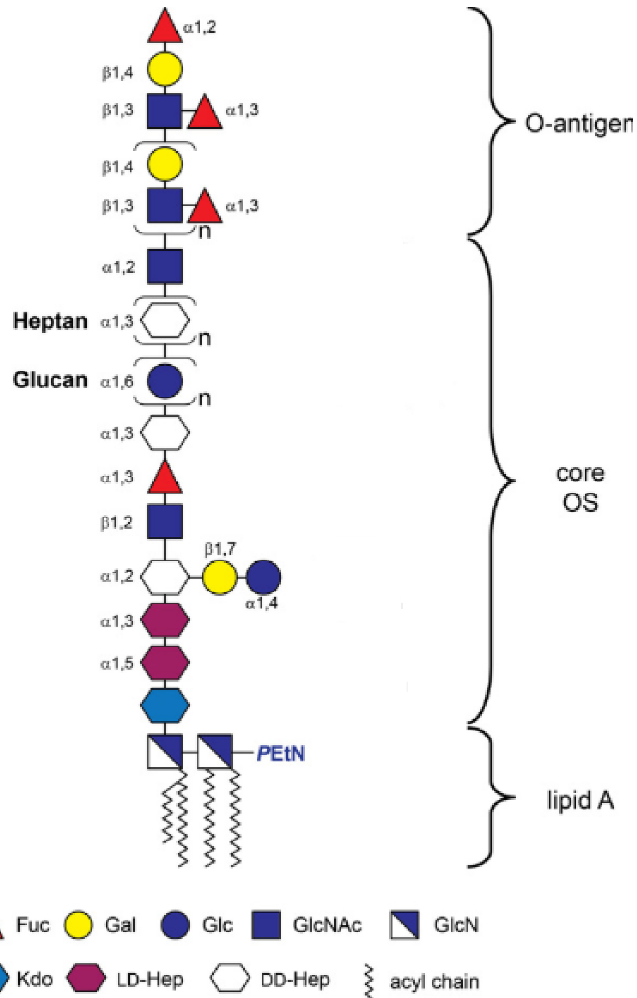


Figure 1.3 Proposed LPS structure of the reference strain *H. pylori* 26695. Modified from Li *et al.* (2016).

The outer part of the core OS structure is a linear arrangement of DD-heptan and α-1,6-glucan linked to the inner core through a trisaccharide (GlcNAc-Fuc-DD-Hep) termed as Trio. This trisaccharide motif is linked to the lipid A through a linear glucan-heptan domain (Li *et al.*, 2018, 2017).

Lipid A, the core-oligosaccharide domain and the Trio were proposed to be conserved among strains and are important for host colonisation, whereas the Lewis antigen, heptan and glucan are variable and non-essential for colonisation (**Figure 1.4**) (Li *et al.*, 2017).

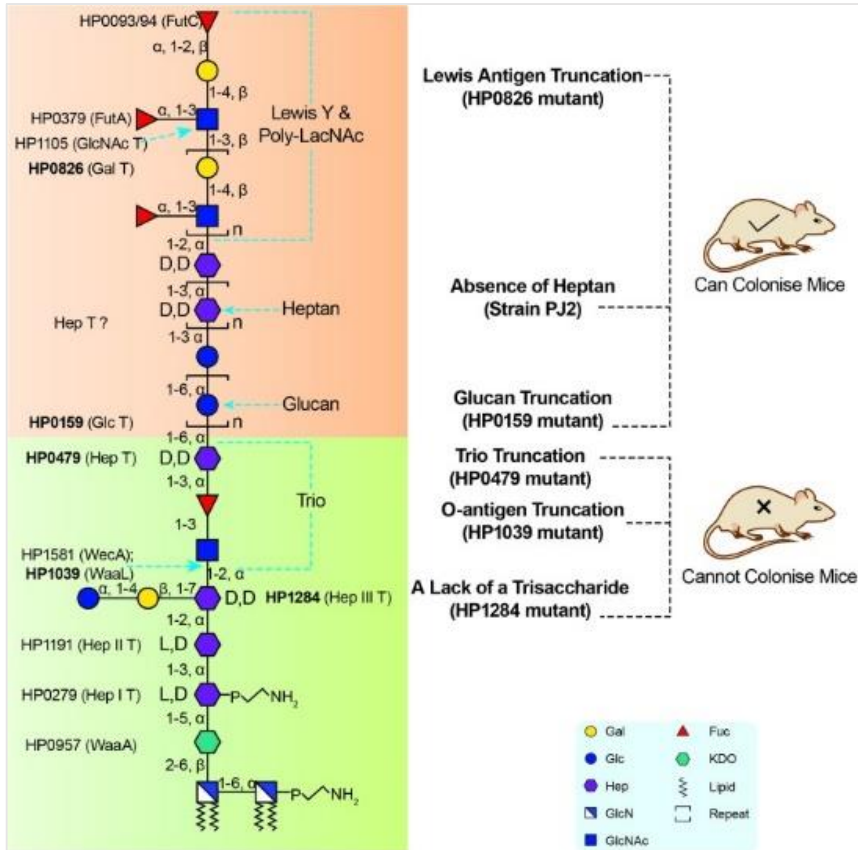


Figure 1.4 LPS structure (left) and importance of its components (right) (Li *et al.*, 2017).

1.3.2.2 Biosynthesis of LPS

The biosynthesis of the LPS begins in the cytoplasm with the transference by WecA of the GlcNAc residue of the Trio onto the UndPP carrier (**Figure 1.5**). The assembly of the O-antigen is then followed by the recruitment of successive glycosyltransferases and the addition of the conserved Trio and variable glucan, heptan and Lewis antigens. Once the O-antigen is completely assembled, it is translocated by flippase Wzk to the periplasm. In parallel, the core-lipid A which is also assembled in the cytoplasm is translocated by flippase Msb to the periplasm as well. Once in the periplasmic space, the core-lipid A suffers modification through dephosphorylation and deacylation, and it is ligated with the O-antigen (being the third heptan residue the attachment site) by ligase WaaL (Li *et al.*, 2018, 2017).

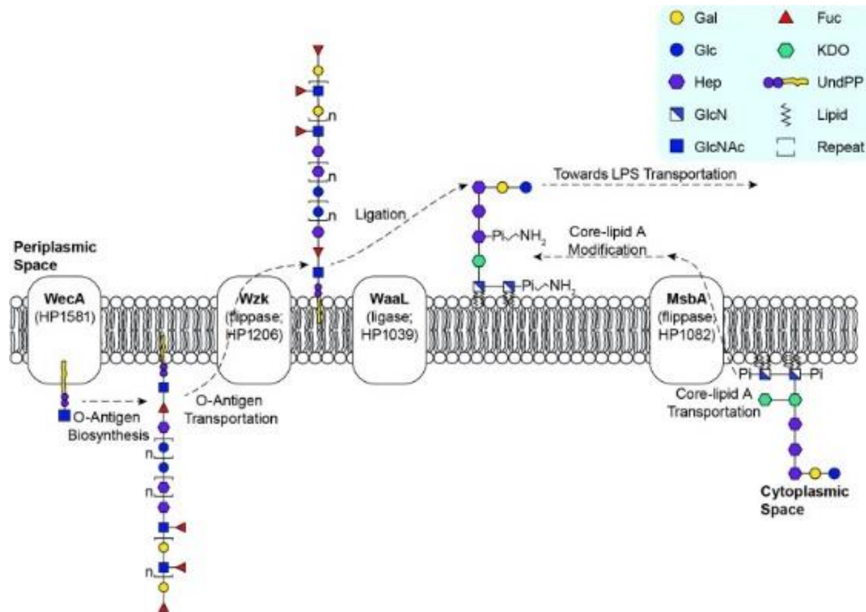


Figure 1.5 Proposed model for the LPS biosynthetic pathways in *H. pylori* (Li *et al.*, 2017).

1.4 Epidemiology

H. pylori has been solely found in humans and other primates. It is hypothesised that when ancestral mammals diverged from reptiles, around 150 million years ago, they were already carrying on their stomachs *Helicobacter* strains, which kept changing with their host. Consequently, the human-adapted strains evolved forming the *H. pylori* species (Blaser *et al.*, 2008).

Human infection by *H. pylori* is an important public health threat as it colonizes the gastric mucosa of approximately half of the world's population. The infection is usually acquired in infancy and early childhood. It is long-lasting and often lasts a lifetime (Quaglia and Dambrosio, 2018). Some epidemiological studies suggested the presence of *H. pylori* in contaminated water supplies, and it has been detected from well water, saliva and feces, although it has been rarely possible to culture from these sources (Testerman *et al.*, 2006).

H. pylori was classified as a class I (definite) carcinogen by the International Agency for Research on Cancer (IARC, 1994), being gastric cancer the third most common cause of cancer-related deaths worldwide. Individuals suffering from *H. pylori* infection were found to have a six-fold higher risk of developing gastric cancer and up to a 50% higher risk for all gastric cancers as compared to uninfected individuals (Meliț *et al.*, 2019). Even though it exists a very strong association between infection and disease, only approximately 10% of infected patients will develop peptic ulcer disease, and less than 1–2% of patients will develop gastric adenocarcinoma as a result of the infection, even though cancer risk varies widely by geographical location (Piqué *et al.*, 2016).

It's important to emphasize that the clinical outcome of *H. pylori* infection toward severe diseases (ulcers, gastric cancer) is the result of a complex interplay between bacterial virulence factors, host immune response and alimentary factors. Therefore, focusing the studies only on virulence factors may sometimes be insufficient (Leja *et al.*, 2016).

H. pylori infection prevalence depends on various factors, including geographic area, age, race, ethnicity and socio-economic status (Brown, 2000). Studies done in both developed and developing countries have shown that low socio-economic status is closely related to the acquisition of *H. pylori* infection, with special emphasis on the socioeconomic status during childhood. The factors that define socio-economic status and which, when being insufficient, lead to an increase in the acquisition of the infection within the population are low levels of education, high density of living, lack of sanitation and low hygiene levels (Eusebi *et al.*, 2014; Mitchell and Katelaris, 2016).

1.4.1 Transmission

The most likely form of human-to-human transmission is either oral (through vomitus or possibly saliva) or fecal-oral. Person-to-person transmission is supported by the increased incidence of infection in institutionalized children and adults and the clustering of *H. pylori* infection within families (Brown, 2000). This concept is also supported by the detection of *H. pylori* DNA in vomitus, saliva, dental plaque, gastric juice and feces.

Transmission through water, probably due to fecal contamination, can be an important source of infection, especially in areas of the world where untreated water is widespread. Knowledge of the epidemiology and mode of transmission of *H. pylori* is important to prevent its spread and may be useful to identify high-risk populations, particularly in areas with high prevalence of gastric lymphoma, gastric cancer and gastric ulcers.

A study by Cervantes and colleagues (2010) about transmission between siblings at the US-Mexican border showed that *H. pylori* infection first occurred in the older siblings and later in the younger ones, but never vice-versa. They identified that having an *H. pylori*-infected older sibling was associated with an increased rate of getting a persistent *H. pylori* infection in a younger sibling. In addition, persistent infections were more likely to be transmitted between siblings closer in age. A 17-fold increased rate of persistent infections was observed when the sibling age difference was three years or less. Mothers also play a key role in transmitting *H. pylori* infection to their children (Mitchell and Katelaris, 2016). Studies show an increased risk of infection of children with an infected mother. Furthermore, the relative risk of getting the infection was 5.3 times more in children with a *H. pylori*-positive mother than with a *H. pylori*-negative mother (Malaty *et al.*, 2000).

1.4.2 Mixed infections

A person may be infected by one or more strains of *H. pylori*. Multiple infection indicates the presence of two or more genetically distinct isolates in a single patient. Microevolution involves an infection with different strains from the same genetic isolate, where a strain derives from the other under selective pressure. The genome of *H. pylori* changes continuously during chronic colonization by importing small pieces of foreign DNA from other *H. pylori* strains during long-term or transient mixed infections (Suerbaum and Michetti, 2002). When the first colonizer strain spreads around the gastric mucosa, it confronts with multiple selective pressures. Therefore, in order to survive them, *H. pylori* is able to generate genetic diversity within the population and to

adapt to this changing environment. A high mixed infection rate was predicted as a risk factor for the development of disease symptoms (Kibria *et al.*, 2015).

These mutations are difficult to find, and probably influence the efficacy of the eradication treatment. Their detection can be performed using different techniques: genotyping of virulence genes –*cagA, vacA, iceA*–, random amplified polymorphism DNA, Multilocus Sequence Typing (MLST) and Whole Genome Sequencing (Mansour *et al.*, 2016). The ratios of mixed infections vary from 0% to 85% in different populations around the world (Lai *et al.*, 2016; Mansour *et al.*, 2016).

H. pylori has long been known to be genetically highly diverse following a panmictic (non-clonal) population structure as a result of horizontal gene transfer and frequent recombination (Langenberg *et al.*, 1986; Palau *et al.*, 2016). It is also genetically more diverse than most bacterial species (Achtman *et al.*, 1999; Falush *et al.*, 2003; Go *et al.*, 1996). Sequence analyses of these multiple housekeeping genes has shown, with very few exceptions, that each isolate of *H. pylori* possesses unique sequences (Achtman *et al.*, 1999). *H. pylori* isolates possess remarkably high genetic heterogeneity and are diverse in different geographic areas (Lai *et al.*, 2016).

1.4.3 *Helicobacter pylori* as a marker for human migrations

H. pylori population structure provides strong evidence of ancient ancestry in Africa and subsequent human migrations. *H. pylori* traveled with humans through their migrations to all corners of the world, and has remained intimately associated with its human host population ever since (Moodley *et al.*, 2009). Accumulated evidence points out to human migration out of Africa to the Middle East circa. 60,000 to 150,000 years ago, and then independently to Europe and Asia. *H. pylori* and its human host underwent variation driven by founder effects, isolation, drift and host selection. As a result, *H. pylori* became a marker of human migrations (Dominguez-Bello and Blaser, 2011).

MLST was first proposed in 1998 as a tool for the epidemiological study of bacteria (Maiden *et al.*, 1998). Serial studies of extensive number of *H. pylori* strains, including strains isolated from aboriginal populations, show that MLST analysis provides more

detailed information on human migration than the analysis of human genetics does (Yamaoka, 2009).

Several models of population genetics have been proposed in the last 20 years. Achtman and colleagues (1999) were the first to apply MLST using seven housekeeping genes (*atpA*, *efp*, *mutY*, *ppa*, *trpC*, *ureI*, and *yphC*) to a collection of 20 strains of *H. pylori* from distinct regions of the world, also including sequences of two virulence-associated genes (*cagA* and *vacA*). Their analysis discriminated between an Asian clone and other uncharacterized clones (**Figure 1.6**). They also hypothesized about the age of *H. pylori*, suggesting a history of millions of years and proving that the bacterium already accompanied the *Homo sapiens* when colonizing Asia (Achtman *et al.*, 1999).

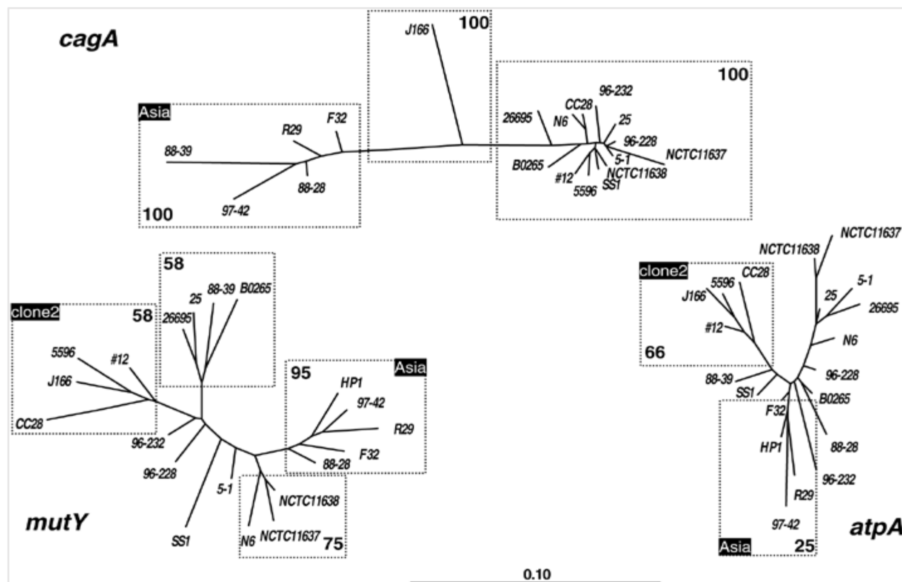


Figure 1.6 Phylogenetic trees of three genes discriminating between an Asia clone and other uncharacterized clones (Achtman *et al.*, 1999).

The group of Falush (2003) repeated the experiment, this time with 370 strains isolated worldwide, the seven housekeeping genes which had been used by Achtman, and the virulence-associated gene *vacA*. A total of 3,850 nucleotides were sequenced from each isolate, and from these, 1,418 nucleotides were polymorphic and were used to define bacterial populations. A Bayesian approach identified four modern populations, labelled hpAfrica1, hpAfrica2, hpEastAsia and hpEurope (**Figure 1.7.A**). Additional

experiments divided the population hpEastAsia into the hspAmerind, hspEAsia and hspMaori subpopulations, and the population hpAfrica1 into hspWAfrica and hspSAfrica (**Figure 1.7.B**). They also defined five ancestral populations, which were named ancestral Africa1, Africa2, EastAsia, Europe1 (AE1) and Europe2 (AE2) (**Figure 1.7.C**).

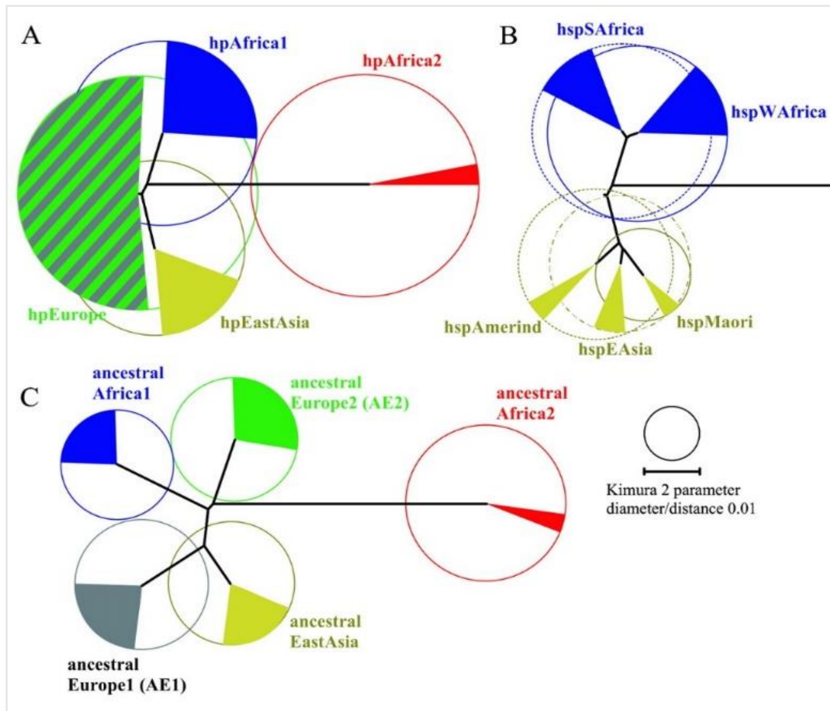


Figure 1.7 Relationships between *H. pylori* populations (Falush *et al*., 2003) **A**. Modern populations **B**. Modern subpopulations **C**. Ancestral populations.

The ancestral population tree points out that Africa2 evolved before the other populations split and that AE1 and ancestral EastAsia diverged from each other more recently (**Figure 1.7.C**). In this model, the circles in modern populations (**Figure 1.7.A**) are larger, reflecting a larger genetic distance due to the admixture between ancestral populations, e.g., modern hpEurope population reflects an admixture of the ancestral Europe1 and Europe2 (Falush *et al*., 2003).

Taking into account the preceding discussion, *H. pylori* has the potential to be informative about human migrations. Falush *et al.* (2003) drew a map showing the specific migrations of humans and *H. pylori* populations, expressing the entry of the population AE1 with the migratory fluxes that brought Uralic speakers to Europe, and

the entrance of AE2 via the arrival of Neolithic farmers to Europe from the Near East. Once the modern hpEurope population was merged, it expanded to different parts of the continent. The hpEastAsia ancestral population spread out with the migration of Austronesian people, which reached New Zealand after sequential island-hopping. It also spread to America with the colonization of the Americas, that began more than 12,000 years ago. Ancestral Africa1 population was expanded due to the migration of slaves and also due to the rapid expansion of Bantu farmers (**Figure 1.8**).

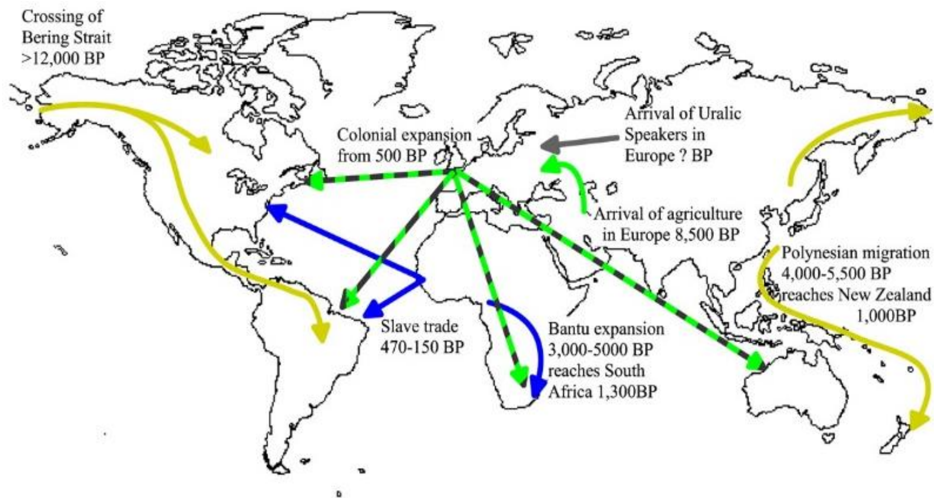


Figure 1.8 Specific migrations of *H. pylori* populations. BP, years before present (Falush *et al.*, 2003).

Linz and collaborators (2007) also built their own approach to the genetic of populations of *H. pylori.* They examined 769 isolates from a total of 51 ethnic sources. The sequences from the seven housekeeping gene fragments were concatenated to form a 3406-bp haplotype, of which 45% (1,552 bp) was polymorphic. This fact reflects the remarkably high frequency of recombinants in this bacterium. Among the already described populations and subpopulations, they identified two new populations and designated them hpAsia2 and hpNEAfrica (**Figure 1.9**). hpAsia2 was isolated in northern India, Thailand, Bangladesh, the Philippines and some other regions in southeastern Asia. hpNEAfrica was predominant among isolates from Ethiopia,

Somalia, Sudan and Nilo-Saharan speakers in northern Nigeria (Linz *et al.*, 2007; Suerbaum and Achtman, 2004).

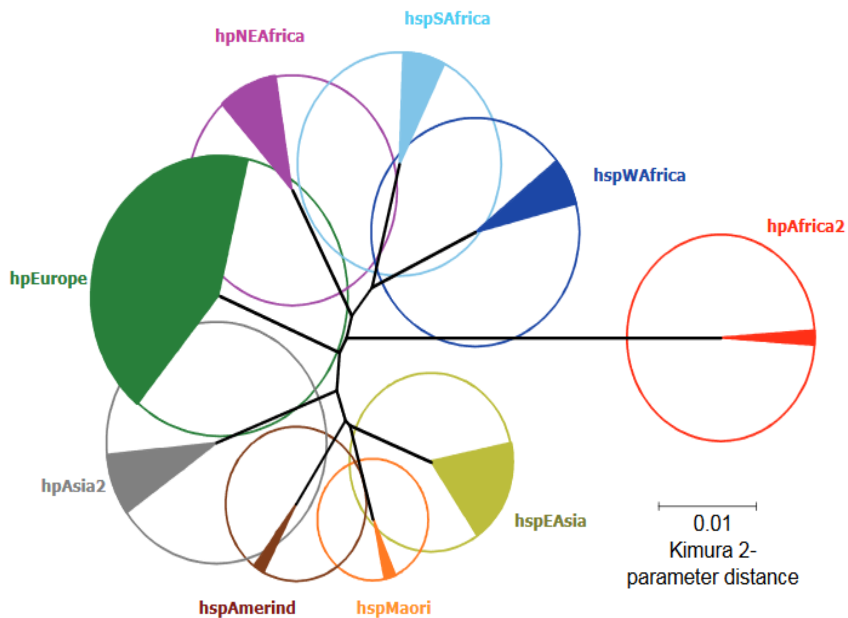


Figure 1.9 Phylogenetic tree with all the new populations and subpopulations (Linz *et al.*, 2007).

More recently, Moodley *et al.* (2009) described the ancestral and modern population hpSahul, which split from the Asian population circa. 35,000 years ago and was the responsible of colonizing the Pacific, specifically New Guinea and Australia.

The actual phylogenetic tree reveals the presence of seven modern populations of *H. pylori* that cluster accordingly to the origin of the bacterium and its host. These seven modern populations are hpAfrica1, hpAfrica2, hpNEAfrica, hpSahul, hpAsia2, hpEurope and hpEastAsia (**Figure 1.10**), which have derived from six ancestral populations which were denoted ancestral Africa1, ancestral Africa2, ancestral Sahul, ancestral EastAsia, ancestral European 1 (AE1) and ancestral European 2 (AE2) (Suzuki *et al.*, 2012).

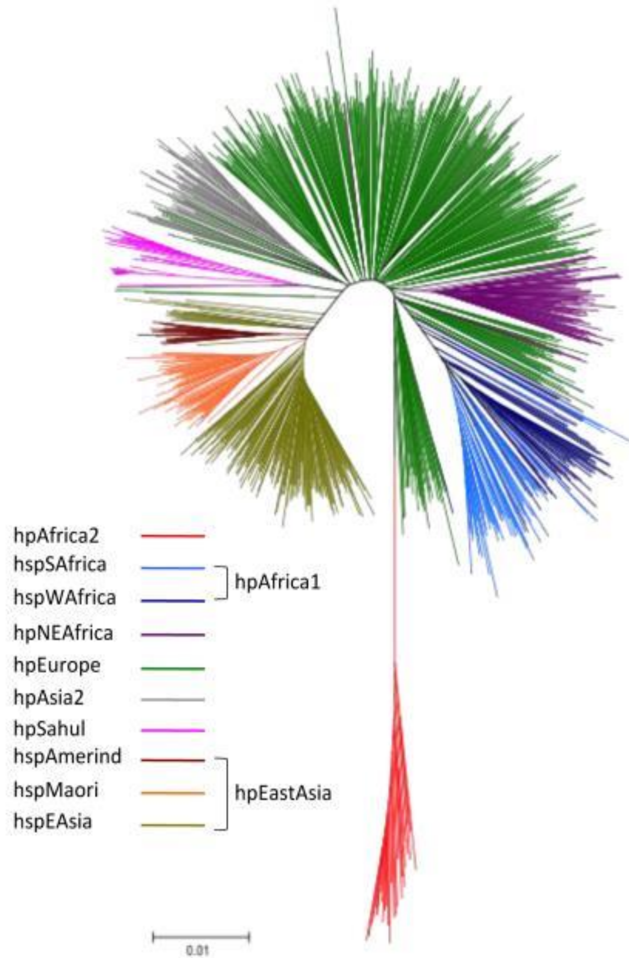


Figure 1.10 Actual *H. pylori* populations (Suzuki *et al.*, 2012).

1.5 Diagnosis, treatment, and antibiotic resistances

1.5.1 Diagnosis

An accurate diagnosis of *H. pylori* infection pre- and post- treatment is a crucial part in the effective management of many gastroduodenal diseases. In other words, to treat the disease appropriately, it is necessary to have suitable diagnostic procedures (Atkinson and Braden, 2016; Sabbagh *et al.*, 2019; Wang *et al.*, 2015).

Although many diagnostic tests are now available, each method has its advantages, disadvantages and limitations. The choice of one method or another may depend on the availability and accessibility of diagnostic tests, the resources of laboratories, the clinical condition of patients and the likelihood of using positive and negative tests in different clinical circumstances (Wang *et al*., 2015). These methods must meet current standards of clinical diagnosis, such as precision, sensitivity and specificity. Additionally, they should also be applicable in areas of development where hygiene standards and medical care are lacking (Sabbagh *et al*., 2019). A differentiation can be made between invasive (endoscopic-based) and non-invasive diagnostic methods (**Table 1.1**).

Invasive diagnostic methods	Non-invasive diagnostic methods
Endoscopic image	Urea breath test
Histology*	Stool antigen test
Rapid urease test*	Serological examination
Culture*	Molecular examination
Molecular methods*	

Table 1.1 Diagnostic tests are classified into invasive and non-invasive methods. *From gastric biopsies (Sabbagh *et al*., 2019).

Invasive methods such as endoscopy are indicated in patients with an outbreak of dyspeptic symptoms and which are above a locally agreed age cut-off (older than 45 in the European guidelines), in patients with alarming symptoms (weight loss, dysphagia, gastrointestinal bleeding, iron-deficient anemia, abdominal mass or persistent vomiting) and in patients who do not respond properly to therapy (Atkinson and Braden, 2016).

Should a patient be under any treatment which makes use of proton pump inhibitors (PPI), this treatment should be stopped two weeks before the endoscopy. In most situations, biopsies from the antrum are sufficient, and biopsies from the body will be necessary when the patient has not stopped the treatment with proton pump inhibitors on time (Atkinson and Braden, 2016). The rapid urease test is another valid alternative; it detects the activity of the urease enzyme in biopsy samples.

Non-invasive techniques are the choice in young patients with uncomplicated dyspepsia and without alarming symptoms, being ^{13}C-urea breath test and the stool antigen test the preferred ones, with a sensitivity and specificity higher than 90% (Atkinson and Braden, 2016). The urea breath test, with prior administration of citric acid, is considered the best method before and after the eradication treatment. Stool antigen test is the alternative technique when the urea breath test is not available. The detection of the antigen in the stool is carried out with a monoclonal method (Sánchez Delgado *et al.*, 2018). Serological methods detect IgG antibodies to *H. pylori* and can show as high accuracy as other non-invasive and invasive biopsies, but do not differentiate between active or past *H. pylori* infections (Sabbagh *et al.*, 2019). It is recommended to repeat the diagnostic test after the eradication treatment, to prove whether the treatment has been successful or not.

Compared to *H. pylori*, there is no other chronic gastrointestinal infection with a comparable collection of diagnostic techniques. Despite the availability of multiple diagnostic methods, it remains unclear which method to select as the gold standard for detecting *H. pylori* infection, especially in epidemiological studies (Sabbagh *et al.*, 2019).

1.5.2 Treatment

The decision to treat or not a *H. pylori* infection should be made in an active manner, taking into account the circumstances and risks of each patient. Current treatment consists in a combination of 2–3 antimicrobial agents, such as amoxicillin, tetracycline, clarithromycin, metronidazole, or levofloxacin and antisecretory agents, such as proton pump inhibitors (PPI). There is currently no single optimal treatment against the infection. For many years, the first-line regimen has relied worldwide on the standard triple therapy, consisting of a PPI and two antibiotics (amoxicillin –or metronidazole– and clarithromycin) for 7 to 10 days treatments (Sánchez Delgado *et al.*, 2018). However, the high rate of resistance of *H. pylori* to clarithromycin and metronidazole has undoubtedly decreased the eradication rate of this therapy. Therefore, the treatment recommendations have changed towards the use of new quadruple guidelines with or

without bismuth, combined with higher doses of proton pump inhibitors and the extension of treatments to 14 days (**Table 1.2**) (Sánchez Delgado *et al.*, 2018). In general, the resistance to amoxicillin tends to be low. Thus, a dual therapy consisting of a PPI and amoxicillin can be an effective first-line or rescue therapy. As a second-line treatment regimen, levofloxacin-based quadruple therapy can be used after failure of first-line therapy (Mitchell and Katelaris, 2016; Yang *et al.*, 2014). In case of the treatment failing three times, therapy focuses on the use of the antibiotic rifabutin.

Moreover, use of probiotics can reduce side effects caused by the exposure to antibiotics and help with *H. pylori* eradication, due to their anti-inflammatory and anti-oxidative properties (Goderska *et al.*, 2018). In addition, the concomitant use of alternative medicine (polyphenol-rich drinks like red wine or green tea, honey, garlic, etc.) has the potential to provide additive or synergistic effects against *H. pylori* infection, though its efficacy still needs to be verified in clinical studies (Ayala *et al.*, 2014).

Therapy	Duration
Classical quadruple therapy (Pylera®)[a]	10 days
High-dose PPI/12 h	
Pylera® 3 capsules/6 h	
Concomitant quadruple therapy	14 days
High-dose PPI/12 h	
Amoxicillin 1 g/12 h	
Clarithromycin 500 mg/12 h	
Metronidazole 500 mg/ 12 h	
Quadruple therapy with Levofloxacin	14 days
High-dose PPI/12 h	
Levofloxacin 500 mg/12 h	
Amoxicillin 1 g/12 h	
Bismuth subsalicylate 120 mg, 2 tablets/ 12 h	
Quadruple therapy with Rifabutin	14 days
High-dose PPI/12 h	
Rifabutin 150 mg/12h	
Amoxicillin 1 g/12h	
Bismuth subsalicylate 120 mg, 2 tablets/ 12 h	

Table 1.2 Dosage and duration of recommended therapies (modified from Sánchez Delgado *et al.* 2018). [a]Pylera® integrates bismuth, tetracycline and metronidazole in a single capsule.

1.5.3 Antibiotic resistances

The eradication of *H. pylori* infection has been unsuccessful in recent years due to a constant increase in the antimicrobial resistance (Meliț *et al.*, 2019) (**Table 1.3**). Actually, *H. pylori* is included in the *WHO priority list for research and development of new antibiotics* with a priority level 2, which is considered high, and just below the level 1 category, which is deemed as critical (World Health Organization, 2017) (**Figure 1.11**). Antibiotic resistance is a key factor in the failure of eradication therapy and recurrence of *H. pylori* infection.

Country	Patients	Year of the study	AMX	CLR	MTZ	LEV
Iceland	Adults	2012-2013	0%	9%	1%	4%
Germany	Children	2002-2015	0,80%	23,20%	28,70%	8,90%
France	Adults	2014	ND ND	Pre: 22,2% Post: 73,9%	Pre: 45,9% Post: 78,3%	Pre: 15,4% Post: 14,8%
Poland	Adults	1998-1999 2013-2014	0% 0%	9% 31%	36% 83%	ND ND
Bulgaria	Adults	2010-2015	4%	28,10%	33,80%	19,40%
Italy	Adults	2010-2014	1-1,2%	72,3-72,3%	34,7-42,2%	42,8-53%
Spain	Children	2007-2014	8,50%	51,20%	39,20%	6,70%
USA	ND	2011-2016	1,20%	70,40%	82,40%	53,50%
Iran	Adults	2014	27,70%	43,10%	73,80%	13,40%
Singapore	Adults	2000-2014	3-4,4%	7,9-17,1%	24,8-48,2%	5-14,7%
China	Children	2009-2015	0.06%	16,40%	75,20%	6,70%
China	Adults	2009-2014	<0.01%	17,80%	95,50%	19,70%
Indonesia	Adults	2012-2015	5,20%	9,10%	46,70%	31,20%
Morocco	Adults	2015-2016	0%	29%	40%	11%

Table 1.3 Percentage of antimicrobial resistance in *H. pylori* in several recent studies, according to the country, patient population and year of the study (modified from Alba *et al.*, 2017). ND: not described. Pre: pre-treatment. Post: post-treatment. AMX: amoxicillin. CLR: clarithromycin. MTZ: metronidazole. LEV: levofloxacin. All values referenced to EUCAST (European Committee on Antimicrobial Susceptibility Testing) standards.

Priority 1: CRITICAL

Acinetobacter baumannii, carbapenem-resistant

Pseudomonas aeruginosa, carbapenem-resistant

***Enterobacteriaceae**,** carbapenem-resistant, 3rd generation cephalosporin-resistant

Priority 2: HIGH

Enterococcus faecium, vancomycin-resistant

Staphylococcus aureus, methicillin-resistant, vancomycin intermediate and resistant

Helicobacter pylori, clarithromycin-resistant

***Campylobacter*,** fluoroquinolone-resistant

***Salmonella** spp.*, fluoroquinolone-resistant

Neisseria gonorrhoeae, 3rd generation cephalosporin-resistant, fluoroquinolone-resistant

Priority 3: MEDIUM

Streptococcus pneumoniae, penicillin-non-susceptible

Haemophilus influenzae, ampicillin-resistant

***Shigella** spp.*, fluoroquinolone-resistant

* Enterobacteriaceae include: *Klebsiella pneumonia*, *Escherichia coli*, *Enterobacter* spp., *Serratia* spp., *Proteus* spp., and *Providencia* spp, *Morganella* spp.

Figure 1.11 Global priority list for research and development of new antibiotics (World Health Organization, 2017).

The antibiotic **amoxicillin** (AMX) is one of the key components of many *H. pylori* eradication therapies. Most *H. pylori* isolates are still susceptible to it, but amoxicillin resistance is increasing in clinical isolates. In Gram-negative bacteria, resistance to β-lactams is mostly due to the production of β-lactamases, either chromosomally encoded, or plasmid borne (Bush, 2001). In contrast, alterations in penicillin-binding proteins (PBPs) have mainly been described in β-lactam-resistant Gram-positive bacteria, but they have also been reported in some Gram-negative bacterial species. Following this patron, in naturally occurring amoxicillin-resistance (Amx^R*)* *H. pylori* isolates, resistance is mediated by various mutational changes in or adjacent to the second and third PBP-

motifs of the *pbp1A* gene (Gerrits *et al.*, 2006). In addition to the already known mutations in PBP1, Rimbara *et al.* (2008) identified the mutations in PBP2 and PBP3 that were detected only in Amx^R or low-susceptibility strains. Multiple mutations in PBP2 and PBP3, in addition to mutations in PBP1, confer higher amoxicillin resistance in *H. pylori.* A study by Nishizawa *et al.* (2011) revealed that resistance to amoxicillin in *H. pylori* was gradually induced after unsuccessful eradication attempts. The data are clearly consistent with the association of resistance rates and eradication failures. If Amx^R *H. pylori* strains would further spread, serious problems would arise, resulting in increasing eradication failures.

In the last 20 years, increasing rates of *H. pylori* resistance to macrolides such as **clarithromycin** (CLR) have been reported (Megraud *et al.*, 2011). Resistance to clarithromycin is due to three point mutations in the gene encoding the 23S ribosomal subunit –23S rDNA– of *H. pylori*; these mutations are A2142G, A2143G and A2142C and may influence resistance levels. Specifically, A2142G point mutations are related to higher minimum inhibitory concentrations (MICs); A2143G is often associated with lower MICs; and A2142C point mutations are rarely seen so no statement can be written on the CLR resistance level transmitted (Beckman *et al.*, 2017; Redondo *et al.*, 2018).

At least four different mechanisms of resistance to **metronidazole** (MTZ) in *H. pylori* have been described (**Figure 1.12**): (1) a reduction in the antibiotic uptake and an enhancement of the efflux pump through the bacteria wall, (2) reduced activity of the nitroreductases (involving the mutations in the genes *rdxA*, *frxA* and *fdxB*), (3) an increase in the activity of the oxygen-radical scavenger system and (4) an increase in the activity of the DNA repair enzymes (Alba *et al.*, 2017; Lee *et al.*, 2018).

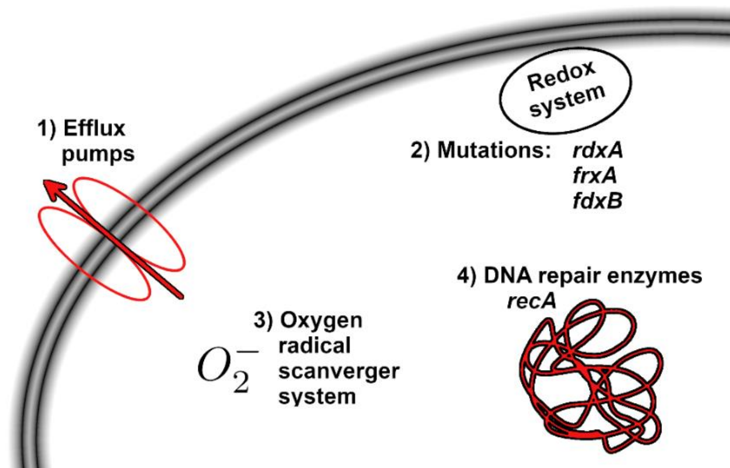


Figure 1.12 Resistance mechanism to metronidazole. Own elaboration.

The broad-spectrum fluoroquinolone **levofloxacin** (LEV) can be used in the rescue treatment to eradicate *H. pylori* infection. This antibiotic inhibits DNA gyrase and thus alters DNA synthesis. Mutations were identified in the *gyrA* DNA gyrase (subunit A or B) (position D91 changes to G, N, A or Y, N87K and A88V) or in the *gyrB* gene (D481E or R484K) in *H. pylori*-resistant isolates (Alba *et al.*, 2017; Hu *et al.*, 2016).

1.6 Pathogenicity

Thanks to the presence of urease, its mobility and its ability to attach to the gastric epithelium, *H. pylori* can survive in the stomach. Among the factors that are advantageous to the successful bacterial colonization in the gastric epithelium are the shape of the bacterium, polar-sheathed flagella, directional motility, chemotaxis, adherence and persistence (Sabbagh *et al.*, 2019; Waskito *et al.*, 2018).

During the 20th century, accumulated evidence has showed that gastric cancer tends to occur in stomachs already affected by chronic inflammation, particularly atrophic gastritis with its associated hypochlorhydria, and that gastric cancer is a consequence and not only a concomitant aspect of the gastritis (Moss, 2017). Indeed, three cohort studies led to the classification of *H. pylori* as a class I carcinogen in 1994

(Forman *et al.*, 1991; Nomura *et al.*, 1991; Parsonnet *et al.*, 1991). In each cohort, evidence of prior *H. pylori* infection was found to be significantly more common in those subjects who subsequently developed gastric cancer compared with a sample of those who had not (Moss, 2017).

The spectrum of clinical diseases associated with *H. pylori* is wide, ranging from gastritis and peptic ulcer to gastric adenocarcinoma and MALT (mucosa associated lymphoid tissue) lymphoma (Suerbaum and Michetti, 2002). The development of gastric cancer is thought to be the accumulation of several alterations, like DNA damage, change in the proliferation and apoptosis of the cell and degradation of tumor suppressors (Cover, 2016).

Concerning the pathogenesis of the disease, multiple virulence factors have been described, such as urease enzyme, flagella, adhesins, cytotoxin-associated gene A (*cagA*), vacuolating cytotoxin A (*vacA*), and the induced by contact with epithelium gene A (*iceA*) (Abu-Taleb *et al.*, 2018). Further detail about these virulence factors is provided in subsequent subsections. In the literature can be found different studies, conducted over the world, addressing the prevalence of virulence factor genes (*vacA*, *cagA*, *cagE*, *oipA*, *babA2*, *babB* and *iceA*) by trying to associate them to clinical outcomes. Pinto-Ribeiro *et al.* (2016) investigated this relationship in 290 patients from Macau, China. The allelic variants *vacA-i1* and *vacA-m1* were detected in 82.5% and 53.6% of the patients that were infected with single genotypes. The prevalence of *cagA*-positive strains was 97.5%. No significant association was observed between *vacA* genotypes or *cagA* and gastric carcinoma. Among the infected patients, 37.5% of them had coexistence of *H. pylori* strains with different *vacA* genotypes (Pinto-Ribeiro *et al.*, 2016). Dabiri *et al.* (2017) detected in 160 *H. pylori* strains isolated from Tehran patients, *cagA*, *cagE*, *oipA*, *iceA1*, *babA2* and *babB* in 69%, 51%, 55%, 26%, 78% and 28% of isolates, respectively. There was no association between the *cagA*, *vacA*, *cagE* or *iceA* status and clinical outcome in patients (Dabiri *et al.*, 2017). Only *babB* and *iceA1* were significantly associated with higher risk of gastric cancer. Additionally, Su *et al.* (2016)

described a higher seropositivity of the adhesins AlpA, OipA, BabA and SabA in patients with gastric cancer than in patients without. Therefore, they were considered as gastric cancer related virulence markers (Su *et al.*, 2016). These results highlight the fact that the relationship between virulence and the bacterium has not always been very clear.

This discrepancy illustrates the controversial results obtained in the last 30 years by trying to establish definitive virulence markers for *H. pylori*. Various studies have found different genes to better match different clinical outcomes. Additionally, there are clinical studies confirming or denying the role of virulence markers in establishing a connection between virulence markers and the clinical outcome (Pormohammad *et al.*, 2018; Yamaoka and Graham, 2014).

Host and environmental factors also are important determinants of *H. pylori*–associated disease risk. On the one hand, it has been shown that a high consumption of fresh fruits, vegetables, enough iron and cholesterol protect against GC. On the other hand, smoked, preserved foods, red meat, and a high level of dietary salt intake might increase the risk of gastric cancer (Cover and Blaser, 2009; González *et al.*, 2006; Raei *et al.*, 2016).

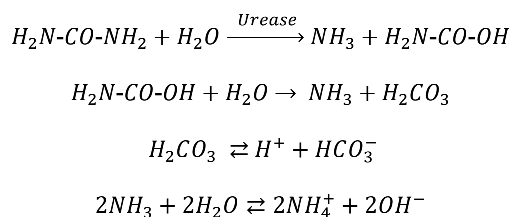
1.6.1 Virulence factors

1.6.1.1 Urease enzyme

Urease activity seems to be essential to the survival and pathogenesis of *H. pylori*, as it allows the bacteria to survive in an acidic milieu (Suerbaum and Michetti, 2002). To date, all *H. pylori* isolated produce large quantities of this enzyme. Through urease the hydrolysis of urea is initiated by generating ammonia which neutralizes the stomach acids and creates the proper pH environment that the bacterium requires to survive and colonize. Urease causes the decomposition and reabsorption of urea, which leads to an abnormal rise in environmental pH and induces a variety of diseases in the human body, i.e., cell damage and inflammation by ammonia (Dunn *et al.*, 1997). The enzyme activity is regulated by UreI, a pH-gated urea channel, which opens at low pH and stops the influx of urea in neutral conditions (Weeks *et al.*, 2000). With the production of ammonia

it is also achieved a solubilization of the mucus gel and therefore, bacterial motility is facilitated (Salama *et al.*, 2013).

The complete reaction of urease encompasses four steps (**Figure 1.13**). First of all, urea is decomposed into one molecule of ammonia and carbamate by the effect of urease. Carbamate is then naturally resolved to another molecule of ammonia and carbonic acid. In water, this carbonic acid is in equilibrium, and the two previous ammonia molecules now become protonated and yield ammonium and hydroxide ions. This leads to an increase in pH of the gastric lumen (Burne and Chen, 2000).



$$H_2N\text{-}CO\text{-}NH_2 + H_2O \xrightarrow{Urease} NH_3 + H_2N\text{-}CO\text{-}OH$$

$$H_2N\text{-}CO\text{-}OH + H_2O \rightarrow NH_3 + H_2CO_3$$

$$H_2CO_3 \rightleftarrows H^+ + HCO_3^-$$

$$2NH_3 + 2H_2O \rightleftarrows 2NH_4^+ + 2OH^-$$

Figure 1.13 Urease reaction. Modified from (Burne and Chen, 2000).

1.6.1.2 Adhesins

Adhesins, bacterial proteins with host cell adhesive properties, mediate in almost all bacterial pathogenesis (Patel *et al.*, 2017). *H. pylori* expresses multiple proteins that act as epithelial cell adhesins, such as BabA, OipA or SabA (**Figure 1.14**). One of the most studied adhesins is the protein encoded by *babA2*, BabA, which binds to Lewis b blood-group antigen (Le^b) found on the gastric epithelial cells. Not all *H. pylori* strains express BabA. Those who express BabA, a correlation with increased risk for gastric cancer has been found (Israel and Peek, 2010). Another *H. pylori* adhesin is OipA, whose expression induces inflammation and cause IL-8 cytokine secretion in host cells. However, the host receptor for OipA has not been yet identified (Teymournejad *et al.*, 2017). On the other hand, SabA binds to the sialylated glycan Le^x. The capability of many *H. pylori* strains to adhere to sialylated glycoconjugates expressed during chronic inflammation might therefore strengthen virulence and the extraordinary chronicity of *H. pylori* infection (Mahdavi *et al.*, 2002).

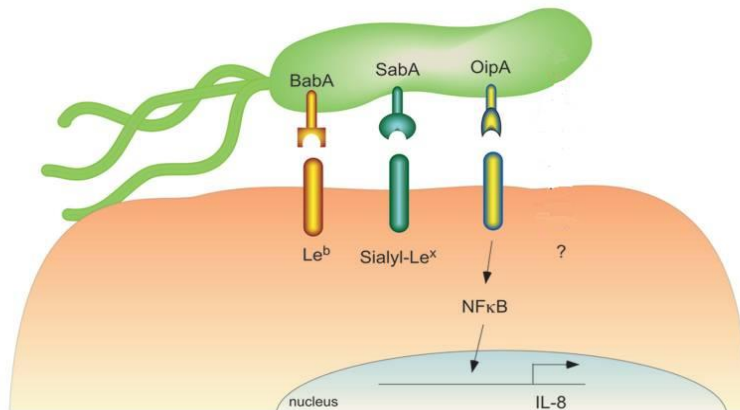


Figure 1.14 Adhesins and host cell receptors. Modified from (Israel and Peek 2010).

1.6.1.3 Induced by contact with epithelium (*iceA*)

The gene *iceA* shows two different alleles, *iceA1* and *iceA2*. The expression of *iceA1* is associated with peptic ulcer disease, enhanced mucosal IL-8 expression and acute antral inflammation. Furthermore, it is enhanced by contact between *H. pylori* and human epithelial cells. On the other hand, *iceA2* is associated with asymptomatic gastritis and non-ulcer dyspepsia, although its function remains unclear (Abu-Taleb *et al.*, 2018; Huang *et al.*, 2016).

1.6.1.4 Vacuolating cytotoxin A (VacA)

VacA is a high-molecular-weight multimeric pore-forming protein, present in all *H. pylori* strains, that possesses no similarity to any other known bacterial or eukaryotic protein. Its ability to persist in the stomach is facilitated through the formation of vacuole-like membrane vesicles in the cytoplasm of gastric cells, the suppression of macrophages and T cells, and the stimulation of apoptosis (Ali *et al.*, 2015; Weng *et al.*, 2019). In more detail, CD4 T cells are one of VacA's targets, causing the suppression of the transcription factor nuclear factor of activated T cells and their proliferation. VacA also restricts antigen presentation by B cells and interrupts the normal behaviour of CD8 T cells. VacA targets as well macrophages and mast cells (Cover and Blaser, 2009)

In relation to its structure, VacA is a polymorphic protein, for which three gene regions of variation have been described: the signal (*s*), intermediate (*i*) and middle (*m*) regions (**Figure 1.15**) (Jones *et al.*, 2010).

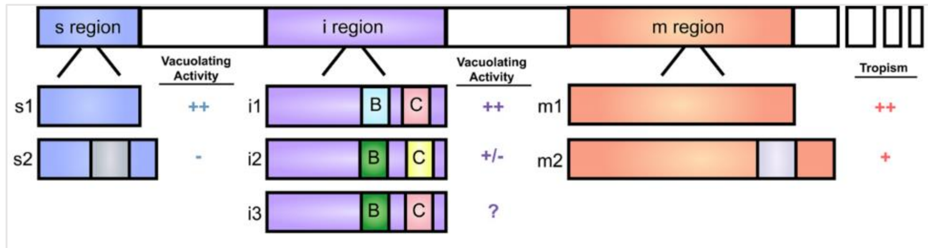


Figure 1.15 Allelic regions of *vacA* (Jones *et al.*, 2010).

The *s* region of *vacA* influences the vacuolating activity and the efficiency of anion channel formation. The s2 variant has an extension of 12 hydrophilic amino acids. The VacA-s1 contains more hydrophobic aminoacids near the cleavage site than the VacA-s2, so the s1 variant is more easily inserted into the host cell membrane. Thus, the presence of the toxigenic allele *vacA-s1* in *H. pylori* strains is commonly associated with an increased risk of peptic ulceration and gastric cancer (Jones *et al.*, 2010; Kim and Blanke, 2012). The *m* region determines host cell tropism, and the *m1* variant is toxic to a broader range of host cells. The *i* region has been proposed to be the best indicator of severity. It is located between the *s*- and *m*- regions, and *vacA-i1* strains are strongly associated with gastric adenocarcinoma and peptic ulcer disease. Taking this into account**,** strains with the *s1*, *i1* and *m1* alleles are generally associated with more severe disease (Jones *et al.*, 2010; Palframan *et al.*, 2012). The *vacA* genotype is recognized as a crucial element in the cellular activity of the toxin (Kim and Blanke, 2012).

1.6.1.5 Cytotoxin-associated gene A (CagA)

H. pylori can be differentiated into *cagA*+ and *cagA*- strains depending on the presence or absence of the *cag* pathogenicity island (*cag*PAI) that encodes for a type IV secretion system which effector is CagA (Hashi *et al.*, 2018). CagA is the only known effector protein to be injected into host cells (Akopyants *et al.*, 1998). Multiple studies have analysed its role, concluding that CagA takes a crucial part in gastric

carcinogenesis (Hashi *et al.*, 2018; Jones *et al.*, 2010). The C-terminal region of CagA contains multiple Glu-Pro-Ile-Tyr-Ala (EPIYA) tyrosine-phosphorylation motifs. The sequences are annotated according to the amino acid sequence surrounding the EPIYA motifs. There are four distinct EPIYA segments, namely EPIYA-A, -B, -C and -D (**Figure 1.16**) (Hashi *et al.*, 2018; Hashinaga *et al.*, 2016).

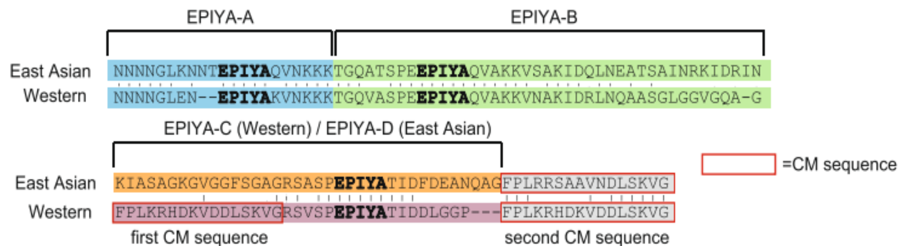


Figure 1.16 EPIYA-repeat region amino acid sequence differences between East Asian and Western types. Modified from (Hashi *et al.*, 2018).

Two subtypes of CagA can be distinguished, the East Asian CagA and the Western CagA. The Western CagA strains carry EPIYA-A, EPIYA-B and between one and three EPIYA-C segments. On the other hand, the East Asian CagA strains contain EPIYA-A, EPIYA-B and EPIYA-D segments (**Figure 1.17**) (Hatakeyama, 2017).

Once inside the host cell cytoplasm, it interacts with different signaling molecules, for example the pro-oncogenic phosphatase SHP-2 and the polarity-regulating kinase PAR1b. CagA binds to SHP-2 through the tyrosine-phosphorylated EPIYA-C or EPIYA-D segments, causing a deregulation of its catalytic activity and promoting, consequently, pro-oncogenic cellular events. CagA binds to PAR1b via the CagA multimerization (CM) sequence, a 16 amino acid sequence. Western subtypes contain at least two CM motifs, one in the N-terminal part of EPIYA-C segment and the other located immediately downstream of EPIYA-C. In contrast, the East Asian CagA possesses only one CM motif, located downstream of EPIYA-D (**Figure 1.17**). Finally, CagA is able to bind to the membrane phospholipid thanks to the K-Xn-R-X-R motif, located upstream of the EPIYA region (Hashi *et al.*, 2018; Hatakeyama, 2017).

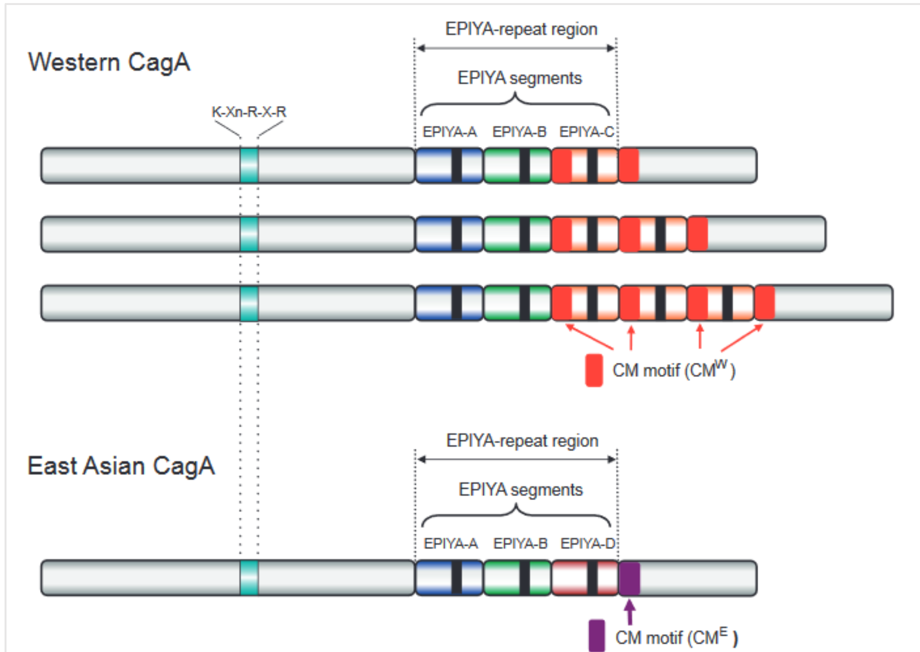


Figure 1.17 Graphic representation of the CagA protein (Hatakeyama, 2017).

1.6.1.6 CagA and VacA interaction

The bacteria that can reach the cell surface alter the nature of the gastric cells through CagA, VacA and CagL (**Figure 1.18**). On the one hand, interactions between the phosphorylated and the non-phosphorylated CagA with different molecules inside the cell cytoplasm lead to multiple consequences for the cell, such as dephosphorylation of cellular proteins, perturbation of epithelial cell differentiation, and modified cell signalling, motility and proliferation. CagA in conjunction with VacA disrupt the polarity of the cell, leading to iron acquisition and cell extrusion. The protein CagL takes a very important part in this series of events. It binds to the host receptors integrins α5β1 and αvβ5. Then, and in combination with CagA, it causes the induction of gastrin and the repression of the H^+/K^+ ATPase. Therefore, the episodes of hypergastrinaemia and hypochlorydria are observed during chronic infection. Moreover, CagA also sets off the synthesis of the inflammatory cytokine interleukin-8 (Cover and Blaser, 2009; Salama *et al.*, 2013).

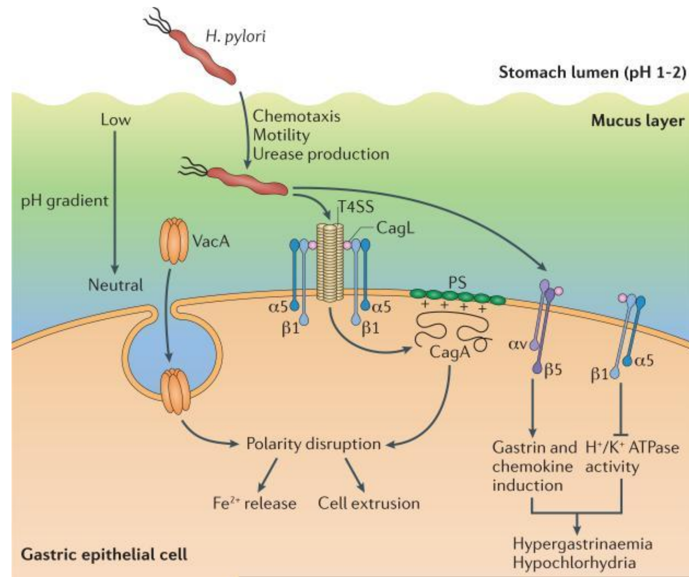


Figure 1.18 Depiction of *H. pylori* colonization of a gastric cell. The different interaction mechanism are shown, with special emphasis on CagA and VacA (Salama *et al.*, 2013).

1.6.2 Correa pathway

The Correa pathway, or precancerous cascade, is the succession of histological changes that *H. pylori* infection produces before cancer becomes clinically apparent. It was described by Correa (Correa *et al.*, 1975) and has five well differentiated steps (**Figure 1.19**). The first stages of gastric cancer are most of the time asymptomatic or associated with non-specific symptoms such as nauseas or heartburn (dyspepsia) (Correa, 2013).

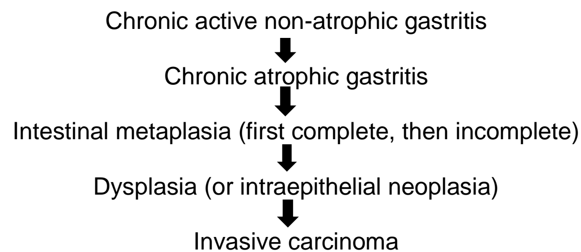


Figure 1.19 Steps in the Correa pathway (Correa, 2013).

The location and type of gastritis is associated with the forthcoming risk of developing gastric pathologies. Patients with antral gastritis are predisposed to have duodenal ulcers due to an abnormal acidity during long periods of time. On the other hand, patients with corpus gastritis probably will mostly develop gastric ulcers, gastric atrophy, intestinal metaplasia and gastric carcinoma (**Figure 1.20**) (Suerbaum and Michetti, 2002).

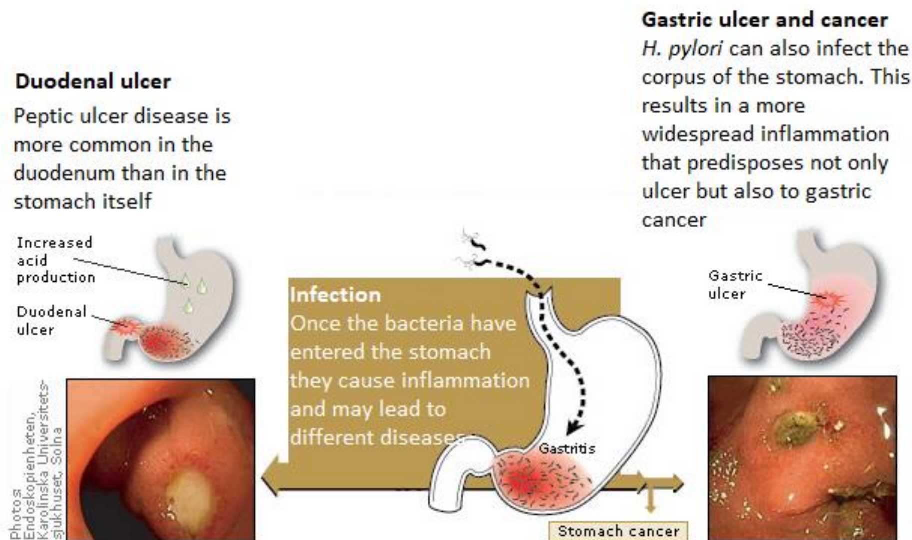


Figure 1.20 Pathogenic outcomes from *H. pylori* infection. Modified from NobelPrize.org (2005c).

1.6.2.1 Chronic active non-atrophic gastritis

The first stage is called “non-atrophic” because it doesn’t show loss of gastric glands, and “active” because polymorphonuclear neutrophils can be found, meaning that an acute inflammation is present. This phase is also distinguished by diffuse infiltration of the gastric mucosa by white blood cells (lymphocytes, plasma cell and macrophages), representing chronic inflammation (**Figure 1.21**) (Correa and Houghton, 2007). In most cases, this kind of gastritis is restricted to the antrum, and as mentioned above, antral gastritis results most of the times in duodenal ulcers, which don’t usually lead to atrophy nor metaplasia.

1.6.2.2 Chronic atrophic gastritis

Atrophy, the loss of glands, takes place around the *incisura angularis*, in the antrum-corpus junction. Cell loss is related to the effects of bacterial products and the cytokine environment in the gastric mucosa. Gaps left by lost glands are then filled with fibrous tissue that, in company with the subsequent loss of cellular cross-talking, can coordinate the influx of blood borne stem cells. These are responsible for tissue changes which can lead finally to cancer. With time, this process can extend to the anterior and posterior gastric walls (Correa, 2013; Suerbaum and Michetti, 2002) (**Figure 1.21**).

1.6.2.3 Intestinal metaplasia

At this point of the pre-cancerous cascade, the original structure of the gastric epithelium is replaced by cells with intestinal phenotype (intestinal metaplasia). The metaplastic intestinal cells first look like the small intestine mucosa: eosinophilic absorptive enterocytes with a well-developed brush border, alternating with well-developed goblet cells. This type of metaplasia receives the name of ‘intestinal complete metaplasia’. At later stages of the pre-cancerous process, the metaplastic cells start losing their small intestinal phenotype and develop and acquire morphological features of large intestine phenotype. This type is called ‘incomplete’ or ‘colonic’ because it is only composed of goblet cells of different sizes and shapes (Correa *et al.*, 2010).

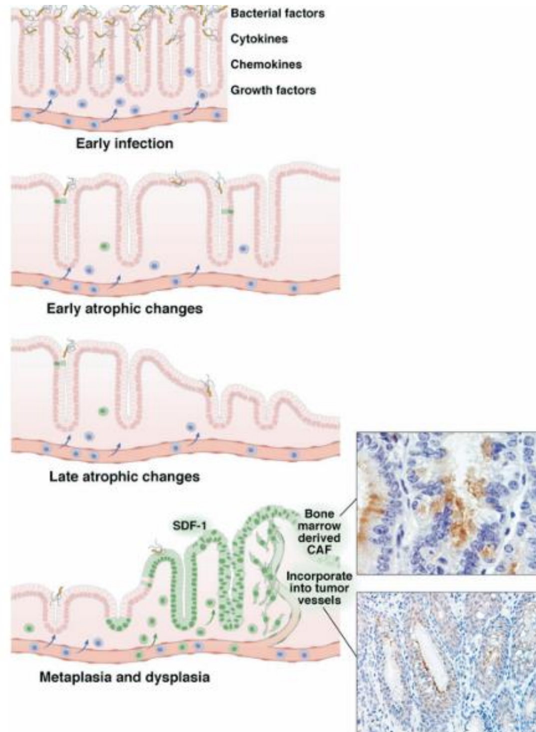


Figure 1.21 Evolution of the histological changes in the gastric mucosa (Correa and Houghton 2007).

1.6.2.4 Dysplasia

Dysplasia is defined as an unequivocal neoplastic non-invasive epithelial alteration (Lauwers and Riddell, 1999). The nuclei of a dysplastic epithelium are enlarged, irregular in shape and lacking of polarity (Correa and Houghton, 2007). These changes are not just limited to the deeper glands, they also extend to the surface epithelium. Dysplasia is firstly low-grade, with a low degree of nuclear atypia and architecture distortion. Later it develops focal points of high-grade with rising rates of nuclear polymorphism and irregular architecture, which increase the cancer risk (Correa, 2013; Correa and Houghton, 2007).

1.6.2.5 Invasive carcinoma

When the dysplastic atypical cells cross the basal membrane, they become invasive carcinomas. The main hypothesis sustaining the appearance of carcinogenesis is likely

owing to oxidative stress, represented by the expression of inducible nitric oxide synthase caused by the infection (Atherton, 2006).

1.6.2.6 Extragastric manifestations

The affectation region of *H. pylori* is not just limited to the gastric environment, but also to extragastric points. Several studies have described disorders of the cardiovascular system, and metabolic, neurological, dermatological, ocular, or allergic diseases. For instance, the infection by *H. pylori* was linked to systemic inflammation and therefore, a higher risk of mortality in patients with chronic obstructive pulmonary disease (COPD) (Sze *et al.*, 2015). Still, some data are controversial. Different studies found a positive correlation between the infection and stroke, while others suggested a lower mortality in people infected by *H. pylori* (Gravina *et al.*, 2018). Similarly, controversial conclusions have been reached with regard to *H. pylori* infection and multiple sclerosis in different studies (Franceschi *et al.*, 2015; Mohebi *et al.*, 2013; Trang *et al.*, 2016); some authors show an inverse association between the infection and the disease while others have described a positive relation.

1.7 Is *Helicobacter pylori* beneficial or detrimental?

Currently, when the infection with *H. pylori* is detected, a treatment is automatically prescribed, but only 10 to 15% of those infected will suffer consequences. Is it the most adequate action to confront all your microbiota to three or four antibiotics during 10 days without knowing if your infection by *H. pylori* will ever develop any disease? Some beneficial and protective effects of the infection have been shown.

There is no doubt that the prevalence of gastric cancer in North America and Western Europe has declined over the years due to the treatment of the infection, high levels of sanitation, and the reduction of direct contact between people. Nevertheless, collaterally, an increase in esophageal adenocarcinoma has been detected in the same population. The presence of Barret’s esophagus and gastroesophageal reflux disease have also been inversely correlated with the presence of *H. pylori* (Labenz *et al.*, 1997;

Richter *et al.*, 1998). This is explained by the fact that uninfected people have a more acidic gastric pH, making it more harmful compared to the more basic pH of infected people. This fact perhaps may be considered as the “unplanned drawback” of the disappearance of *H. pylori* from Western stomachs (Cover and Blaser, 2009; Moss, 2017).

Other scientists have been working on the protective role of the bacterium against asthma and allergy in children. It has been studied that as the predominance of *H. pylori* has receded, the prevalence of asthma and allergic disorders in children have expanded. In other words, scientists showed that children who acquired *H. pylori* in childhood were less likely to present allergy symptoms and asthma (Chen and Blaser, 2007).

Several researchers studied its protective function against diarrheagenic gastrointestinal infections, suggesting that it may protect the body from infections by exogenous intestinal pathogens, a probable hypothesis to explain this fact would be that the immune system is activated by the production of antibacterial peptides, produced by *H. pylori* and the host (Cover and Blaser, 2009; Rothenbacher *et al.*, 2000).

Ghrelin, known as the hunger hormone, is 65% produced in the stomach, and studies have shown that individuals infected by the bacteria produce lower levels than people without *H. pylori* in their stomachs. Because ghrelin has effects throughout the body, it is likely to have long-term metabolic consequences affecting the height and weight (Cover and Blaser, 2009).

To conclude, maybe it is time to stop thinking of *H. pylori* as an absolute pathogen and to start thinking of it as a normal member of the human microbiota that, under some circumstances yet to be completely established, can yield to pathogenic states of the gastric mucosa.

CHAPTER 2 **OBJECTIVES**

The main objectives of this PhD have been the following:

1. To evaluate the usefulness of housekeeping genes in the detection of *H. pylori* infection, in the genetic discrimination at strain level, and in the detection of multiple infection.
2. To characterize the events of multiple infection and microevolution by means of amplicon sequencing of housekeeping genes in gastric biopsies.
3. To analyse and compare the whole genome sequences of three *H. pylori* strains isolated from the tumoral and the non-tumoral tissue obtained from a single patient with gastric cancer, and to find possible relevant differences between them.

CHAPTER 3 **Usefulness of housekeeping genes for the diagnosis of *Helicobacter pylori* infection, strain discrimination and detection of multiple infection**

Main contents of this chapter have been published in an International Journal:

Palau M, Kulmann M, Ramírez-Lázaro MJ, Lario S, Quílez ME, Campo R, Piqué N, Calvet X, Miñana-Galbis D. Usefulness of housekeeping genes for the diagnosis of Helicobacter pylori infection, strain discrimination and detection of multiple infection. Helicobacter 2016; 21 (6): 481–487. doi: 10.1111/hel.12304

3.1 Introduction

Helicobacter pylori chronically infects more than half of the world's population because the immune response of the host is unable to eliminate the infection. In most infected people, the bacterium acts as a commensal organism inducing chronic asymptomatic gastritis that can last for life. In other cases, however, it is responsible for a heavy toll of morbidity and mortality as a consequence of peptic ulcers and gastric cancer. Chronic gastritis may progress to intestinal metaplasia, dysplasia and eventually gastric cancer. This multi-step process is known as the Correa pathway (Correa and Houghton, 2007). The clinical outcome of *H. pylori* infection depends on strain virulence, host response, and environmental factors.

Several virulence factors have been identified in *H. pylori*: gastric colonization mediated by flagella, urease, outer membrane proteins (OMPs), phospholipids, glycolipids, and other adhesins (*babA*, *hpA*, *napA*, *sabA*, etc.), toxins such as lipopolysaccharide (LPS) and the vacuolating cytotoxin VacA, a type IV secretion system encoded by the cag pathogenicity island (*cag*PAI) containing the effector protein CagA, a γ-Glutamyl transpeptidase (GGT) associated with colonization and cell apoptosis, the duodenal ulcer-promoting gene A (*dupA*), and others (Backert and Clyne, 2011; Mascellino *et al.*, 2009; Mobley *et al.*, 2001). Among virulence factors identified in *H. pylori*, *cagA* and *vacA* genes are the main virulence markers as shown in a study conducted by González *et al.* (González *et al.*, 2011), in which patients infected with high virulence strains (*cagA*+, *vacA* s1i1m1) had a higher risk of progression to preneoplastic lesions and gastric cancer in comparison with patients infected with low virulence strains.

Strains of *H. pylori* exhibit considerable genetic diversity following a panmictic (non-clonal) population structure due to horizontal gene transfer and frequent recombination. In contrast, since *H. pylori* does not spread epidemically, phylogenetic studies based on multilocus sequence analysis (MLSA) of housekeeping genes are contributing to trace human migrations (Linz *et al.*, 2007; Moodley *et al.*, 2009). On the other hand,

intrapatient diversity of *H. pylori* needs to be re-evaluated because results about the prevalence of multiple infection are controversial, and it could have implications related to the mode of transmission, antibiotic resistance and virulence of *H. pylori* (Mobley *et al.*, 2001; Sheu *et al.*, 2009; Talarico *et al.*, 2009). Moreover, there is a lack of studies regarding *H. pylori* detection from specimens other than gastric biopsies (Khalifa *et al.*, 2010).

Other genes have likewise been related to *H. pylori* pathogenesis: *amiA* (N-acetylmuramoyl-L-alanine amidase) and *luxS* (autoinducer-2 synthase) genes are involved in biofilm formation although the former is also related to bacterial adhesion, cell morphology and immune escape (Chaput *et al.*, 2006). Others such the cholesterol-α-glucosyltransferase gene (*cgt* or *capJ*) glycosylates host cholesterol contributing to pathogenicity and antimicrobial resistance while *cpn40* (*dnaJ*), *cpn60* (*groEL*) and *cpn70* (*dnaK*) genes have been suggested as a risk factor for oncogenesis and are recommended as serological markers of *H. pylori* infection (Di Felice *et al.*, 2005; Mascellino *et al.*, 2009; McGee *et al.*, 2011).

In this study, *H. pylori* specific PCR amplification and sequencing of *amiA*, *cgt*, *cpn60*, *cpn70*, *dnaJ*, and *luxS* genes have been developed in order to evaluate their usefulness in the detection of *H. pylori* infection, to achieve genetic discrimination at strain level, to detect cases of multiple infection and to perform MLSA.

3.2 Material and Methods

3.2.1 *Helicobacter pylori* clones and DNA extraction

Ten *H. pylori* strains (APP134, B247, B271, B319, B355, B491, B508S, B508T, CRL122 and SVC135) from the *H. pylori* collection of the Digestive Diseases Department of the Hospital Taulí (Sabadell, Barcelona) were included in this study. Outpatients sent to the Endoscopy Unit of the Hospital Taulí for evaluation of dyspeptic symptoms from February 2006 to November 2015 were recruited in this collection. Patients were contacted prior to the endoscopy and were asked to participate. Before

the endoscopy, the patients signed an informed consent. During endoscopy, antral and corpus biopsies were obtained for histology, rapid urease test (RUT) and molecular studies. Isolation, culture and identification of *H. pylori* were performed after a positive RUT test. The RUT biopsy was plated on Pylori Agar (bioMérieux) in microaerophilic conditions in microaerophilic jars (Jar Gassing System; Don Whitley Scientific Limited, UK). After a maximum of a week, *H. pylori* isolates were subcultured on Columbia plates (bioMérieux) and identified by colony morphology, Gram-negative staining and a positive result for urease, catalase, and oxidase tests. The strains were frozen in Brucella Broth with 10% glycerol and stored at –80ºC until analysis (Lario *et al.*, 2012).

The strains were recovered on Columbia agar with 5% sheep blood (bioMérieux) and incubated at 37ºC under microaerophilic conditions. These strains were previously isolated from antral biopsies of different patients, except for B508S and B508T, which were isolated from the same patient, B508S from normal tissue and B508T from gastric adenocarcinoma (**Table 3.1**). For this study, these *H. pylori* strains were not considered as pure cultures because they were originally obtained by picking up most of the growth on Pylori agar (bioMérieux) after seeding of gastric biopsies, and not from individual colonies. In order to obtain pure cultures, between three and six isolated colonies were selected from each blood agar culture.

Further *H. pylori* colonies were also isolated from antral and corpus biopsies of two different patients (B657A / B657C and B659A / B659C) (**Table 3.1**) and seeded onto Pylori agar.

Gastric biopsies (n = 14)	**Endoscopic / Histopathological diagnosis**	**Clones analysed (n = 52)**
APP134	duodenal ulcer / moderate gastritis	APP134-1, -2, -3
B247[a]	neoplasia / adenocarcinoma	B247-1, -2, -3
B271	duodenal ulcer / moderate gastritis	B271-1, -2, -3
B319	normal / mild gastritis	B319-1, -2, -3
B355	duodenal ulcer / mild gastritis	B355-1, -2, -3
B491[a]	neoplasia / adenocarcinoma	B491-1, -2, -3

B508S[a]	neoplasia / adenocarcinoma	B508S-1, -2, -3, -4, -5, -6
B508T	neoplasia / adenocarcinoma	B508T-2A, -2B, -3, -4, -5, -6
B657A	normal / mild gastritis	B657A-1, -2, -3, -4
B657C	normal / mild gastritis	B657C-1, -2, -3, -4
B659A	normal / moderate gastritis	B659A-1, -2, -3, -4
B659C	normal / moderate gastritis	B659C-1, -2, -3, -4
CRL122	normal / mild gastritis	CRL122-1, -2, -3
SVC135	normal / moderate gastritis	SVC135-1, -2, -3

Table 3.1 Samples included in this study. [a]Gastric biopsies obtained from normal tissue of patients with adenocarcinoma.

H. pylori strains ATCC 49503 and ATCC 51932 were also included as controls in the present study.

Subcultures of individual colonies were performed on Columbia blood agar or Brucella agar (BD Diagnostics, Franklin Lakes, NJ, USA) supplemented with 10% fetal bovine serum (FBS; Invitrogen, Waltham, MA, USA) and incubated 7 days at 37ºC in microaerophilic conditions. Several subcultures were done, reaching 52 clones.

3.2.2 DNA extraction, PCR amplification and sequencing

DNA of each clone was extracted from 4 ml of the cultures in Brucella broth supplemented with 10% FBS following the manufacturers' instructions (REAL pure genomic DNA extraction kit; Durviz, Paterna, València). Briefly, cultures were centrifuged, 80 µl of Lysis Solution was added to the sediment and samples were incubated at 80ºC for 5 min. 3 µl of RNase was added and after an incubation of 30 min, 300 µl of Protein Precipitation Solution was added and vigorously vortexed for 30 s. After 7 min of centrifugation at 14,000 rpm, the supernatant was transferred to a new tube, cleaned with 600 µl of isopropanol, then with 600 µl of ethanol, and finally resuspended with 100 µl of ultrapure water.

Primers for *H. pylori* specific PCR amplification and sequencing of the housekeeping genes *amiA*, *cgt*, *cpn60*, *cpn70*, *dnaJ* and *luxS* (**Table 3.2**) were designed from 43 *H. pylori* complete genome sequences (**Table AI.1**). To evaluate the specificity of the PCR assays, the following four strains, belonging to different bacterial species, were included

as negative controls: *Campylobacter jejuni* ATCC 33291, *Pseudomonas aeruginosa* 42A2, *Aeromonas molluscorum* 848T[T] and *Aeribacillus pallidus* DR03.

Gene	Length (pb)[a]	Primer	Sequence 5'→3'	*Ta*	Amplicon (pb)
amiA	1323	amiA-658	GTTTTRGACGCYGGGCATGG	64ºC	635
		amiA-1292	CCATCAGCAATGCCCTTAGC		
cgt	1170	cgt-252	GGCTTTTAAGGGAGCGGATA	60ºC	615
		cgt-866	ATCGCTTCGCTYTCCACATT		
cpn60	1641	Hp156	CGTGAGCGTGGCTAARGAG	54ºC	801
		Hp956	GCTTTGCCTAAAAACTCYACTT		
cpn70	1863	cpn70-982	ATTTCAGAAGTGGTGATGGT	54ºC	612
		cpn70-1593	GTTTCTCGCTTCAATCACTT		
dnaJ	1110	dnaJ-259	TTTGAAGATTTAGGCTCGTT	54ºC	588
		dnaJ-846	TAAAGAYGGCACTTTAATCG		
luxS	468	luxS-38	TGGATCACACYAAAGTCAAAG	54ºC	429
		luxS-466	AAACCCCCACTTCAGACCA		

Table 3.2 Primers designed for specific PCR amplification and partial sequencing of *H. pylori* housekeeping genes. *Ta*, annealing temperature. [a]Data from the complete genome of *Helicobacter pylori* 26695 (NC_000915).

PCR amplifications were carried out in a total volume of 50 µl containing 50 mM KCl, 15 mM Tris/HCl, pH 8.0, 2.5 mM $MgCl_2$, 0.2 mM dNTPs (Amersham Biosciences, Little Chalfont, Buckinghamshire, UK), 2.5 U AmpliTaq Gold DNA polymerase (Applied Biosystems), 40 pmol of each primer (Sigma, St. Louis, MO, USA) and 250 ng DNA. The reaction mixtures were subjected to the following thermal cycling program in a 2720 Thermal Cycler (Applied Biosystems): denaturation at 95 ºC for 5 min, 35 cycles of 95 ºC for 1 min, 54–64 ºC (**Table 3.2**) for 1 min and 72 ºC for 1 min and a final extension step at 72 ºC for 5 min.

The amplified products were purified using either the MSB® Spin PCRapace kit (Stratec, Birkenfeld, Germany) or ExoSAP-IT® (Affymetrix, Santa Clara, CA, USA) and sequencing was performed using the ABI PRISM BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) with the following thermal cycling conditions: 96 ºC

for 1 min and 25 cycles of 96 ºC for 10 s, 50 ºC for 5 s and 60 ºC for 4 min. Nucleotide sequences were determined in an ABI PRISM 3730 DNA analyser by the Genomics Unit of Scientific and Technological Centers from University of Barcelona (CCiTUB).

3.2.3 Data analyses

Partial sequences of *amiA*, *cgt*, *cpn60*, *cpn70*, *dnaJ*, and *luxS* genes were aligned independently and phylogenetic trees were constructed by neighbor-joining method and Jukes–Cantor distance estimation model using MEGA 6.0 (Tamura *et al.*, 2013). Concatenated sequences of all genes obtained by the online sequence toolbox FaBox (Villesen, 2007) were also analysed (**Table 3.3**). The topological robustness of the phylogenetic trees was evaluated by a bootstrap analysis through 1,000 replicates.

Gene	Sequence (nt)	Distances values		
		Range	Minimum[a]	Mean ± SE
amiA	576	0−0.052	0.019	0.031 ± 0.004
cgt	558	0−0.044	0.009	0.025 ± 0.004
cpn60	555	0−0.039	0.009	0.023 ± 0.004
cpn70	588	0−0.040	0.015	0.027 ± 0.004
dnaJ	564	0−0.061	0.013	0.032 ± 0.004
luxS	405	0−0.078	0.028	0.046 ± 0.006
concatenated	3246	0−0.036	0.024	0.030 ± 0.002

Table 3.3 Genetic distance values of individual and concatenated genes. [a]Minimum distance values between sequences from clones isolated from biopsies of different patients.

3.3 Results

A total of 52 *H. pylori* clones were isolated from 14 gastric biopsies from 11 patients (**Table 3.1**). Concentration measured with Nanodrop ranged from 39,7 to 375,4 ng/μl, and even with the lower concentration values, PCR and sequencing reactions were successful.

3.3 Results

All housekeeping genes (*amiA*, *cgt*, *cpn60*, *cpn70*, *dnaJ*, and *luxS*) were specifically amplified for the bacterial species *H. pylori* with the primers designed in this study and no gene was amplified for the other bacterial species analyzed. Amplicons and partial gene sequences were 429–635 nt and 405–588 nt long, respectively (**Table 3.2** and **Table 3.3**). The GenBank/EMBL/DDBJ accession numbers for the *amiA*, *cgt*, *cpn60*, *cpn70*, *dnaJ* and *luxS* gene sequences are KU053341–58, KU053359–76, KU053377–94, KU053395–412, KU053413–30 and KU053431–48, respectively.

We compared the gene sequences from the 52 *H. pylori* clones, *H. pylori* ATCC 49503 and ATCC 51932. All gene sequences from clones isolated from biopsies of different patients were different, their gene distances ranged from 0.9% to 7.8%, and the highest distance values were obtained for *luxS*. Distance rates calculated from concatenated sequences ranged from 2.4% to 3.6%, with a mean distance of 3% (**Table 3.3**).

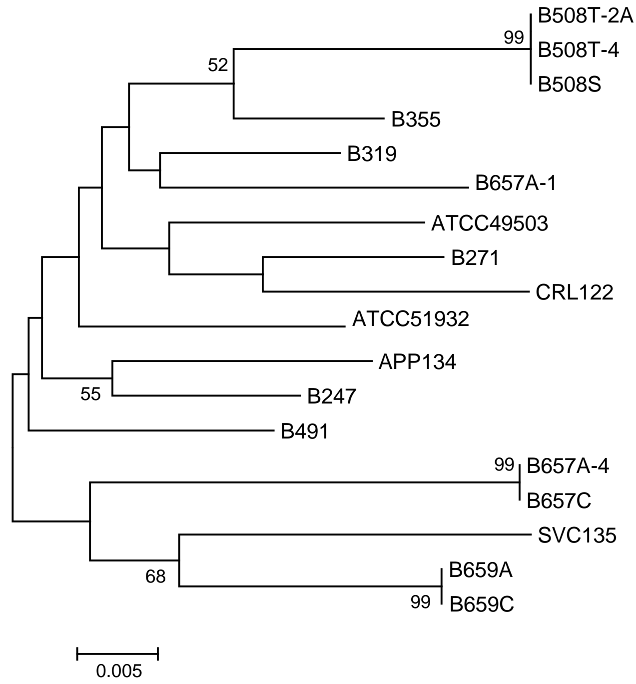
All clones isolated from the same patient showed identical sequences except in the case of B508S/B508T and B657A/B657C. Two different *amiA* and *cpn60* sequences were detected from B508T clones: B508T-2A, B508T-2B, B508T-3, B508T-5 and B508T-6 showed identical sequences, which were different to those from clone B508T-4. A unique nucleotide difference (0.2% divergence) was detected in the gene *amiA*, meanwhile six nucleotide differences (1.1% divergence) were found in *cpn60*. B508S clones presented identical *amiA* and *cpn60* sequences to those from B508T-4. However, on the basis of *cgt*, *cpn70*, *dnaJ*, and *luxS* sequence comparison, all B508T and B508S clones analysed were identical. Phylogenetic trees for *amiA*, *cpn60*, *luxS*, *cgt*, *cpn70*, and *dnaJ* are shown in **Figure 3.1** A, B, C, D, E, and F, respectively.

76 B508S (=B508S1–6)
99 B508T-4
B508T-2A (=-2B,-3,-5,-6)
B355 (=B355-1,-2,-3)
B271 (=B271-1,-2,-3)
B491 (=B491-1,-2,-3)
APP134 (=APP134-1,-2,-3)
SVC135 (=SVC135-1,-2,-3)
ATCC 49503
B247 (=B247-1,-2,-3)
B319 (=B319-1,-2,-3)
B657A-1 (=-2,-3)
B659A (=B659A-1,-2,-3,-4)
99 B659C (=B659C-1,-2,-3,-4)
ATCC 51932
CRL122 (=CRL122-1,-2,-3)
B657A-4
99 B657C (=B657C-1,-2,-3,-4)
0.005

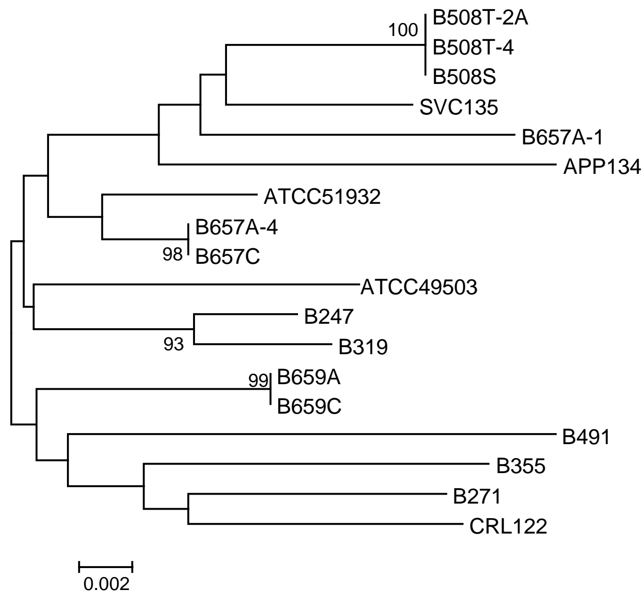
(A) *amiA*

99 B508T-4
57 B508S
B508T-2A
ATCC49503
B657C
99 B657A-4
B247
CRL122
B355
56 B271
APP134
B657A-1
B659C
99 B659A
ATCC51932
B319
B491
SVC135
0.002

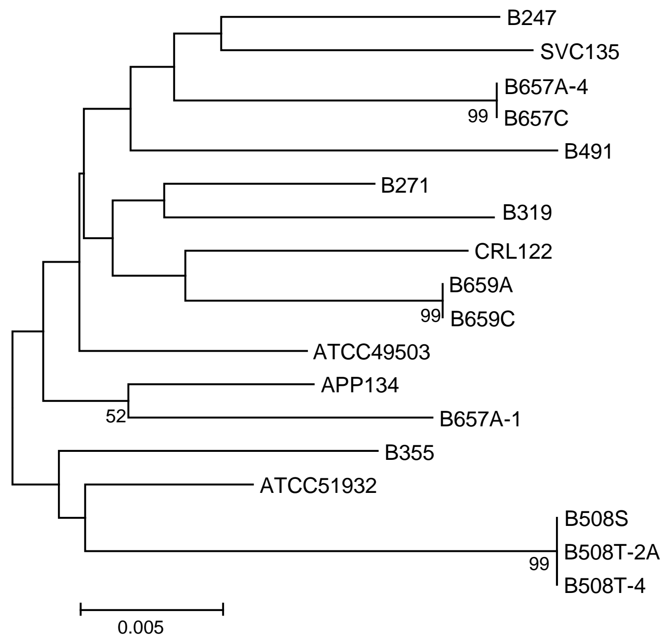
(B) *cpn60*



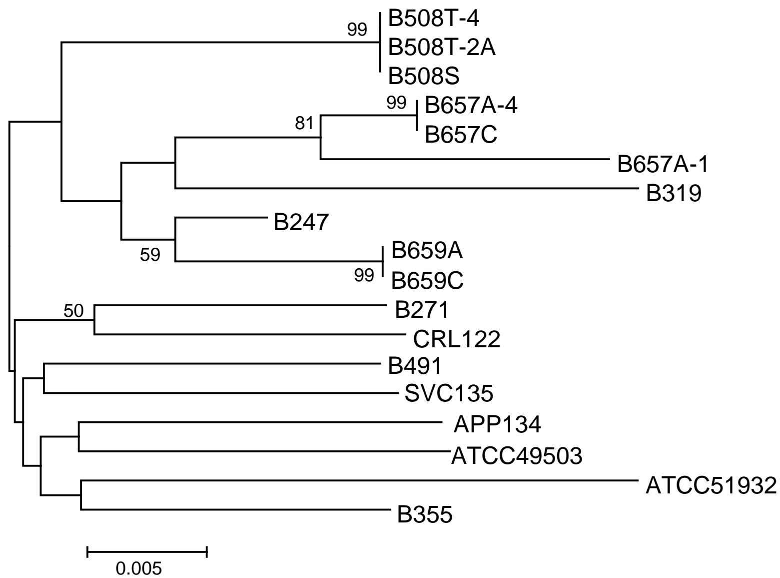
(C) *luxS*



(D) *cgt*



(E) *cpn70*



(F) *dnaJ*

Figure 3.1 Consensus neighbor-joining phylogenetic trees obtained from 18 *Helicobacter pylori* sequences of genes *amiA* (A), *cpn60* (B), *luxS* (C), *cgt* (D), *cpn70* (E), and *dnaJ* (F). Bar, distance values as calculated by MEGA 6.0. Bootstrap values (>50%) after 1,000 replicates are shown as percentages.

On the other hand, all six genes allowed us to differentiate two groups of clones from B657A and B657C biopsies. A group was formed by B657A-1, B657A-2 and B657A-3 and another group by B657A-4 and all four clones from B657C (**Figure 3.1**B, C, D, E, F), with the following intergroup distance rates: 3.6%, 2.2%, 3.5%, 3.1%, 1.6%, and 7.3% for the genes *amiA*, *cgt*, *cpn60*, *cpn70*, *dnaJ*, and *luxS*, respectively. Moreover, two nucleotide differences (0.3% divergence) were observed between B657A-4 and B657C clones for *amiA* (**Figure 3.1**A).

The phylogenetic tree obtained from the concatenated sequences (3,246 nt) is shown in **Figure 3.2**.

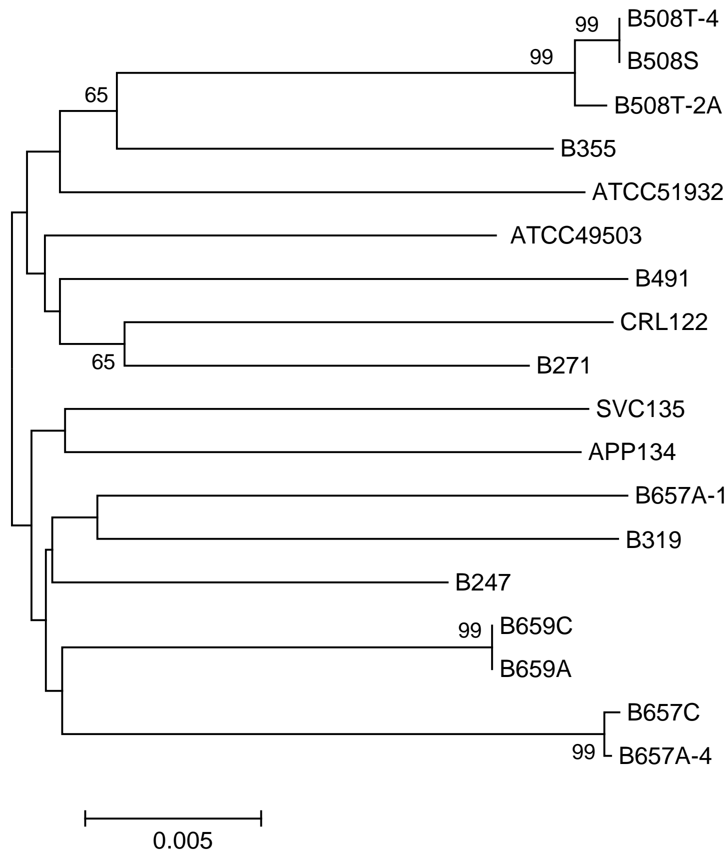


Figure 3.2 Consensus neighbor-joining phylogenetic tree obtained from *amiA*, *cgt*, *cpn60*, *cpn70*, *dnaJ*, and *luxS* concatenated sequences. Bar, distance of 0.005 substitutions per nucleotide position as calculated by MEGA 6.0. Bootstrap values (>50%) after 1,000 replicates are shown as percentages.

3.4 Discussion

In this study, specific PCR detection of *H. pylori* has been achieved with the housekeeping genes *amiA*, *cgt*, *cpn60*, *cpn70*, *dnaJ,* and *luxS*. These genes are potential candidates to detect *H. pylori* infection in gastric biopsies and other specimens (as gastric juice, stool, saliva, dental plaque, water, and food samples), together with others, mainly used for gastric biopsies, as *ureA*, *vacA*, 16S rRNA and 23S rRNA genes, previously described (Ramírez-Lázaro *et al.*, 2011; Wang *et al.*, 2015).

Sequencing of any of these six genes has allowed a clear differentiation between all clones isolated from different patients (0.9–7.8% divergence). These results suggest that these housekeeping genes could be useful to elucidate the mode of transmission of *H. pylori*, an issue that is still controversial. While oral-oral, gastro-oral and fecal-oral transmission are the most probable ways, *H. pylori* isolation from stools or the oral cavity is difficult and different environmental sources of *H. pylori* (food, animals, water) could play an important role as reservoirs (Khalifa *et al.*, 2010).

Two different strains isolated from the antral biopsy B657A were clearly detected by all genes with high distance values (1.6–7.3%). One strain was represented by the identical clones B657A-1, -2 and -3, and the other strain by the clone B657A-4, which was identical to clones isolated from B657C (corpus biopsy from the same patient), except in the case of *amiA* as mentioned below. This high divergence observed between both strains indicates an event of multiple infection in the antrum of the stomach. Horizontal gene transfer and genetic recombination by mixed infection is crucial for the acquisition of the high genetic diversity of *H. pylori* (Falush *et al.*, 2001). Although some authors have detected mixed infections, different studies have reported very variable prevalence rates and little is known about the role of multiple infections on disease outcome (Ghose *et al.*, 2005; Raymond *et al.*, 2004; Sheu *et al.*, 2009).

Although *luxS* gene showed the highest distance values, the genes *amiA* and *cpn60* were more useful to discriminate clones isolated from the same biopsy and patient. Two different clones isolated from the sample B508T, the gastric biopsy collected from tissue

with adenocarcinoma, were only detected by partial sequencing of *amiA* and *cpn60* (one and six nucleotide differences, respectively). One of both clones (B508T-4) was identical to clones isolated from B508S, the gastric biopsy from the same patient but collected from normal tissue. On the other hand, two nucleotide differences in *amiA* were observed between B657A-4 and B657C clones. These small differences between clones suggest events of microevolution rather than related to multiple infection. Evidences that *H. pylori* can establish a chronic infection after infection with a single strain have been described (Kennemann *et al.*, 2011; Toita *et al.*, 2013) and, recently, Linz *et al.* (Linz *et al.*, 2014) have revealed that the mutation rate during the acute infection phase is over 10 times faster than during chronic infection.

Concatenation of *amiA*, *cgt*, *cpn60*, *cpn70*, *dnaJ*, and *luxS* partial sequences allowed us to conduct a multilocus sequence analyses (MLSA) approach (**Figure 3.2**), which is useful to elucidate intra- and interspecies phylogenetic relationships (Gevers *et al.*, 2005; Tindall *et al.*, 2010) and, in the case of *H. pylori*, the phylogeographic differentiation of bacterial populations associated to the migration of human populations (Linz *et al.*, 2007; Suzuki *et al.*, 2012). Most of these genes have never been used for these purposes except *dnaJ* and, especially, *cpn60*, that is useful for microbial phylogeny, detection and identification, ecology, and evolution through the analysis of the 555 bp region known as universal target (UT) analysed in this study (Hill *et al.*, 2006; Miñana-Galbis *et al.*, 2010).

In conclusion, *H. pylori* specific detection has been developed, all *H. pylori* clones isolated from different patients have been discriminated, and microevolution and mixed infection have been detected by means of amplification and sequencing of housekeeping genes *amiA*, *cgt*, *cpn60*, *cpn70*, *dnaJ,* and *luxS* for the first time. Further studies based on these housekeeping genes are needed to explore their potential applications for *H. pylori* detection in different specimens, the mode of transmission of *H. pylori*, the role associated to virulence or the phylogeographic differentiation of *H. pylori* populations.

CHAPTER 4 Detection of *Helicobacter pylori* microevolution and multiple infection from gastric biopsies by housekeeping gene amplicon sequencing

Main contents of this study have been presented in International Conferences:

Detection of Helicobacter pylori microevolution and multiple infection from gastric biopsies by amplicon sequencing. Palau M, Comeau AM, Piqué N, Ramírez-Lázaro MJ, Lario S, Calvet X, Langille MGI, Miñana-Galbis D. Federation of European Microbiological Societies (FEMS), 9-13 July 2017. València.

4.1 Introduction

Helicobacter pylori is a Gram-negative bacterium characterized by a high level of intraspecies genetic diversity. This species has been intensely investigated with more than 43,000 research articles cited in PubMed[1]. One reason for such interest is that *H. pylori* has evolved with humans for about 100,000 years. Having coexisted with humans for this long period of time, tracing human migrations is still possible today through multilocus sequence analysis (MLSA) of housekeeping genes of *H. pylori* strains isolated from human populations with distinct geographical origins (Moodley *et al.*, 2012; Suzuki *et al.*, 2012).

Humans and some other primates are the unique hosts of *H. pylori*, which infects gastric mucosa as the primary niche and persists for the lifetime of the host in the absence of treatment. *H. pylori* infects about half of the world's population and the infection is mainly acquired in early childhood (Hashi *et al.*, 2018; Piqué *et al.*, 2016). Infection results in a life-long gastric colonization that usually is asymptomatic. However, it can lead to the development of several gastro-intestinal diseases, such as gastric ulcers, mucosa-associated lymphoid tissue (MALT) lymphoma, or chronic inflammation. The latter could potentially progress to multifocal atrophy, intestinal metaplasia, dysplasia, and gastric adenocarcinoma (Correa and Houghton, 2007; Kabamba *et al.*, 2018; Quaglia and Dambrosio, 2018).

It has been suggested that the transmission of *H. pylori* occurs directly via oral-oral, gastro-oral and fecal-oral routes, although food- and waterborne transmissions could play an important role (Khalifa *et al.*, 2010; Quaglia and Dambrosio, 2018). Another controversial issue related to transmission is the prevalence of *H. pylori* infection whether by a single strain or by multiple strains, since different methodologies and discordant results have been reported (Mansour *et al.*, 2016; Palau *et al.*, 2016). Studies of *H. pylori* infection at strain level using high-throughput analyses could be key to

[1] https://www.ncbi.nlm.nih.gov/pubmed/?term=helicobacter+pylori

elucidate the emergence or transmission of antibiotic resistances and the development of gastric diseases.

The finding of multiple strains in most patients has important clinical and research implications. For example, the presence of minority strains bearing antibiotic resistances –that will remain undetected in culture and conventional antibiogram– may explain why rescue treatment with the same antibiotic (especially clarithromycin) fails even in patients with sensitive strains in the culture (Baylina *et al.*, 2019).

The aim of this study was to evaluate the usefulness of amplicon sequencing methodology for the detection of *H. pylori* microevolution and multiple infection from gastric biopsies of patients with dyspeptic symptoms and different histopathological findings (from atrophy to adenocarcinoma).

4.2 Material and Methods

4.2.1 Gastric biopsies

Patient recruitment and gastric biopsies obtained by the Digestive Diseases Department of the Hospital Taulí (Sabadell, Barcelona) were described in previous studies (Lario *et al.*, 2017; Palau *et al.*, 2016). The study was approved by the Ethics Committee in accordance with the Declaration of Helsinki. Biopsies were performed after obtaining the patients' written informed consent. For the present study, DNA from five gastric biopsies corresponding to four different patients infected with *H. pylori* were analyzed. Regarding the patients, two suffered from gastric cancer (B247, B508), one from duodenal ulcer (B373) and one was without medical conditions, here labelled as *normal* stomach (B601). Histological diagnoses from the biopsies were: chronic active gastritis, atrophy, intestinal metaplasia, and adenocarcinoma. Biopsies B508S and B508T were taken from the same patient with gastric adenocarcinoma (B508). Specifically, biopsy B508S was taken from normal gastric tissue and biopsy B508T from the tumoral tissue (**Table 4.1**).

Patient ID	Age and gender	Endoscopic diagnosis	Anatomical location of biopsy specimen	Histopathological diagnosis	Specimen ID
B247	63 y/o male	Gastric cancer	Non-neoplastic stomach	Chronic active gastritis	B247S
B373	56 y/o male	Duodenal ulcer	Antrum	Intestinal metaplasia	B373A
B508	77 y/o male	Gastric cancer	Non-neoplastic stomach	Not available	B508S
			Tumor tissue	Adenocarcinoma	B508T
B601	43 y/o female	Normal	Antrum	Atrophy	B601A

Table 4.1 Samples included in this study.

4.2.2 DNA extraction and PCR amplification

Extraction and sequential purification of DNA were performed using the MasterPure kit (Epicentre, Illumina). Isolated DNA was quantified with a NanoDrop spectrophotometer (Nano-Drop Technologies, USA).

Using *H. pylori* specific primers, PCR amplifications of six housekeeping genes (*amiA*, *cgt*, *cpn60*, *cpn70*, *dnaJ*, and *luxS*) were performed as described in a previous work of the authors (Palau *et al.*, 2016), in which these genes were useful for *H. pylori* clones discrimination. From these genes, *cgt* and *luxS* demonstrated the best PCR results (a single, clear band with the corresponding length revealed in agarose gel) and therefore, were selected for amplicon sequencing.

Fusion *cgt* and *luxS* primers (**Table 4.2**) were designed for unidirectional sequencing following the 454 sequencing system guidelines for amplicon experimental design (**Figure 4.1**) (454 Life Sciences Corp., Branford, CT, USA). Briefly, adaptor sequences are dictated by the requirements of the 454 Sequencing platform and they must always end with the sequence key “TCAG”. Immediately after the key, Multiplex Identifier (MID) are used to barcode the amplicons, in this case, being a unidirectional sequencing, it was just included on the forward strands. The 3’-portion of the primer is designed to bind itself to a specific sequence of the template DNA. PCR amplifications were carried out

with the FastStart High Fidelity PCR System (Roche Diagnostics GmBH, Mannheim, Germany) in a total volume of 50 µL containing 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 2.5 U FastStart polymerase, 40 pmol of each primer (Isogen Life Science, PW De Meern, The Netherlands), and 200 ng DNA. The reaction mixtures were subjected to the following thermal cycling program in a PrimeG Thermal Cycler (Bibby Scientific, Staffordshire, UK): initial denaturation at 95ºC for 5 min; 35 cycles of 95ºC for 1 min, 60ºC (*cgt*) or 54ºC (*luxS*) for 1 min, and 72ºC for 1 min; final extension at 72ºC for 5 min. The amplified products were purified using ExoSAP-IT® kit (Affymetrix, Santa Clara, CA, USA). Briefly, 5 µl of PCR product are mixed with 2 µl of ExoSAP enzyme. Then, the mixture is incubated for 15 min at 37ºC followed by 15 min at 80ºC. With this incubation steps the excess of nucleotides and primers from the PCR reaction is eliminated.

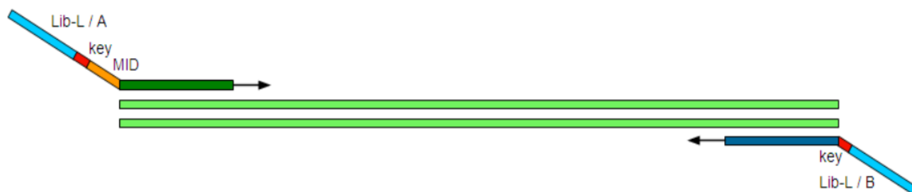


Figure 4.1 Schematic representation of the PCR reaction components (Roche, 2014), depicting fusion primers and the template DNA.

4.2.3 Preparation of libraries

Following the Amplicon Library Preparation Manual (454 Life Sciences Corp., Branford, CT, USA), the amplicons including fusion primers of each sample were sequenced using the 454 GS Junior instrument (Roche, Basel, Switzerland). After PCR purifications, library quantification was done using a Qubit™ Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA).

Each amplicon was diluted separately to 1×10^9 molecules/µL in 1xTE buffer. Diluted amplicons were mixed to prepare the amplicon pool. As the size of *cgt* and *luxS* amplicons were different (685 bp for *cg*t and 499 bp for *luxS*), mixed volumes were adjusted to the ratio 1.2:1 *cgt*:*luxS*. The amplicon pool was diluted to 1×10^7 molecules/µL in molecular biology grade water. Once the libraries were constructed and pooled, emPCR amplification and sequencing were performed in the facilities of the Genomics Unit of Scientific and Technological Centers from University of Barcelona (CCiTUB).

Gene	Fusion primer	**Sequence** (Adaptor–*key*–**MID**–template-specific sequence)
cgt	cgt-252-B247S	CCATCTCATCCCTGCGTGTCTCCGAC*TCAG***AGACGCACTC**GGCTTTTAAGGGAGCGGATA
	cgt-252-B373A	CCATCTCATCCCTGCGTGTCTCCGAC*TCAG***AGCACTGTAG**GGCTTTTAAGGGAGCGGATA
	cgt-252-B508S	CCATCTCATCCCTGCGTGTCTCCGAC*TCAG***ACGAGTGCGT**GGCTTTTAAGGGAGCGGATA
	cgt-252-B508T	CCATCTCATCCCTGCGTGTCTCCGAC*TCAG***ACGCTCGACA**GGCTTTTAAGGGAGCGGATA
	cgt-252-B601A	CCATCTCATCCCTGCGTGTCTCCGAC*TCAG***ATCAGACACG**GGCTTTTAAGGGAGCGGATA
	cgt-866 (reverse)	CCTATCCCCTGTGTGCCTTGGCAGTC*TCAG*ATCGCTTCGCTYTCCACATT
luxS	luxS-38-B247S	CCATCTCATCCCTGCGTGTCTCCGAC*TCAG***AGACGCACTC**TGGATCACACYAAAGTCAAAG
	luxS-38-B373A	CCATCTCATCCCTGCGTGTCTCCGAC*TCAG***AGCACTGTAG**TGGATCACACYAAAGTCAAAG
	luxS-38-B508S	CCATCTCATCCCTGCGTGTCTCCGAC*TCAG***ACGAGTGCGT**TGGATCACACYAAAGTCAAAG
	luxS-38-B508T	CCATCTCATCCCTGCGTGTCTCCGAC*TCAG***ACGCTCGACA**TGGATCACACYAAAGTCAAAG
	luxS-38-B601A	CCATCTCATCCCTGCGTGTCTCCGAC*TCAG***ATCAGACACG**TGGATCACACYAAAGTCAAAG
	luxS-466 (reverse)	CCTATCCCCTGTGTGCCTTGGCAGTC*TCAG*AAACCCCCACTTCAGACCA

Table 4.2 Fusion primers designed for *H. pylori* specific PCR and amplicon sequencing.

4.2.4 Data processing

Raw data was processed by the GS Run Processor, which resulted in FASTQ and Standard Flowgram Format (SFF) files that were used in subsequent analyses. Amplicon sequences were assigned to each corresponding sample using the multiplex identifiers (MIDs), generating five files for each gene (**Table 4.2**). Thereafter, MIDs, sequences <50 bp or containing >5% Ns were removed from FASTQ files (Schmieder and Edwards, 2011). This processing was carried out by the CCiTUB.

Sequences in FASTA files were aligned with *cgt* or *luxS* reference sequences (Palau *et al.*, 2016) using the MEGA7 software (Kumar *et al.*, 2016). The aim was to check that amplicon sequences corresponded to *cgt* or *luxS* sequences. Additionally, in this process the length of the sequences was adjusted.

Sequences were analysed with Mothur (Schloss *et al.*, 2009) following the subsequent protocol (all commands were run with the default parameters except when explicitly specified): first, the *unique.seqs* command was used in order to obtain the unique sequences of the sequence file by dereplication. Next, the *merge.files* command concatenates multiple files, which is useful for merging multiple FASTA-formatted sequence files. The *make.group* command reads a series of FASTA files and creates a group file, which is used to assign sequences to a specific group. The *dist.seqs* command then calculates uncorrected pairwise distances between aligned DNA sequences. By default, *dist.seqs* penalizes terminal gaps. This option was excluded and the default value of *countends* was changed to False. Afterwards, the *cluster* command was used to assign sequences to Operational Taxonomic Units (OTU). The *furthest neighbor* clustering method was chosen, as it is the most conservative way to select OTUs. The *make.shared* and the *summary.shared* commands were used successively to obtain a summary file with the calculated values for each line in the OTU data and for all possible comparisons between the different groups in the group file. The *get.oturep* command generated a FASTA-formatted sequence file containing only one

representative sequence for each OTU. Finally, the *classify.seqs* command was used to taxonomically classify the sequences.

4.3 Results

4.3.1 Samples and amplicon sequences

As mentioned above, the genes *cgt* and *luxS* were selected from six housekeeping genes for amplicon sequencing. The sequencing results of the amplicon libraries (**Table 4.3** and **Table 4.4**) yielded a much larger number of *cgt* sequences (11,833 in total and 2,042 unique sequences) than *luxS* sequences (403 in total and 112 unique sequences). This is probably due to the increased dilution of *luxS* amplicons in the preparation of the pooling. All sequences from each sample and gene were processed in order to adjust for the length of the sequences (438 nt for *cgt* sequences and 366 nt for *luxS* sequences) and to detect duplicate sequences. As a result, a significant number of identical sequences and a low number of unique sequences (11–43%) were obtained (**Table 4.3** and **Table 4.4**). All sequences were deposited in Genbank under the project accession number PRJNA434670.

The distribution of total and unique *cgt* sequences by sample was as follows (**Table 4.3**): 5,185 total and 593 (11%) unique sequences from sample B247S; 2,422 and 422 (17%) from B373A; 968 and 234 (24%) from B508S; 1,414 and 339 (24%) from B508T; and 1,844 and 454 (25%) from B601A. In the case of *luxS*, the distribution of total and unique sequences by sample was as follows (**Table 4.4**): 111 total and 34 (31%) unique sequences from sample B247S; 72 and 16 (22%) from B373A; 56 and 18 (32%) from B508S; 115 and 23 (20%) from B508T; and 49 and 21 (43%) from B601A.

***cgt* OTUs**	**B247S sequences**		**B373A sequences**		**B508S sequences**		**B508T sequences**		**B601A sequences**		***cgt* sequences**	
	total	**unique**	**total**	**unique**	**total**	**unique**	**total**	**unique**	**total**	**unique**	**total**	**unique**
OTU-01	5,184	592	–	–	–	–	3	3	–	–	5,187	595
OTU-02	1	1	2,416	418	–	–	4	4	–	–	2,421	423
OTU-03	–	–	–	–	963	230	1,391	321	–	–	2,354	551
OTU-04	–	–	–	–	–	–	–	–	1,808	426	1,808	426
OTU-05	–	–	–	–	–	–	–	–	30	25	30	25
OTU-06	–	–	6	4	4	3	16	11	–	–	26	18
OTU-07	–	–	–	–	1	1	–	–	4	1	5	2
OTU-08	–	–	–	–	–	–	–	–	1	1	1	1
OTU-09	–	–	–	–	–	–	–	–	1	1	1	1
TOTAL	5,185	593 (11%)	2,422	422 (17%)	968	234 (24%)	1,414	339 (24%)	1,844	454 (25%)	11,833	2,042 (17%)

Table 4.3 Distribution of *cgt* amplicon sequences by samples and OTUs.

***luxS* OTUs**	**B247S sequences**		**B373A sequences**		**B508S sequences**		**B508T sequences**		**B601A sequences**		***luxS* sequences**	
	total	**unique**	**total**	**unique**	**total**	**unique**	**total**	**unique**	**total**	**unique**	**total**	**unique**
OTU-01	98	25	–	–	–	–	–	–	–	–	98	25
OTU-02	–	–	70	14	–	–	–	–	–	–	70	14
OTU-03	–	–	–	–	52	14	107	15	5	2	164	31
OTU-04	–	–	–	–	–	–	–	–	33	12	33	12
OTU-05	–	–	–	–	2	2	1	1	8	4	11	7
OTU-06	7	6	–	–	–	–	1	1	–	–	8	7
OTU-07	–	–	2	2	–	–	1	1	1	1	4	4
OTU-08	6	3	–	–	–	–	–	–	–	–	6	3
OTU-09	–	–	–	–	1	1	2	2	–	–	3	3
OTU-10	–	–	–	–	–	–	2	2	–	–	2	2
OTU-11	–	–	–	–	1	1	1	1	–	–	2	2
OTU-12	–	–	–	–	–	–	–	–	1	1	1	1
OTU-13	–	–	–	–	–	–	–	–	1	1	1	1
TOTAL	111	34 (31%)	72	16 (22%)	56	18 (32%)	115	23 (20%)	49	21 (43%)	403	112 (28%)

Table 4.4 Distribution of *luxS* amplicon sequences by samples and OTUs.

4.3.2 OTU assignment and distribution

All unique sequences belonging to the same gene were merged in a FASTA file and pairwise distances were calculated to generate a PHYLIP-formatted distance matrix. A cutoff distance of 3% for both genes was defined in order to assign sequences to OTUs. In the previous study (Palau *et al.*, 2016), it was shown that for the *luxS* gene this value was the mean distance value between sequences from clones isolated from different patients. The distant value of the *cgt* gene was calculated to be 1%. However, due to the inherent error rates of Roche 454 platform, it was decided to raise the value to 3%. Therefore, sequences with distance <3% were considered as originating from the same strain. A total of 9 and 13 OTUs were obtained from the *cgt* (**Table 4.3**) and *luxS* (**Table 4.4**) distance matrix, respectively. In all samples at least two different OTUs were detected. Additionally, most of the total (99% from *cgt* and 91% from *luxS*) and unique (98% from *cgt* and 73% from *luxS*) sequences clustered in OTUs 01–04. Some sequences from different samples belonged to the same OTU, especially in the case of OTU-03 (containing sequences from samples B508S and B508T) (**Table 4.3** and **Table 4.4**).

As seen in the tables, every sample had a dominant OTU. For example, for the gene *cgt* in the sample B247S, 5,184 sequences out of 5,185 of the total were in OTU-01. When referring to unique sequences, 592 of 593 are in this OTU. Regarding the gene *luxS* in the same sample, 98 sequences of a total of 111 were in OTU-01. And referring to unique sequences, 25 of a total of 34 were in this OTU.

4.3.3 Amplicon sequence identification

As *cgt* and *luxS* PCRs were specific for *H. pylori*, all amplicon sequences matched this species. In order to classify the amplicon sequences obtained in this study, two FASTA-formatted sequence files for each gene were generated. One file contained the representative sequence for each OTU (by means of the *get.oturep* Mothur command) and the other file with reference sequences as detailed below.

For the reference sequences file, *cgt* and *luxS* sequences were downloaded from 75 *H. pylori* complete genomes hosted in GenBank (**Table AII.1**). These sequences were processed using Mothur (*unique.seqs* command) in order to avoid duplicate sequences, resulting in 53 *cgt* and 59 *luxS* unique sequences. Additional 15 *cgt* and *luxS* sequences were also incorporated as reference sequences (**Table AII.1**). These additional sequences had been previously described in Palau *et al.* (2016). Furthermore, some of them were obtained from strains isolated from the same biopsies (B247S, B373A, B508S, and B508T) used in this study. All reference sequences were adjusted to the length of 438 nt for *cgt* and 366 nt for *luxS*.

Representative OTU sequences from this study were classified using Mothur (*classify.seqs* command) with the above reference sequences as templates. On the other hand, all sequences (representative OTUs and references) were joined in a single file, one for each gene. Thereafter, the sequence distances were calculated and phylogenetic trees were constructed using MEGA7 (Kumar *et al.*, 2016). Moreover, representative OTU sequences were also compared using Nucleotide BLAST on the NCBI website.

The representative sequence of *cgt* OTU-01, obtained from sample B247S, showed one nucleotide difference (the first nucleotide of the sequence) with respect to the *cgt* reference sequence. The reference was obtained from *H. pylori* strain B247 (**Table 4.5** and **Figure 4.2**), which was isolated from the replicate biopsy B247S in a previous study (Palau *et al.*, 2016). The representative sequence of *cgt* OTU-02 (from sample B373A) was identical to the *cgt* reference sequence. In this case, the reference was obtained from *H. pylori* strain B373, isolated from replicate biopsy B373A in the above mentioned study. The representative sequence of *cgt* OTU-03, that comprised most of the sequences from samples B508S and B508T, was identical to the *cgt* reference sequence that was obtained from *H. pylori* strain B508S, isolated from replicate biopsy B508S. The representative sequence of the other principal OTU (OTU-04) (from sample B601A) could not be compared to the corresponding strain sequence because no *H.*

pylori strain could be isolated from biopsy B601A. The other five *cgt* OTUs were different from any reference sequence, except in the case of *cgt* OTU-06. Its sequence was identical to the sequence of *H. pylori* ATCC 51932 (**Table 4.5** and **Figure 4.2**).

***cgt* OTUs**	**Representative sequence**	**Mothur**	**MEGA7 (% similarity)**	**BLAST (% similarity)**
OTU-01	JVDD3PU01CBX2Q	B247	B247 (99.8)	B247 (100)
OTU-02	JVDD3PU01CG848	B373A	B373A (100)	B373A (100)
OTU-03	JVDD3PU01A0MHF	B508S	B508S (100)	B508S (100)
OTU-04	JVDD3PU01A0A1O	SVC135	OTU-05 (99.5), SVC135 (98.6)	SVC135 (99)
OTU-05	JVDD3PU01BIJRV	SVC135	OTU-04 (99.5), SVC135 (98.2)	SVC135 (98)
OTU-06	JVDD3PU01A46IT	ATCC 51932	ATCC 51932 (100)	ATCC 51932 (100)
OTU-07	JVDD3PU01B6AVS	B508S	B508S, B657A-4, B38 and OTU-03 (98.4)	B508S, B657A-4 and B38 (98)
OTU-08	JVDD3PU01ABYSA	SVC135	OTU-04 (95.8), SVC135 (94.3)	SVC135 (98)
OTU-09	JVDD3PU01C89ZH	SVC135	OTU-05 (97.2), OTU-04 (96.7), SVC135 (95.3)	SVC135 (97)

Table 4.5 Classification of *cgt* OTUs by Mothur, MEGA7 and BLAST.

Similarly, the representative sequences of *luxS* OTU-01, -02, and -03 were identical to the reference sequences that were obtained from the above-mentioned strains, isolated from replicate gastric biopsies of the same patient (B247, B373, and B508, respectively) (**Table 4.6** and **Figure 4.3**). The representative sequence of *luxS* OTU-04 could not be compared either to the corresponding strain sequence because the lack of *H. pylori* strain isolation from biopsy B601A. The other nine *luxS* OTUs were different from any reference sequence (**Table 4.6** and **Figure 4.3**).

***luxS* OTUs**	**Representative sequence**	**Mothur**	**MEGA7 (% similarity)**	**BLAST (% similarity)**
OTU-01	JVDD3PU01A1FQ2	B247	B247 (100)	B247 (100)
OTU-02	JVDD3PU01A13LE	B373A	B373A (100)	B373A (100)
OTU-03	JVDD3PU01A15L2	B508S	B508S (100)	B508S (100)
OTU-04	JVDD3PU01ANTJ4	7C	7C (99.2)	7C (99)
OTU-05	JVDD3PU01A6Z3D	oki112	OTU-07 (97.8), OTU-06, B247 and B319 (96.9)	B247 and B319 (97)
OTU-06	JVDD3PU01AN58Z	B247	OTU-08 (98.6), B247 and OTU-02 (98.3)	B247 (98)
OTU-07	JVDD3PU01ASHDN	B373A	OTU-05 (97.8), ELS37 (97.2)	ELS37 (97)
OTU-08	JVDD3PU01BK98M	B491	OTU-06 (98.6), B319 and B491 (97.5)	B319, B491 and ATCC 51932 (97)
OTU-09	JVDD3PU01CMBEF	B508S	B508S and OTU-03 (97.2)	B508S (98)
OTU-10	JVDD3PU01CPAE2	ATCC 51932	ATCC 51932 (99.7)	ATCC 51932 (99)
OTU-11	JVDD3PU01DOSQL	G27	G27 (97.2)	G27 (97)
OTU-12	JVDD3PU01COY1E	7C	OTU-04 (97.5), 7C, B319, B508S, ELS37, HPAG1, HUP-B14 and OTU-03 (96.6)	7C, B319, B508S, ELS37, HPAG1 and HUP-B14 (97)
OTU-13	JVDD3PU01CSY7G	B508S	OTU-04 (97.8), 7C (97.5)	7C (98)

Table 4.6 Classification of *luxS* OTUs by Mothur, MEGA7 and BLAST.

Figure 4.2 Neighbor-joining phylogenetic tree obtained from representative OTU sequences and reference sequences of gene *cgt*. Bar, distance values as calculated by MEGA 7.0. Bootstrap values (>50%) after 1,000 replicates are shown as percentages.

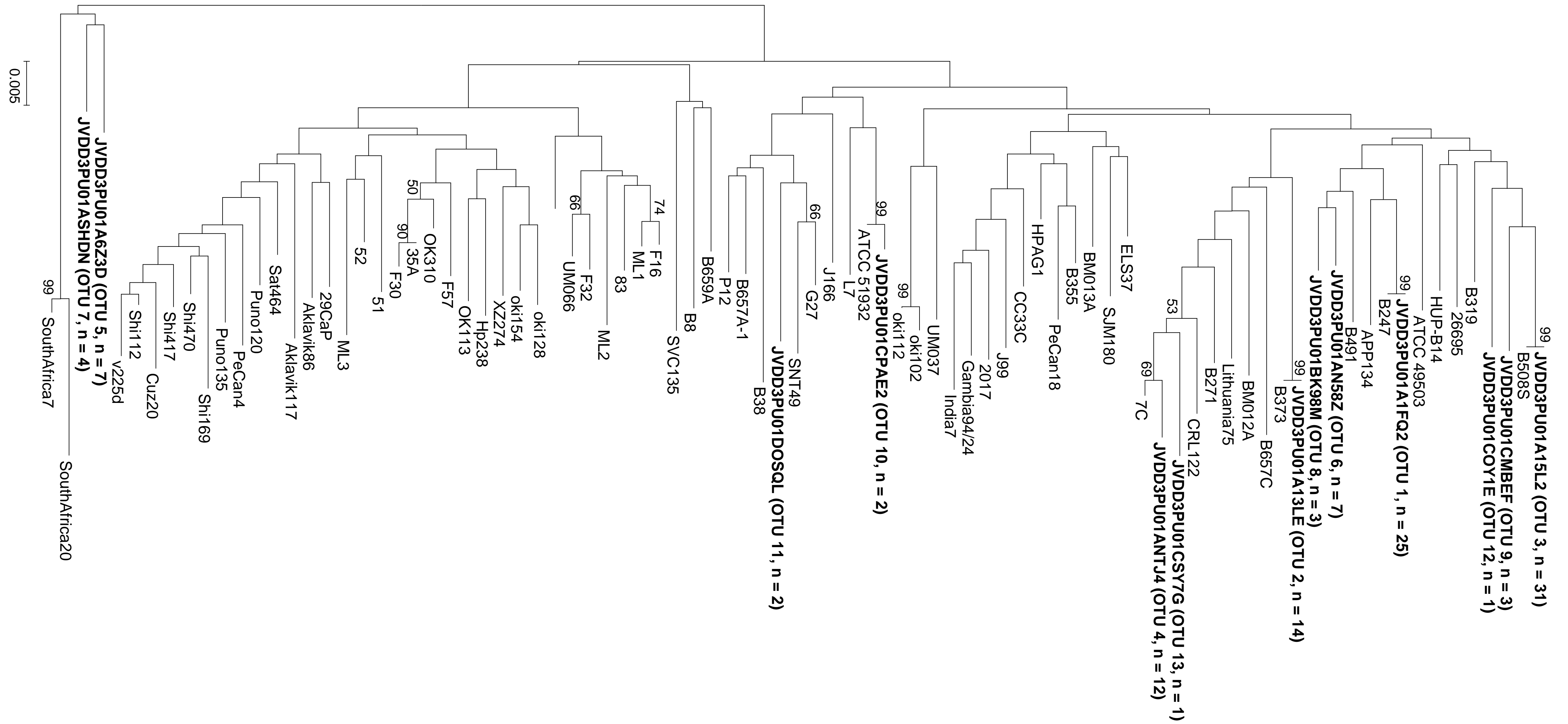


Figure 4.3 Neighbor-joining phylogenetic tree obtained from representative OTU sequences and reference sequences of gene *luxS*. Bar, distance values as calculated by MEGA 7.0. Bootstrap values (>50%) after 1,000 replicates are shown as percentages.

4.4 Discussion

Most amplicon sequences were repeated sequences that were largely represented by four OTUs. Representative sequences from these OTUs matched with sequences of strains isolated from replicate gastric biopsies obtained from the same patient (except in the case of sample B601A because of the lack of corresponding isolate). However, although in the minority, other OTUs were detected in all five gastric biopsies for both genes. These results suggest that *H. pylori* colonizes the human stomach through different infection events that lead to a gastric multi-infection with a predominant strain together with other minority strains.

Other studies have evaluated the prevalence of mixed *H. pylori* infections, but with a high disparity of results, even when using similar methodologies. Using randomly amplified polymorphic DNA (RAPD) fingerprinting, Toita *et al.* (2013) reported that all patients were infected with a single *H. pylori* clone. Contrarily, Sheu *et al.* (2009) found a 23.3% prevalence of mixed infections. Analyzing the *cag*-PAI status and the s-region or m-region of *vacA* in *H. pylori* isolates by PCR, Lai *et al.* (2016) described a prevalence of 28% for mixed infections. Additionally, Kibria *et al.* (2015) studied the prevalence of mixed *H. pylori* infection using both methods mentioned above (RAPD fingerprinting and multiplex PCR amplification for *cagA* and *vacA* alleles). An overall prevalence of 60.2% was obtained. Recently, an event of multiple infection was detected from 52 *H. pylori* clones isolated from 11 patients by sequencing of six housekeeping genes (*amiA*, *cgt*, *cpn60*, *cpn70*, *dnaJ*, and *luxS*) (Palau *et al.*, 2016). However, Raymond *et al.* (2004) detected mixed infection in all six members of a family analysing 107 clones by MLST of two housekeeping genes (*hspA* and *glmM*).

To our knowledge, the present work is the first study that uses amplicon sequencing of housekeeping genes to detect *H. pylori* multi-infections. Amplicon sequencing is a high-throughput technique with greater resolution when compared to other methods (i.e. RAPD fingerprinting, multiplex *cagA* and *vacA* PCR, MLST) used in previous studies. In

the present study, multiple *H. pylori* infections with a predominant strain were detected in all the gastric biopsies analyzed, in accordance with results obtained for Raymond *et al.* (2004). Although further studies are needed to confirm our results, these suggest that mixed *H. pylori* infections are the main status in the colonization of the human gastric mucosa. In fact, a high frequency of recombination between unrelated strains in mixed colonization is the main cause of the panmictic population structure of *H. pylori* (Cao *et al.*, 2015; Falush *et al.*, 2001; Morelli *et al.*, 2010). In this study, microevolution by mutation of original strains infecting the gastric mucosa (Cao *et al.*, 2015; Linz *et al.*, 2014; Morelli *et al.*, 2010) was also found, since different unique sequences were obtained in each OTU (if it consisted of two or more amplicon sequences).

Amplicon sequencing of housekeeping, virulence or antibiotic resistant genes could be useful in different fields. It could be useful for elucidating the mode of transmission of *H. pylori*, to perform epidemiological studies. The impact of mixed *H. pylori* infection in the gastric pathogenesis and the failure of antibiotic treatments deserves also additional investigation.

In conclusion, our study strongly suggests that multiple infections including a predominant strain and multiple minority *H. pylori* strains is the predominant pattern of *H. pylori* infections in humans. The clinical and biological relevance of this finding deserves further study.

CHAPTER 5 **Whole-genome sequencing and comparative genomics of three *Helicobacter pylori* strains from a stomach with adenocarcinoma**

5.1 Introduction

Helicobacter pylori is a Gram-negative bacterium that persistently infects the human stomach inducing chronic inflammation. The pathogenicity of *H. pylori* ranges from asymptomatic colonization to bacterially mediated oncogenesis (Baltrus *et al.*, 2009). This bacterium can cause several gastrointestinal diseases, such as gastritis, peptic ulcer disease, gastric adenocarcinoma, and a mucosa-associated lymphoid tissue (MALT) lymphoma.

The genomes of *H. pylori* are highly variable, with considerable allelic diversity. This genomic plasticity is thought to aid in its adaptation, which is essential for its survival in different human populations (Chiurillo *et al.*, 2013). It has been reported that isolates associated with different geographic areas, different diseases and different individuals might have variable genomic features (Cao *et al.*, 2016; Gressmann *et al.*, 2005; You *et al.*, 2015). Several genes, such as *vacA*, *cagA*, and *iceA*, among others, have been identified as markers for enhanced pathogenicity of *H. pylori* (Junaid *et al.*, 2016).

All *H. pylori* strains carry the vacuolating cytotoxin gene (*vacA*). This gene contains three highly variable polymorphic regions, which are the signal sequence region (*s1*, *s2*), the intermediate region (*i1*, *i2*, *i3*) and the mid region (*m1*, *m2*). Cytotoxicity is related to variations in the signal and mid regions. Strains of *H. pylori* with the *s1/m1* genotype are more frequently associated with severe disease symptoms than strains carrying other combinations of these alleles (Atherton *et al.*, 1995). Specifically, *s* region variations are associated with the vacuolating activity of the protein VacA, whereas variations in the *m* region determine the cell specificity of vaculation, by affecting the binding of the toxin to the host cells (Rudi *et al.*, 1998). Toxin production is highest in *s1/m1* strains, moderate in *s1/m2* strains, and scarce or null in *s2/m2* strains (Rhead *et al.*, 2007). The *i* region is positioned between the *s* and *m* regions and is the most recent region to be described. The *i1* variants of VacA have been shown to have a stronger vacuolating activity than toxins containing the *i2* regions (Mascellino *et al.*, 2009).

The *cagA* gene is a marker for the presence of the *cag* pathogenicity island (*cag*-PAI) of approximately 40 kb, whose presence is associated with more severe clinical outcomes. A type IV secretion system translocates CagA protein into gastric epithelial cells, where it is phosphorylated. When this modification occurs, CagA affects various cellular processes and signal transduction pathways, such as disruption of tight and adherent junctions that lead to proinflammatory and mitogenic responses–effects (Junaid *et al.*, 2016). Studies on *H. pylori* heterogeneity have proved that the strongest virulence factors were among the genes within the *cag*-PAI.

The more recently described 'induced by contact with epithelium A' (*iceA*) gene is another virulence factor of relevance (Peek *et al.*, 1996). It has two main allelic variants, *iceA1* and *iceA2*. Allele *iceA1* shows sequence homology with a gene from *Neisseria lactamica*, *nlaIIIR*, which encodes a CTAG-specific restriction endonuclease. Furthermore, it has been associated with peptic ulcer disease. On the other hand, *iceA2* has been associated with gastritis, although homology to other genes has not been described and the function of its product remains unclear (da Costa *et al.*, 2015).

Flagella, urease, lipopolysaccharide (LPS) and peptidoglycan are other critical factors related to the pathogenicity of *H. pylori*. Regarding the first factor, *H. pylori* can have between two and six unipolar flagella, which confer motility to the cell and are considered an absolutely essential virulence factor for colonization. Although the presence of the flagellum alone has been shown to be insufficient for colonization, non-mobile mutants lacking the flagellum are unable to establish an infection in animal models (Piqué *et al.*, 2016). Concerning the urease, this enzyme transforms urea into ammonium and bicarbonate to counteract the acidic environment of the stomach. Furthermore, it also plays an important role in the pathogenesis of *H. pylori*, as it is correlated with its adhesion and its immunogenicity (Piqué *et al.*, 2016). Regarding the LPS, its structural characteristics provide the cell with two mechanisms of persistence in its niche. On the one hand, Lewis antigens present in the O-antigen mimic the glycan structures of human cells thus, facilitating the immune escape. On the other hand, lipid

A unique structure provides the cell resistance to host cationic antimicrobial peptides (CAMPs) (Li *et al.*, 2016; Sycuro *et al.*, 2010). It has been proposed that the coordinated action of different proteins causes a crosslinking relaxation of the peptidoglycan, which confers its characteristic helical curvature, necessary for reliable bacterial colonization of the stomach (Li *et al.*, 2016; Piqué *et al.*, 2016).

The high number of these factors and allelic variation of the involved genes generates a highly complex scenario and reveals the difficulties in testing the contribution of each individual factor. The association of gastric cancer with the *H. pylori* infection has led to its systematic eradication. But the progression from the infection to the development of cancer continues to be unclear, promoting a great interest in clarifying this issue. The aim of this study was to investigate possible relevant differences between genomes of strains isolated from tumoral gastric tissue and from non-tumoral tissue. Both samples were obtained from the same patient who suffered from gastric cancer.

5.2 Materials and Methods

5.2.1 *Helicobacter pylori* strains and DNA sequencing

This study is part of a larger project, which encompasses the analysis of 51 complete genomes of different strains of *Helicobacter pylori* (GenBank accession number PRJNA449871) from the *H. pylori* collection of the Digestive Diseases Department of the Hospital Taulí (Sabadell, Barcelona) (**Table 5.1**). Genomic DNA from all of them was extracted using the genomic DNA extraction kit (REAL, Durviz S.L., València) following the procedure explained in Palau *et al.* (2016) (see **section 3.2.2**). Genome sequencing was performed using Illumina technology (San Diego, USA). The sequencing library preparation and sequencing of the whole genome was done by the Centre for Genomic Regulation (CRG, Barcelona) using Illumina Hi-Seq Sequencing v4 Chemistry. Following, a *de novo* assembly of all the set of genomes was carried out using the tool

a5-assembler (Coil *et al.*, 2015), which also generates a statistics summary table with the length of the genome and the GC%.

Pan- and core-genome analyses for all strains were performed using Roary, the rapid large-scale prokaryote pan-genome analysis pipeline from the Sanger Institute (https://github.com/sanger-pathogens/Roary; Page *et al.*, 2015). A phylogenetic tree of pan-genomes was constructed using the tool PanX (Ding *et al.*, 2018), a software package that allows for comprehensive analysis, interactive visualization and dynamic exploration of bacterial pan-genomes.

Strain	Biopsy site	Age and gender	Histopathological diagnosis
AAP164	antrum	64 y/o male	severe active gastritis
AFR58	antrum	37 y/o male	metaplasia
APR133	antrum	67 y/o male	mild active gastritis
B126	antrum	43 y/o male	metaplasia
B247A	antrum	63 y/o male	adenocarcinoma*
B274	antrum	80 y/o male	mild active gastritis
B297	antrum	66 y/o female	moderate active gastritis
B314	antrum	53 y/o male	metaplasia
B319	antrum	70 y/o female	mild active gastritis
B335	corpus/fundus	75 y/o female	adenocarcinoma
B344	antrum	63 y/o male	moderate active gastritis
B345	antrum	59 y/o male	moderate active gastritis
B355	antrum	42 y/o male	mild active gastritis
B360	antrum	65 y/o male	moderate active gastritis
B366	antrum	54 y/o male	metaplasia
B368	antrum	69 y/o male	mild active gastritis
B373	antrum	56 y/o male	metaplasia
B400	antrum	63 y/o male	moderate active gastritis
B444A	antrum	35 y/o male	metaplasia
B448	antrum	68 y/o female	mild active gastritis
B455	antrum	62 y/o male	moderate active gastritis
B464A	antrum	42 y/o female	metaplasia
B491	corpus	60 y/o male	adenocarcinoma
B497A	antrum	56 y/o female	moderate active gastritis
B508A-S1	antrum	78 y/o male	adenocarcinoma*
B508A-T2A	antrum	78 y/o male	adenocarcinoma

B508A-T4	antrum	78 y/o male	adenocarcinoma
B518	corpus	42 y/o male	adenocarcinoma
B528A	antrum	51 y/o male	metaplasia
B529	antrum	66 y/o male	mild active gastritis
B547F	fundus	63 y/o male	adenocarcinoma
B572A	antrum	63 y/o female	metaplasia
B630	antrum	60 y/o male	moderate active gastritis
B657-A1	antrum	36 y/o male	mild active gastritis
B657-A4	antrum	36 y/o male	mild active gastritis
B657-C1	corpus	37 y/o male	mild active gastritis
B659-A1	antrum	60 y/o male	moderate active gastritis
B659-C2	corpus	60 y/o male	moderate active gastritis
B661A	antrum	67 y/o male	mild active gastritis
B679	antrum	61 y/o male	moderate active gastritis
B712A	antrum	47 y/o male	mild active gastritis
BMG112	antrum	75 y/o female	moderate active gastritis
CRL122	antrum	45 y/o female	mild active gastritis
CRM21	antrum	43 y/o male	severe active gastritis
JDX15	antrum	47 y/o male	severe active gastritis
JGF25	antrum	41 y/o female	metaplasia
JMM43	antrum	34 y/o male	metaplasia
JSS185-B120	antrum	62 y/o male	moderate active gastritis
MMV242	antrum	44 y/o female	metaplasia
MSL190	antrum	62 y/o male	metaplasia
VCT187-B122	antrum	54 y/o male	metaplasia

Table 5.1 *H. pylori* strains included in the project PRJNA449871. *From healthy tissue.

5.2.2 Bioinformatic analysis for genomic comparison of strains B508A-S1, B508A-T2A, and B508A-T4.

From all the studied genome sequences, the present study focused on just three of them, specifically strains B508A-S1, B508A-T2A, and B508A-T4. These were isolated from the stomach of a single patient with adenocarcinoma (Palau *et al.*, 2016). In particular, strains B508A-T2A and B508A-T4 were obtained from cancerous tissue and strain B508A-S1 from non-tumoral tissue. The interest on using these strains lies in the

finding described in the third chapter of this thesis, where the strain B508A-T4 was found genetically closer to the strain B508A-S1 than to the strain B508A-T2A.

The initial step in the genomic analysis was to linearize and order the sequences obtained through the Illumina sequencing. The alignment and rearrangement of the sequences was then achieved using the software Mauve (Darling *et al*., 2004). The draft genome annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (Tatusova *et al*., 2016).

In order to fully characterize the genomes, different analyses were carried out. Average Nucleotide Identity (OrthoANIu algorithm) (Lee *et al*., 2016), which is a value to measure the genomic similarity at nucleotide level, was calculated using EzBioCloud's ANI calculator (www.ezbiocloud.net/tools/ani). OrthoVenn2 (Xu *et al*., 2019) was used for the comparison of orthologous gene clusters. Their functions could be identified by performing a BLAST of the results. Snippy (Seemann, 2015) was used to find both substitutions and insertions/deletions (indels). The substitutions encompassed single nucleotide polymorphisms (SNP), multi-nucleotide polymorphisms (MNP) and complex polymorphisms (combinations of SNP and MNP). Later, snippy-core was used to combine the previous obtained snippy outputs into a core SNP alignment. Snippy and snippy-core were run on the Galaxy platform (Afgan *et al*., 2016) with the default parameters.

In order to visualize the results, CGView server (Stothard *et al*., 2017) was used to obtain a circular representation of the genomes. Additionally, Rapid Annotation Subsystem Technology (RAST) and SEED (Overbeek *et al*., 2014) servers were used to compare the subsystem distribution statistics between all strains. In this context, a subsystem is defined as a set of functional roles that together implement a specific biological process or structural complex, generally refereed as a pathway.

With the aim of finding pathogenic genes, Diamond (Buchfink *et al*., 2015) was used as a search tool against the Virulence-Factor database (VFDB). In addition, PCR was used to determine *H. pylori cagA* status and to genotype the *vacA* gene (signal,

intermediate and midregion polymorphisms), as described by (Lario *et al.*, 2012). PCR results were verified by looking for the primers in the sequence of the whole genomes and calculating the size of the fragments to determine the corresponding allele (Atherton *et al.*, 1999, 1995; Rhead *et al.*, 2007) (**Figure AIII.1**). Furthermore, PILER-CR (Edgar, 2007) and CRISPRCasFinder (Couvin *et al.*, 2018) were used for CRISPR identification.

ResFinder (Zankari *et al.*, 2012) was used to identify acquired antimicrobial resistance genes in the total sequenced isolates of bacteria. ResFinder consists of two subprograms, one for identifying acquired genes (*ResFinder.pl*), and the other for identifying chromosomal mutations (*PointFinder.py*). Sequences of genes involved in amoxicillin, clarithromycin, tetracycline, metronidazole and levofloxacin resistances were also manually checked –*pbp1* (Gerrits *et al.*, 2006; Nishizawa *et al.*, 2011; Rimbara *et al.*, 2008), 23S rRNA (Agudo *et al.*, 2011; Furuta *et al.*, 2007; Mahachai *et al.*, 2011; Stone *et al.*, 1997), 16S rRNA (Glocker *et al.*, 2005; Lawson *et al.*, 2005), *rdxA* (Binh *et al.*, 2015; Kwon *et al.*, 2000) and *gyrA* (Aires *et al.*, 2017; Trespalacios-Rangél *et al.*, 2016), respectively–.

For multilocus sequence typing (MLST) of the three genomes, seven housekeeping genes (*atpA*, *efp*, *mutY*, *ppa*, *trpC*, *ureI* and *yphC*) were extracted using BLAST and the JGI-database (https://img.jgi.doe.gov/). A total of 243 representative sequences from different *H. pylori* populations were downloaded with PubMLST (https://pubmlst.org/helicobacter/), and MAFFT (https://mafft.cbrc.jp/alignment/server/) was used to align all the sequences and to predict a neighbor-joining phylogenetic tree.

5.3 Results

5.3.1 Basic comparison of all the genomes

Concentrations after DNA extractions were measured with a Qubit fluorimeter using Broad Range standards. The obtained values for the 51 *H. pylori* genomes ranged from 20 to 330 ng/µl.

Basic characteristics of these genomes are shown in **Table 5.2**. The average of the genome size is 1,637,551 bp and the GC content 38.9%, ranging from 38.7% to 39.1%, perfectly fitting with the standards of the species, namely lower GC contents, which range from 38 to 39% (Ali *et al.*, 2015). On average 456-fold genome coverage of sequence data was used to create draft genome assemblies using the a5 assembler. The number of scaffolds per genome in each assembly ranged from 19 to 349 (average: 57.41 scaffolds) and a median N50 of 90,877 bp. Using Roary, a total of 1,077 clusters of orthologous sequences were identified as core genes (>95% of the strains) and 5,799 were present just in a subset of isolates (**Table 5.3**) The whole gene set, defined as the pan genome, was of 6,876 (**Table 5.3**). Between 1,553 and 2,033 genes were identified per genome. The studied genomes, genome statistics and number of unique genes are showed in **Table 5.2**.

Strain	Genome Size (bp)	N50	Genes	Unique genes	GenBank Accession number
AAP164	1,617,565	148,944	1,598	78	QDJV00000000
AFR58	1,654,636	102,441	1,655	106	QEHJ00000000
APR133	1,625,889	135,772	1,619	89	QEHI00000000
B126	1,674232	60,328	1,687	83	QEHH00000000
B247A	1,674,625	94,380	1,669	94	QDJU00000000
B274	1,666,918	85,845	1,685	146	QEHG00000000
B297	1,674,839	88,441	1,661	48	QEHF00000000
B314	1,686,647	77,520	1,683	82	QEHE00000000
B319	1,651,823	82,099	1,658	113	QELC00000000
B335	1,628,927	56,300	1,679	89	QDJT00000000
B344	1,548,362	101,134	1,551	60	QEHD00000000
B345	1,638,825	97,466	1,651	79	QEHC00000000
B355	1,666,644	73,617	1,665	113	QDJS00000000
B360	1,645,792	90,586	1,624	79	QEHB00000000
B366	1,625,481	109,955	1,624	69	QEGZ00000000
B368	1,669,066	77,651	1,684	68	QEGY00000000
B373	1,679,821	79,805	1,678	94	QDJR00000000
B400	1,655,874	61,632	1,666	69	QEGX00000000
B444A	1,644,603	19,989	1,683	77	QDJQ00000000

B448	1,666,264	80,874	1,675	87	QEGW00000000
B455	1,635,818	68,253	1,642	81	QEGV00000000
B464A	1,654,509	79,817	1,647	53	QDJP00000000
B491	1,689,659	65,438	1,701	100	QDJO00000000
B497A	1,610,841	100,745	1,615	87	QEGU00000000
B508A-S1	1,586,749	200,200	1,577	2	QDJN00000000
B508A-T2A	1,584,784	106,266	1,577	13	QDJM00000000
B508A-T4	1,585,256	124,923	1,581	3	QDJL00000000
B518	1,842,284	47,080	2,033	308	QELB00000000
B528A	1,573,376	202,365	1,573	67	QEGT00000000
B529	1,673,033	148,962	1,654	76	QDJK00000000
B547F	1,677,546	111,880	1,671	57	QDJJ00000000
B572A	1,667,087	112,418	1,639	59	QEGS00000000
B630	1,620,852	109,298	1,618	43	QEGR00000000
B657-A1	1,631,165	235,871	1,614	65	QDJI00000000
B657-A4	1,589,699	110,673	1,577	9	QDJH00000000
B657-C1	1,591,707	116,109	1,577	2	QDJG00000000
B659-A1	1,562,015	152,617	1,553	3	QDJF00000000
B659-C2	1,568,140	202,500	1,564	6	QDJE00000000
B661A	1,660,453	73,506	1,666	63	QEGQ00000000
B679	1,665,663	96,927	1,664	96	QEGP00000000
B712A	1,571,835	175,687	1,565	51	QEGO00000000
BMG112	1,568,426	175,883	1,568	93	QEGN00000000
CRL122	1,617,097	129,791	1,635	66	QDJD00000000
CRM21	1,650,793	87,678	1,636	136	QEGM00000000
JDX15	1,613,710	74,251	1,606	64	QEGK00000000
JGF25	1,629,619	104,106	1,630	56	QEGJ00000000
JMM43	1,618,025	72,594	1,615	73	QEGI00000000
JSS185-B120	1,640,940	70,448	1,660	38	QEGH00000000
MMV242	1,641,887	81,075	1,643	67	QEGG00000000
MSL190	1,620,073	141,775	1,588	92	QEGF00000000
VCT187-B122	1,675,252	83,036	1,656	68	QEGD00000000
Average	1,637,551	90,877	1,640	73	

Table 5.2 Main features of the 51 *H. pylori* genomes analysed.

Core genes	(99% <= strains <= 100%)	979
Soft core genes	(95% <= strains < 99%)	98
Shell genes	(15% <= strains < 95%)	720
Cloud genes	(0% <= strains < 15%)	5,079
Total genes	(0% <= strains <= 100%)	6,876

Table 5.3 Pan-genome statistics.

Phylogenetic analysis of the 51 sequenced genomes revealed that the evolutionary relationships of the selected strains were not entirely determined by the human host gastric health. As shown in **Figure 5.1**, all *H. pylori* isolates had relatively large unique branches, except for the samples from the same stomach, which were grouped together. Here, strains ATCC49504 and ATCC26695 were used as controls.

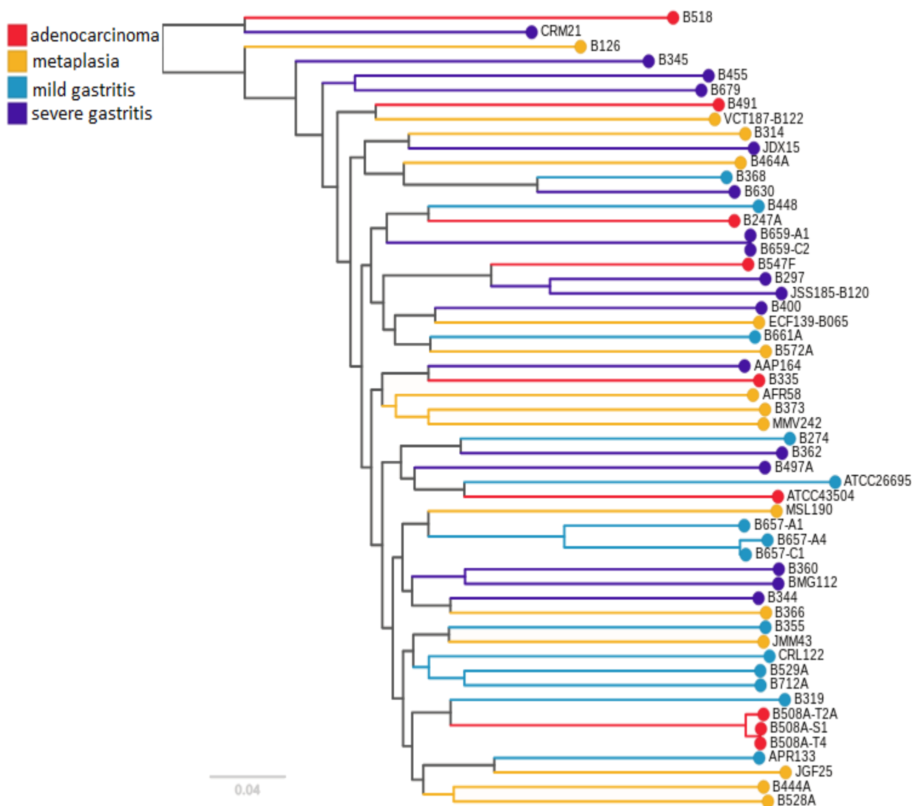


Figure 5.1 Phylogenetic tree from pan-genomes obtained with PanX.

5.3.2 B508A-S1, B508A-T2A, and B508A-T4 genome comparison

5.3.2.1 Additional characteristics of the genomes B508A-S1, B508A-T2A and B508A-T4

Focusing on the three strains selected for further analysis, **Table 5.4** shows additional features of interest, such as the total number of coding and RNA genes, *vacA* alleles, etc. The average genomic GC content for each strain was 39%. The genome length of B508A-S1, B508A-T2A and B508A-T4 was of 1,580,553 bp, 1,578,500 bp and 1,585,256 bp, respectively. The total number of genes was 1,577 for B508A-S1 and B508A-T2A, and 1,581 for B508A-T4. The genomes of B508A-S1, B508A-T2A and B508A-T4 contained a total of 1,445, 1,442 and 1,444 proteins or coding genes, respectively (**Table 5.4**).

Genome characteristics	B508A-S1	B508A-T2A	B508A-T4
Genes (total)	1,577	1,577	1,581
CDS (total)	1,535	1,535	1,536
CDS (coding)	1,445	1,442	1,444
Genes (RNA)	42	42	45
rRNAs (5S, 16S, 23S)	1,1,1	1,1,1	2,1[a],1
tRNAs	36	36	36
ncRNAs	3	3	3
Pseudo Genes	90	93	92
vacA (PCR)	s1-s2 / m1-m2	s1-s2 / m1-m2	s1-s2 / m1-m2
vacA (genome sequence)	s2 / m2/ i2	s1 / m2 / i2	s2 / m2 / i2
cagA	-	-	-

Table 5.4 Additional characteristics of the genomes B508A-S1, B508A-T2A and B508A-T4. [a]There is one complete sequence (1,500 bp) and two partial sequences (circa. 150 bp).

The ANI values were higher than 99.50% in all comparisons, as expected for strains of the same species. The most similar strains, with an ANI value of 99.94%, were B508A-S1 and B508A-T4. Regarding the B508A-T2A isolate, it had an ANI value of 99.58% with the other two strains.

5.3.2.2 OrthoVenn2

Comparison of the three genomes using OrthoVenn2 showed 1,435 clusters of orthologous genes, where 1,394 of them were shared among all strains, two between B508A-T2A and B508A-T4, 21 between B508A-T4 and B508A-S1, and 12 between B508A-S1 and B508A-T2A (**Figure 5.2**). The differences between them were mainly in outer membrane proteins, methylases and flagellar biosynthesis proteins (**Table AIII.1**). Regarding unique clusters, strain B508A-S1 possessed one, strain B508A-T4 had four and strain B508A-T2A also had just one. Each of the unique clusters had a double copy of the protein. The four unique clusters of B508A-T4 consisted of two Hop family outer membrane proteins, one transposase and one hypothetical protein. The unique cluster of B508A-T2A corresponded to an outer membrane beta-barrel protein. Additionally, the unique cluster of B508A-S1 was found to be a restriction endonuclease. In total, B508A-S1, B508A-T2A and B508A-T4 showed 1,428, 1,409 and 1,421 gene clusters, respectively (**Figure 5.2**).

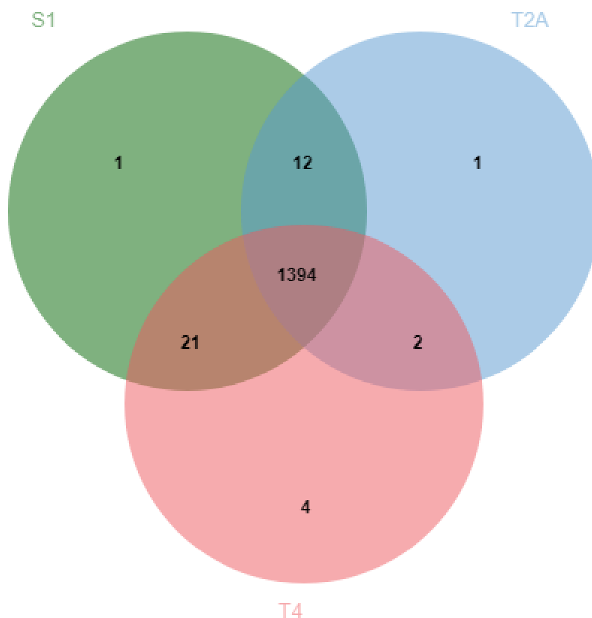


Figure 5.2 Venn Diagram of the three genomes obtained with OrthoVenn.

5.3.2.3 Roary

Focusing on the analysis of the unique genes using Roary, while the isolate B508A-T2A had 13 unique genes in the comparison of the complete set of genomes (**Table 5.2**), in the comparison of the three it was found to have 38 specific genes. These 13 genes were included on the 38, and they consist of 10 hypothetical proteins, a GTPase Era, a GTPase Der and the flagellar biosynthesis protein FlhA. For the isolate B508A-S1 there were four unique genes in the comparison of the three, and two in the comparison of the whole set. Regarding strain B508A-T4, it had twelve unique genes in the comparison of the three and three in the whole comparison. A graphical representation of the distribution of the unique genes of the comparison of the three genomes is shown in **Figure 5.3**. The complete annotation of every one of them can be found in the **Table AIII.2**. As can be observed, *H. pylori* genome contains a high proportion of unique genes that codify for hypothetical proteins with unknown function.

A further aspect to highlight is that the values above mentioned do not match the ones in **Table 5.2.** The reason for the mismatch is as follows. The number of unique genes obtained when performing the comparison of the full dataset, the 51 strains, will always be either equal or lower than the number of unique genes obtained when using a subset of the data, as in the current case. The reason being that some unique genes in the subset may be shared, and thus not unique, in the complete dataset. Comparison of the unique genes values in **Table 5.2** and **Figure 5.3** shows this trait.

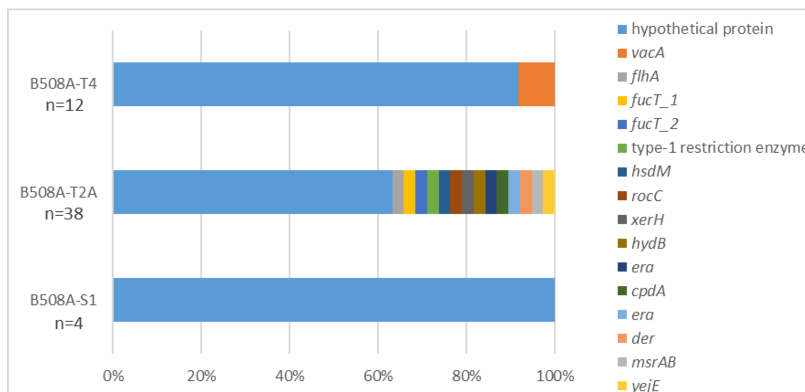


Figure 5.3 Relative distribution of the unique genes found with Roary.

5.3.2.4 Polymorphisms

Nucleotide polymorphisms and indels between genomes were obtained using Snippy. On the one hand, few variants, 169 in total, were found between B508A-S1 and B508A-T4. On the other hand, the number of variants between B508A-T2A and the other genomes was higher, specifically 3,344 with B508A-S1 and 3,316 with B508A-T4 (**Table 5.5**). This fact reflects the special distinction of the strain B508A-T2A with the other strains. Most of the variability was found in coding genes with unknown function (data not shown). Regarding the coding genes with known function, the greatest variability was found in flagelar proteins, proteins from the LPS and OMPs.

Type	**Number of polymorphisms**		
	B508A-S1 *v.* -T4	**B508A-S1 *v.* -T2**	**B508A-T2 *v.* -T4**
SNP[a]	113	2,454	2,418
MNP[b]	10	198	201
Complex	29	618	619
DEL[c]	12	35	38
INS[d]	5	39	40
Total	169	3,344	3,316

Table 5.5 Nucleotide polymorphisms between genomes. [a]Single nucleotide polymorphism, [b]Multi-nucleotide polymorphism, [c]Deletions, [d]Insertions.

5.3.2.5 Genome representation

The circular representation of the genomes obtained with GC-View is available in **Figure AIII.2**. They are markedly similar between them and no visual difference can be detected on this representation. It would be necessary to zoom-in the image significantly in order to be able to perceive the minor differences.

5.3.2.6 Subsystem Distribution Statistics

The functional distribution of the genomes was obtained with RAST server (**Figure AIII.3**). It showed the same subsystem distribution statistics for all the strains, with few differences in the categories of DNA and protein metabolism, motility and chemotaxis. Remarkably, in the genome B508A-T2A a unique protein was found, which was

classified as a type I restriction modification system, specifically subunit S (EC 3.1.21.3), which could be of interest because it is present on the virulence strain.

5.3.2.7 Virulence factors

The analysis of the results from Diamond shows that all three strains contain all the urease genes, the flagella forming genes and most of the virulence factors analyzed, including *vacA* and *iceA* (**Table 5.6**). On the contrary, they are *babA/hopS* and *sabB/hopO* negative. There are some differences in the number of copies in the *babB/hopT* adherence gene and in the genes involved in immune evasion *(futA*, *futB* and *futC* genes). All isolates have the *vacA i2/m2* genotype (**Table 5.4**). Furthermore, on the one hand the strain B508A-T2A has the *s1* allele. On the other hand, the isolates B508A-S1 and B508A-T4 are *vacA s2*, as evidenced by the size of the fragment within the primers (**Figure AIII.1**). It should be noted that conventional PCRs couldn't discriminate between alleles. None of the three isolates has the cagPAI nor any other genomic island (**Table 5.4**).

Virulence factors	related genes	B508A-S1	B508A-T2A	B508A-T4
Urease	[a]	1	1	1
Flagella	[b]	1	1	1
Lipopolysaccharide Lewis antigens	*futA*	1	2	1
	futB	1	0	1
	futC	2	2	1
Neutrophil-activating protein	*napA*	1	1	1
	oipA/hopH	1	1	1
Vacuolating cytotoxin	*vacA*	1	1	1
Adherence-associated lipoproteins	*alpA/hopC*	1	1	1
	alpB/hopB	1	1	1
Blood group antigen binding adhesins	*babA/hopS*	0	0	0
	babB/hopT	3	2	3
H. pylori adhesin A	*hpaA*	1	1	1
HopZ	*hopZ*	1	1	1
HorB	*horB*	1	1	1
Sialic acid binding proteins	*sabA/hopP*	1	1	1
	sabB/hopO	0	0	0
Duodenal ulcer promoting	*dupA*	1	1	1
Induced by contact with epithelium A[c]	*iceA*	1	1	1

Table 5.6 Virulence factors analyzed with Diamond. Gene copy number is indicated.
[a]Include all these genes: *ureA*, *ureB, ureI*, *ureE*, *ureF*, *ureG*, *ureH*.
[b]Include all these genes: flaA, flaB, flaG, fliR, flgI, flgL, flgH, flaG, fliF, fliG, fliH, flgG_1, flhA, flhF, fliA, fliM, fliY, fliN, fliP, fliD, fliS, flhB_1, fliL, motA, motB, flgE_1, flgD, flgE_2, flgK, flA, fliQ, fliI, flgA, fliE, flgC, flgB, flhB_2, flgG_2.
[c]Found with JGI IMG (https://img.jgi.doe.gov/).

5.3.2.8 CRISPR

Possible CRISPR sequences were sought using CRISPR-Finder. No confirmed sequence was found, only two questionable ones in each of the genomes. The sequences of the two spacers were identical in all cases and located in the middle of a coding gene which was annotated as a toxin. The locus tags for B508A-S1, B508A-T2A and B508A-T4 were DDP45_03600, DDP35_03380 and DDP36_03845, respectively. Furthermore, PILER-CR could not find any putative CRISPR array.

5.3.2.9 Antibiotic resistances

No resistance genes were found using ResFinder. The sequences of the genes *pbp1* (amoxicillin), 23S rRNA (clarythromicin), 16S rRNA (tetracycline), *rdxA* (metronidazole) and *gyrA* (levofloxacin) were manually checked to confirm this statement and, indeed, no mutation responsible for the resistance was found (**Table 5.7**).

Gen	Nucleotide position	Susceptible strains[a]	B508A-S1	B508A-T2	B508A-T4
pbp1	1120–2	GTG (= V)	GTG (= V)	GTG (= V)	GTG (= V)
	1216–8	GAA (= E)	GAA (= E)	GAA (= E)	GAA (= E)
	1240–2	AGC (= S)	AGC (= S)	AGC (= S)	AGC (= S)
	1777–9	ACA (= T)	ACA (= T)	ACA (= T)	ACA (= T)
	1795–7	GCG (= A)	GCG (= A)	GCG (= A)	GCG (= A)
	1801-3	GTG (= V)	GTG (= V)	GTG (= V)	GTG (= V)
23S rRNA	2143	A	A	A	A
	2144	A	A	A	A
	2183	T	T	T	T
	2196	C	C	C	C
	2224	A	A	A	A
16S rRNA	926	A	A	A	A
	927	G	G	G	G
	928	A	A	A	A
rdxA	3	G	G	G	G
	46	C	C	C	C
	238	G	G	G	G
	352	G	G	G	G
gyrA	256-8	GAT (= D)	GAT (= D)	GAT (= D)	GAT (= D)
	259-261	AAC (= N)	AAC (= N)	AAC (= N)	AAC (= N)
	261-3	GCT (= A)	GCG (= A)	GCG (= A)	GCG (= A)
	271-3	GAT (= D)	GAT (= D)	GAT (= D)	GAT (= D)

Table 5.7 Nucleotides found in the described mutational positions of each antibiotic. [a]Codons and/or amino acids are shown.

5.3.2.10 Genetic populations

In the phylogenetic tree obtained with MLST analysis four of the actual populations can be observed. The most distant is hpAfrica2, due to its ancestral evolution.

Population hpAfrica1 is clearly separated from hpAfrica2 and the other ones. The population hpEastAsia expands from within the hpEurope population, showing the fact that East Asian *H. pylori* diverged from European lineages, as suggested by Kawai *et al.* (2011). Finally, circled in this phylogenetic tree are the three strains of this study, which lie within the hpEurope population (**Figure 5.4**).

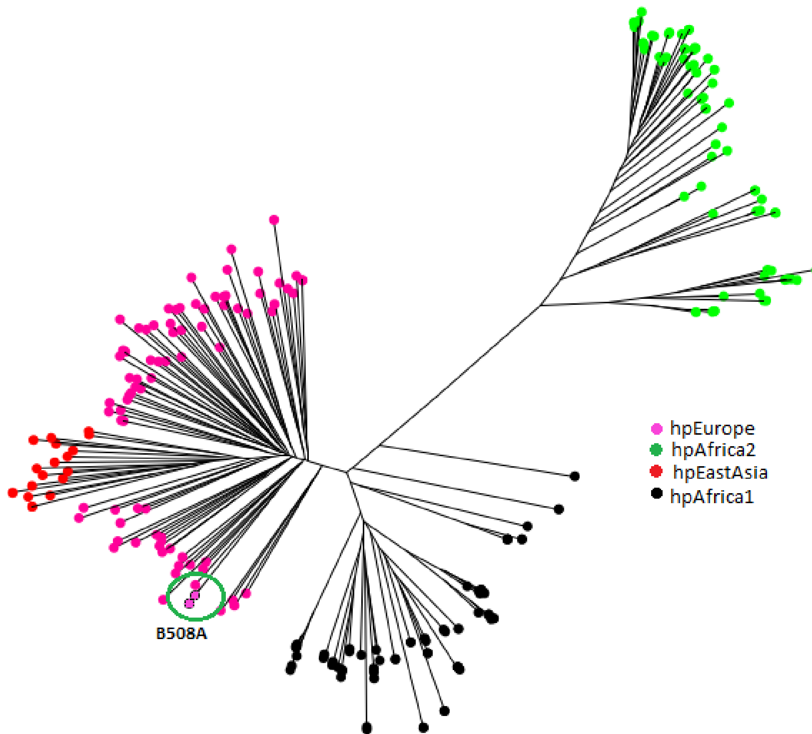


Figure 5.4 Phylogenetic tree showing the position of the three strains within the hpEurope population.

5.4 Discussion

5.4.1 Basic comparison of all the genomes

The characteristics of the 51 genomes included in this study (GC content, genome size, predicted genes number) (**Table 5.2**) suit with the description of *H. pylori* (Goodwin *et al.*, 1989; Vandamme *et al.*, 1991) and the 1,474 *H. pylori* genomes published until now (www.ncbi.nlm.nih.gov/genome/microbes/). Just one isolate of this study had a genome relatively larger than the rest, B518 is on the upper edge of the rang (1.35–1.91 Mb) with 1.84 Mb.

On this study, as shown in **Table 5.3**, the core genome is formed of 1,077 genes, where the core genome consists of the core and the soft core genes. The threshold was set at 95% because it must be taken into account that the gene prediction process may have missed some CDS and/or some CDS may not have been included in the final assembly due to the incompleteness of the sequencing. The core genome represented the 65% of the average gene number (1,640), which was calculated from the total number of genes for each genome (**Table 5.2**). The pan genome, known as the whole gene set of all the strains of the study, was of 6,876.

The core genome consists of the gene families that are shared by all the strains of a species. Additionally, its study is important to understand the minimum requirements for cellular life and to determine the major phenotypic characteristics. The percentage of core genes in prokaryotes varies widely from species to species, i.e. the core genome of the intracellular pathogen *Chlamydia trachomatis* represents the 84% of its pan genome. On the other hand, the core genome of the cyanobacteria *Prochlorococcus marinus* represents only an 18%; and in the case of *Escherichia coli*, this percentage falls to 3% (McInerney *et al.*, 2017). Bacteria that have a low percentage are those with higher genetic variability, and therefore, have more ability to migrate to new environments and can adapt to a whole variety of niches. On the other hand, those organisms that have a high percentage do not have this variability, and their adaptation is strictly preserved for a specific niche.

Regarding the core genome of *Helicobacter pylori*, in the study by Cao *et al.* (2016), the percentage of core genes on a set of 75 genomes was of 74%, and in other studies, when studying a group of 56 globally representative strains, the core essential genes were found to represent a 73% of their pan genome (Ali *et al.*, 2015; Gressmann *et al.*, 2005). On the other hand, McInerney *et al.* (2017) found the core genome of *H. pylori* to be the 48% of the total pan genome (0.95% cutoff).The discrepancies observed between the different studies, including ours (core genome of 66%), could be attributed to the number of genomes analysed, the origin of the *H. pylori* population and the panmictic structure of this species.

From the phylogenetic tree (**Figure 5.1**) it can be inferred that the current isolates cannot be distinguished by reason of pathogenesis on the basis of phylogenetic relationships of the genes included in the pan genome. Furthermore, strains B518 and CRM21 are remarkably distant from all the rest, pointing towards an interesting focus of study. When calculating the ANI values of these isolates with the type strain ATCC 43504^{T}, a value of 94.19% and 93.48% for CRM21 and B518 were obtained, respectively. These values bordered on the species definition (95–96% ANI) and further studies would be needed to know if they belong to another species or they represent a new one (Chun *et al.*, 2018).

Additionally, when looking at the genomes from strains isolated from gastric biopsies from the same patient, short distances can be observed (<0.04) (**Figure 5.1**), suggesting that they were the result of events of microevolution. It was also suggested in our previous study (see **CHAPTER 3**) that the group of strains B659 and B508 were considered to be the result of microevolution events. On the other hand, concerning the B657 isolates, the strain B657-A1 was suggested as a result of multi infection, because it was very distant from the rest (B657-A4 and B657-C1), as inferred from phylogenetic relationships of housekeeping genes (Palau *et al.*, 2016). However, the present results show that B657-A1 grouped with B657-A4 and B657-C1 although with a relatively high

distance (>0.16) (**Figure 5.1**). Therefore, in this case, is unclear if B657-A1 has evolved from microevolution events or represents a multi infection.

5.4.2 B508A-S1, B508A-T2A, and B508A-T4 genome comparison

Even though hundreds of *H. pylori* genomes have been published, few studies focused on different strains obtained from the same patient at a single point of time (Cao *et al.*, 2015). This study focused directly on three strains, isolated from a stomach suffering gastric cancer. The ANI values of the B508 isolates were ≥99.58%, being consistent with the definition of species (Goris *et al.*, 2007), and it is even more consistent considering they come from microevolution events. As shown in the pan genome tree and in Chapter 3, the strain B508A-T2A is the most distant among them. Therefore, the aim of the study was to search within these three strains for genetic markers that could be linked with the level of virulence, the pathogenicity or the risk to develop gastric cancer.

5.4.2.1 Diamond

Different virulence factors have been described until now, related to persistent colonization of the gastric mucosa, toxin expression or immune evasion (Kim and Blanke, 2012; Mascellino *et al.*, 2009; Salama *et al.*, 2013). Some of these pathogenicity factors have been associated with increased risk of gastric cancer: *cagA* + and *vacA s1i1m1* genotypes and the protein expression of AlpA, OipA, BabA, and SabA (Berthenet *et al.*, 2018; Pormohammad *et al.*, 2018; Su *et al.*, 2016).

The three studied genomes are *cagA*, *babA* and *sabB/hopO* negative. This is a surprising result for two main reasons. First, as mentioned above these genes have been clearly defined in the literature as gastric cancer markers and second, the strains B508A-T2A and B508A-T4 were isolated from tumoral tissue. Strain B508A-T2A is the only presenting the genotype *vacA s1*, showing a specific feature directly linked with pathogenicity, while the others are *s2*. On the other hand, the three strains have the coding genes for the proteins AlpA, OipA and SabA and other genes related to pathogenicity (**Table 5.6**).

The strain B508A-T2A is the most divergent isolate, as seen by multiple factors. One of them is the length of its genome, being the smallest of the three with a difference of 2,053 bp with B508A-S1 and 6,756 bp with B508A-T4 (**Table 5.2**), respectively. It also has fewer coding genes than the rest (**Table 5.4**), indicating a gene loss. Any change in selection pressure might contribute to a gene loss, which might be done in several ways: selection for smaller a genome, some genes become actively deleterious or some genes become less necessary (Weinert and Welch, 2017). Adoption of a pathogenic lifestyle can lead to recurrent changes in selection pressures, due to the host adaptive immune response (Brodsky and Medzhitov, 2009; Schulte *et al.*, 2013). Such changes have been observed to occur even within the same strain during the course of infection (Draper *et al.*, 2017). Hence, the present results are consistent with the previous studies where the event of microevolution was also detected (Palau *et al.*, 2016 and **CHAPTER 4**).

5.4.2.2 Roary

Unique genes obtained with Roary (**Figure 5.3**) show the relative abundance of the unique genes found in the three isolates. The strain B508A-T2A is the one with more unique genes. It can clearly be seen, that the area describing hypothetical proteins is much bigger than the rest, showing there is a high amount of undiscovered functions. Strain B508A-T4 showed a unique vacuolationg cytotoxin autotransporter, although caution must be taken when assigning its real function, because just a small percentage of these proteins have been proved to be able to translocate proteins (Fischer *et al.*, 2001). Consequently, further studies have to be done in this direction to understand the role of this protein.

5.4.2.3 RAST and OrthoVenn

As seen in the results, RAST and OrthoVenn analyses point at restriction modification systems, which have been described to have a relevant role in the regulation of gene expression and in modulating virulence (Ershova *et al.*, 2015). These restriction modification systems are important providers of defense against foreign DNA.

Furthermore, in order to avoid destroying its own DNA, methyl groups are added to the sequences by methyltransferases, which also appear in our results.

Also important to mention is that the amount of shared genes of strain B508-T2A with any of the other two strains is lower than the amount of shared genes between those other two strains (**Figure 5.2**). Thus, making strain B508-T2A a more differentiated strain.

5.4.2.4 Snippy

After studying all the types of polymorphisms between the three strains, this work continued with the focus on the differences between the most different strain, B508A-T2A, with the other two. Specifically, nonsense polymorphisms were sought. Being B508A-T2A the most differentiated strain and isolated from tumoral tissue, the search for these differences serves to highlight possible signs of differentiations between virulent and none-virulent strains. Even though only one strain is here considered, the current results provide hints to narrow the search for possible candidates.

Changes in the expression of bacterial surface structures, such as OMPs, are anticipated to facilitate adaption of the bacterium to the new human host (Linz *et al.*, 2013). A total of 21 non-synonymous or missense SNPs were related to outer membrane proteins. Specifically, on the *hof* family of adhesins, six were situated in the *hofB* gene, two in the *hofG* gene and one in the *hofH* gene. The other twelve involved *bamA*, an outer membrane protein assembly factor.

Additionally, eight non-synonymous SNPs were detected on the *fliI* gene and a single one in the *flgG* gene. This differences may be of relevance since colonization is the basis of the inflammatory reaction induced by *H. pylori* (Gu, 2017). Consequently, the motility of *H. pylori*, and in particular the flagellum, is a critical colonization determinant that affects the infection outcome.

The LPS of this organism plays a key role in its colonization and persistence in the stomach. In addition, the LPS of *H. pylori* modulates pathogen-induced host inflammatory responses. These responses may result in chronic inflammation within the

gastrointestinal tract. Very little is known about the LPS compositions of different strains of *H. pylori* with varied degree of virulence in human (Leker *et al.*, 2017). Here, 27 missense SNPs in the same LPS biosynthesis protein, a glycosyltransferase, were found in the strain B508A-T2A when comparing them with the other two strains.

5.4.2.5 CRISPR

Although CRISPR-Cas systems are of vital importance in the immunological defense of certain bacteria, as they confer resistance to foreign genetic elements, no *cas* genes could be detected in this work, and neither CRISPR-like sequences could be defined. From the bibliography, CRISPR-like loci have been identified, but no CRISPR-Cas system has been found so far (García-Zea *et al.*, 2019).

5.4.2.6 Antibiotic resistances

Regarding the study of antibiotic resistances, all three strains were defined as susceptible to the four antibiotics studied (amoxicillin, clarithromycin, tetracycline and levofloxacin), as none of them contained any mutation in the sites responsible of the resistances. As commented, *H. pylori* is nowadays in the high priority list of the WHO in the research and discovery of new antibiotics. Resistances have arisen in the last years, leading to suboptimal eradication rates (Alba *et al.*, 2017; World Health Organization, 2017). Here, however, these three strains show no mutations linked with resistances This could be explain by multiple factors, as for example the patient history of antibiotics intake.

5.4.2.7 MLST

Regarding the MLST analysis and its predicted phylogenetic tree (**Figure 5.4**), the three isolates of study were from the group within the hpEurope population, so the statement that the patient was European can be made.

With all the accumulated information no significant differences were found between the isolates regarding the virulence and the origin of the isolates. Still, considering that B508A-S1 and B508A-T4 are more closely related and that they originate from non-

tumoral and tumoral tissue, respectively, it could be inferred that their characteristics/genes are not a cause for developing a cancerous process. On the other hand, regarding B508A-T2, being markedly distant from the other two and having been found in the tumoral tissue, it can be hypothesised that some of its characteristics could be linked to the pathogenicity of *Helicobacter pylori*. A further consideration to take into account is that even though the strains were extracted from a cancerous tissue, they are all missing some relevant virulence genes like *cagA*, *babA* or *sabB.* Nonetheless, more exhaustive analyses are needed.

CHAPTER 6 **CONCLUSIONS**

In the first part of this study, we aimed to study the usefulness of six housekeeping genes for the diagnosis of *Helicobacter pylori* infection, strain discrimination and detection of multiple infection. The main findings from this part are:

1. A specific detection method for *H. pylori* infection was developed by using PCR amplification of six housekeeping genes (*amiA*, *cgt*, *cpn60*, *cpn70*, *dnaJ*, and *luxS*) with primers designed specifically in this study.
2. All the studied clones isolated from biopsies of different patients were discriminated between them with the analysis of the housekeeping gene sequences, as individual genes (0.9–7.8% distances) and as concatenated (2.4–3.6% distances).
3. Microevolution and multiple infection events were detected via the amplification and sequencing of the six housekeeping genes for the first time.
4. Microevolution was detected from clones isolated from the gastric biopsy B508T for the genes *amiA* and *cpn60*.
5. Multiple infection was detected by all genes from clones isolated from gastric biopsy B657A.
6. Results obtained suggest that studies based on these housekeeping genes could be useful for *H. pylori* detection in gastric biopsies and other specimens (such as gastric juice, stool, saliva, dental plaque, water and food samples) and for elucidation of the mode of *H. pylori* transmission.

In the second part of this study, we aimed to characterize the events of multiple infection and microevolution by means of amplicon sequencing of two housekeeping genes (*cgt* and *luxS*) from gastric biopsies. The most relevant findings from this study are:

7. This work is the first study that uses amplicon sequencing of housekeeping genes to detect *H. pylori* multi infections and microevolution.

8. From the five gastric biopsies analyzed, obtained from four different patients, between 11–25% of *cgt* amplicon sequences were unique sequences distributed in nine OTUs. In the case of *luxS*, between 20–43% of amplicon sequences were unique sequences distributed in 13 OTUs.
9. For both genes, at least two different OTUs were detected for all samples and most amplicon sequences (>90%) clustered in four OTUs, each one of them corresponding to a different patient. Most amplicon sequences from gastric biopsies B508S and B508T, obtained from the same patient, clustered in the same OTU.
10. The representative sequences of OTU-01, OTU-02, and OTU-03, both in *cgt* and *luxS*, matched with sequences of strains isolated from replicate gastric biopsies of the same patient (B247, B373, and B508, respectively).
11. These results suggest that *H. pylori* colonizes the human stomach through diverse infection circumstances that lead to a gastric multi-infection with a predominant strain together alongside other minority strains, where all strains are subjected to microevolution events.
12. Evidence of mixed infection is of crucial importance because the minority strains, which are the most difficult to detect, could be carrying antibiotic resistance genes, and be in charge of the failure of the treatment.
13. Amplicon sequencing of housekeeping, virulence or antibiotic resistant genes could be useful in different fields as epidemiological, pathogenicity or treatment studies.

In the last part of this thesis we aimed to compare three related *H. pylori* genomes extracted from the same stomach using whole genome sequencing. We also performed a basic comparison of 51 *H. pylori* whole genome sequences. The outcome of the study led the following conclusions:

14. The analysis of the 51 H. pylori whole genomes led to a core genome of 1,077 genes representing the 66% of the pan genome (average of 1,640 genes and

total of 6,876 genes). The average genome size and GC content were 1.6 Mb and 38.9%, respectively. These results were in concordance with those described in the literature.

15. The genome analysis of the strains B508A-S1, B508A-T2A and B508A-T4 revealed that they were cagA, babA and sabB/hopO negatives. These results were unexpected since these genes are described as important H. pylori virulence markers and strains analyzed were isolated from a patient with gastric adenocarcinoma. The strain B508A-T2A is the only presenting the genotype vacA s1, showing a specific feature linked with pathogenicity.
16. The differences between the three genomes were mainly in outer membrane proteins, methylases, restriction modification systems and flagellar biosynthesis proteins.
17. B508-T2A genome was the most distinct genome as lower amount of shared genes, higher number of unique genes, and more polymorphisms were found in this genome.
18. No mutation responsible for the resistances to the antibiotics amoxicillin, clarythromicin, tetracycline, metranidazone and levofloxacin were found in any genome.
19. The three strains were grouped together and lie within the hpEurope population
20. With all the accumulated information no significant differences were found between the isolates regarding the virulence and the origin of the isolates. Still, considering that B508A-S1 and B508A-T4 are more closely related and that they were isolated from non-tumoral and tumoral tissue, respectively, it could be inferred that their characteristics/genes are not a cause for developing a cancerous process. On the other hand, regarding B508A-T2A, being distant from the other two and having been found in the tumoral tissue, it can be hypothesised that some of its characteristics could be linked to the pathogenicity of H. pylori.

FUTURE DIRECTIONS

The work developed in this thesis can be regarded as a first approach to study and compare whole genome sequences by means of bioinformatics at the Microbiology Section of the Faculty of Pharmacy and Food Sciences of the Universitat de Barcelona. The next step would be to expand the study and make a comparison using the whole set of isolates, namely the 51 genomes. It would also be very interesting to compare the *Helicobacter pylori* genome of an 80-years-old man with no pathology and a *s2m2 vacA*, with the *H. pylori* genome of a 40-years-old man with cancer and a *s1m1 vacA*. The focus would be on looking for polymorphisms and also trying to elucidate the functions of the hypothetical proteins. Overall, we would continue attempting to find new markers of virulence and of progression to gastric diseases. Additionally, regarding the first work of this thesis, it would be interesting to study the usefulness of the housekeeping genes in a bigger and a more diverse quantity of samples to improve the detection and elucidate the mode of transmission. Finally, regarding the study of amplicon sequencing, it would be interesting to use the Illumina technology in a broader set of samples.

REFERENCES

Abu-Taleb, A.M., Abdelattef, R.S., Abdel-Hady, A.A., Omran, F.H., El-Korashi, L.A., Abdel-Aziz El-Hady, H., El-Gebaly, A.M., 2018. Prevalence of *Helicobacter pylori* cagA and *iceA* genes and their association with gastrointestinal diseases. *International Journal of Microbiology* 2018, 1–7. doi:10.1155/2018/4809093

Achtman, M., Azuma, T., Berg, D.E., Ito, Y., Morelli, G., Pan, Z.J., Suerbaum, S., Thompson, S.A., Van Der Ende, A., Van Doorn, L.J., 1999. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Molecular Microbiology* 32(3), 459–470. doi:10.1046/j.1365-2958.1999.01382.x

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., Grüning, B., Guerler, A., Hillman-Jackson, J., Von Kuster, G., Rasche, E., Soranzo, N., Turaga, N., Taylor, J., Nekrutenko, A., Goecks, J., 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses. *Nucleic Acids Research* 44(W1), W3–W10. doi:10.1093/nar/gkw343

Agudo, S., Pérez-Pérez, G., Alarcón, T., 2011. Rapid detection of clarithromycin resistant *Helicobacter pylori* strains in Spanish patients by polymerase chain reaction-restriction fragment length polymorphism. *Revista Española de Quimioterapia* 24(1), 32–36.

Aires, B., Palma, G.Z. De, Mendiondo, N., Viola, L., Ibarra, D., Campitelli, E., Salim, N., Corti, R., Goldman, C., Catalano, M., 2017. Occurrence of mutations in the antimicrobial target genes related to levofloxacin, clarithromycin, and amoxicillin resistance in *Helicobacter pylori*. *Microbial Drug Resistance* 23(3), 351–358. doi:10.1089/mdr.2015.0361

Akopyants, N.S., Clifton, S.W., Kersulyte, D., Crabtree, J.E., Youree, B.E., Reece, C.A., Bukanov, N.O., Drazek, E.S., Roe, B.A., Berg, D.E., 1998. Analyses of the *cag* pathogenicity island of *Helicobacter pylori*. *Molecular Microbiology* 28(1), 37–53.

doi:10.1046/j.1365-2958.1998.00770.x

Alba, C., Blanco, A., Alarcón, T., 2017. Antibiotic resistance in *Helicobacter pylori*. *Current Opinion in Infectious Diseases* 30(5), 489–497. doi:10.1097/QCO.0000000000000396

Ali, A., Naz, A., Soares, S.C., Bakhtiar, M., Tiwari, S., Hassan, S.S., Hanan, F., Ramos, R., Pereira, U., Barh, D., Figueiredo, H.C.P., Ussery, D.W., Miyoshi, A., Silva, A., Azevedo, V., 2015. Pan-genome analysis of human gastric pathogen *H. pylori*: Comparative genomics and pathogenomics approaches to identify regions associated with pathogenicity and prediction of potential core therapeutic targets. *BioMed Research International* 2015, 1–17. doi:10.1155/2015/139580

Allan, E., Dorrell, N., Foynes, S., Anyim, M., Wren, B.W., 2000. Mutational analysis of genes encoding the early flagellar components of *Helicobacter pylori*: Evidence for transcriptional regulation of flagellin A biosynthesis. *Journal of Bacteriology* 182(18), 5274–5277. doi:10.1128/JB.182.18.5274-5277.2000

Atherton, J.C., 2006. The pathogenesis of *Helicobacter pylori*–induced gastro-duodenal diseases. *Annual Review of Pathology: Mechanisms of Disease* 1(1), 63–96. doi:10.1146/annurev.pathol.1.110304.100125

Atherton, J.C., Cao, P., Peek, R.M., Tummuru, M.K.R., Blaser, M.J., Cover, T.L., 1995. Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. Association of specific *vacA* types with cytotoxin production and peptic ulceration. *Journal of Biological Chemistry* 270(30), 17771–17777. doi:10.1074/jbc.270.30.17771

Atherton, J.C., Cover, T.L., Twells, R.J., Morales, M.R., Hawkey, C.J., Blaser, M.J., 1999. Simple and accurate PCR-based system for typing vacuolating cytotoxin alleles of *Helicobacter pylori*. *Journal of Clinical Microbiology* 37(9), 2979–2982.

Atkinson, N.S.S., Braden, B., 2016. *Helicobacter pylori* infection: Diagnostic strategies in primary diagnosis and after therapy. *Digestive Diseases and Sciences* 61, 19–24. doi:10.1007/s10620-015-3877-4

Ayala, G., Escobedo-Hinojosa, W.I., de la Cruz-Herrera, C.F., Romero, I., 2014.

Exploring alternative treatments for *Helicobacter pylori* infection. *World Journal of Gastroenterology* 20(6), 1450–1469. doi:10.3748/wjg.v20.i6.1450

Backert, S., Clyne, M., 2011. Pathogenesis of *Helicobacter pylori* infection. *Helicobacter* 16(s1), 19–25. doi:10.1111/j.1523-5378.2011.00876.x

Baltrus, D.A., Blaser, M.J., Guillemin, K., 2009. *Helicobacter pylori* genome plasticity. In: Microbial Pathogenomics. KARGER, Basel, pp. 75–90. doi:10.1159/000235764

Baylina, M., Muñoz, N., Sánchez-Delgado, J., López-Góngora, S., Calvet, X., Puig, I., 2019. Systematic review: Would susceptibility-guided treatment achieve acceptable cure rates for second-line *Helicobacter pylori* therapy as currently practiced? *Helicobacter* 24(3), e12584. doi:10.1111/hel.12584

Beckman, E., Saracino, I., Fiorini, G., Clark, C., Slepnev, V., Patel, D., Gomez, C., Ponaka, R., Elagin, V., Vaira, D., 2017. A novel stool PCR test for *Helicobacter pylori* may predict clarithromycin resistance and eradication of infection at a high rate. *Journal of Clinical Microbiology* 55(8), 2400–2405. doi:10.1128/jcm.00506-17

Berthenet, E., Yahara, K., Thorell, K., Pascoe, B., Meric, G., Mikhail, J.M., Engstrand, L., Enroth, H., Burette, A., Megraud, F., Varon, C., Atherton, J.C., Smith, S., Wilkinson, T.S., Hitchings, M.D., Falush, D., Sheppard, S.K., 2018. A GWAS on *Helicobacter pylori* strains points to genetic variants associated with gastric cancer risk. *BMC Biology* 16(1), 84. doi:10.1186/s12915-018-0550-3

Binh, T.T., Suzuki, R., Huyen, T., Kwon, H., 2015. Search for novel candidate mutations for metronidazole resistance in *Helicobacter pylori* using next-generation sequencing. *Antimicrobial Agents and Chemotherapy* 59(4), 2343–2348. doi:10.1128/AAC.04852-14

Blaser, M.J., Chen, Y., Reibman, J., 2008. Does *Helicobacter pylori* protect against asthma and allergy? *Gut* 57(5), 561–567. doi:10.1136/gut.2007.133462

Brodsky, I.E., Medzhitov, R., 2009. Targeting of immune signalling networks by bacterial pathogens. *Nature Cell Biology* 11(5), 521–526. doi:10.1038/ncb0509-521

Brown, L.M., 2000. *Helicobacter pylori*: epidemiology and routes of transmission. *Epidemiologic Reviews* 22(2), 283–297. doi:https://doi.org/10.1093/oxfordjournals.epirev.a018040

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 12, 59–60. doi:10.1038/nmeth.3176

Buck, G.E., Smith, J.S., 1987. Medium supplementation for growth of *Campylobacter pyloridis*. *Journal of Clinical Microbiology* 25(4), 597–599.

Burne, R.A., Chen, Y.-Y.M., 2000. Bacterial ureases in infectious diseases. *Microbes and Infection* 2(5), 533–542. doi:10.1016/S1286-4579(00)00312-9

Bush, K., 2001. New beta-lactamases in gram-negative bacteria: diversity and impact on the selection of antimicrobial therapy. *Clinical Infectious Diseases* 32(7), 1085–1089. doi:10.1086/319610

Cao, D.M., Lu, Q.F., Li, S.B., Wang, J.P., Chen, Y.L., Huang, Y.Q., Bi, H.K., 2016. Comparative genomics of *H. pylori* and non-pylori *Helicobacter* species to identify new regions associated with its pathogenicity and adaptability. *BioMed Research International* 2016, 1–15. doi:10.1155/2016/6106029

Cao, Q., Didelot, X., Wu, Z., Li, Zongwei, He, L., Li, Y., Ni, M., You, Y., Lin, X., Li, Zhen, Gong, Y., Zheng, M., Zhang, M., Liu, J., Wang, W., Bo, X., Falush, D., Wang, S., Zhang, J., 2015. Progressive genomic convergence of two *Helicobacter pylori* strains during mixed infection of a patient with chronic gastritis. *Gut* 64(4), 554–561. doi:10.1136/gutjnl-2014-307345

Cervantes, D.T., Fischbach, L.A., Goodman, K.J., Phillips, C. V., Chen, S., Broussard, C.S., 2010. Exposure to *Helicobacter pylori*-positive siblings and persistence of *Helicobacter pylori* infection in early childhood. *Journal of Pediatric Gastroenterology and Nutrition* 50(5), 481–485. doi:10.1097/MPG.0b013e3181bab2ee

Chaput, C., Ecobichon, C., Cayet, N., Girardin, S.E., Werts, C., Guadagnini, S., Prévost, M.-C., Mengin-Lecreulx, D., Labigne, A., Boneca, I.G., 2006. Role of AmiA in the

morphological transition of *Helicobacter pylori* and in immune escape. *PLoS Pathogens* 2(9), e97. doi:10.1371/journal.ppat.0020097

Chen, Y., Blaser, M.J., 2007. Inverse associations of *Helicobacter pylori* with asthma and allergy. *Archives of Internal Medicine* 167(8), 821. doi:10.1001/archinte.167.8.821

Chiurillo, M.A., Moran, Y., Cañas, M., Valderrama, E., Granda, N., Sayegh, M., Ramírez, J.L., 2013. Genotyping of *Helicobacter pylori* virulence-associated genes shows high diversity of strains infecting patients in Western Venezuela. *International Journal of Infectious Diseases* 17(9), e750–e756. doi:10.1016/j.ijid.2013.03.004

Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D.R., da Costa, M.S., Rooney, A.P., Yi, H., Xu, X.W., De Meyer, S., Trujillo, M.E., 2018. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology* 68(1), 461–466. doi:10.1099/ijsem.0.002516

Coil, D., Jospin, G., Darling, A.E., 2015. A5-miseq: An updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 31(4), 587–589. doi:10.1093/bioinformatics/btu661

Correa, P., 2013. Gastric cancer: overview. *Gastroenterology Clinics of North America* 42(2), 211–217. doi:10.1016/j.gtc.2013.01.002

Correa, P., Haenszel, W., Cuello, C., Tannenbaum, S., Archer, M., 1975. A model for gastric cancer epidemiology. *The Lancet* 306(7924), 58–60. doi:10.1016/S0140-6736(75)90498-5

Correa, P., Houghton, J., 2007. Carcinogenesis of *Helicobacter pylori*. *Gastroenterology* 133(2), 659–672. doi:10.1053/j.gastro.2007.06.026

Correa, P., Piazuelo, M.B., Wilson, K.T., 2010. Pathology of gastric intestinal metaplasia: Clinical implications. *American Journal of Gastroenterology* 105(3), 493–498. doi:10.1038/ajg.2009.728

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B.,

Rocha, E.P.C., Vergnaud, G., Gautheret, D., Pourcel, C., 2018. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Research* 46(W1), W246–W251. doi:10.1093/nar/gky425

Cover, T.L., 2012. Perspectives on methodology for *in vitro* culture of *Helicobacter pylori*. *Methods in Molecular Biology* 921(2), 11–15. doi:10.1007/978-1-62703-005-2

Cover, T.L., 2016. *Helicobacter pylori* diversity and gastric cancer risk. *mBio* 7(1), e01869-15. doi:10.1128/mBio.01869-15

Cover, T.L., Blaser, M.J., 2009. *Helicobacter pylori* in Health and Disease. *Gastroenterology* 136(6), 1863–1873. doi:10.1053/j.gastro.2009.01.073

Cullen, T.W., Giles, D.K., Wolf, L.N., Ecobichon, C., Boneca, I.G., Trent, M.S., 2011. *Helicobacter pylori* versus the host: Remodeling of the bacterial outer membrane is required for survival in the gastric mucosa. *PLoS Pathogens* 7(12), e1002454. doi:10.1371/journal.ppat.1002454

da Costa, D.M., dos Santos Pereira, E., Rabenhorst, S.H.B., 2015. What exists beyond *cagA* and *vacA*? *Helicobacter pylori* genes in gastric diseases. *World Journal of Gastroenterology* 21(37), 10563–10572. doi:10.3748/wjg.v21.i37.10563

Dabiri, H., Jafari, F., Baghaei, K., Shokrzadeh, L., Abdi, S., Pourhoseingholi, M.A., Mohammadzadeh, A., 2017. Prevalence of *Helicobacter pylori* vacA, *cagA*, *cagE*, *oipA*, *iceA*, *babA2* and *babB* genotypes in Iranian dyspeptic patients. *Microbial Pathogenesis* 105, 226–230. doi:10.1016/j.micpath.2017.02.018

Darling, A.C.E., Mau, B., Blattner, F.R., Perna, N.T., 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* 14, 1394–1403. doi:10.1101/gr.2289704

Di Felice, V., David, S., Cappello, F., Farina, F., Zummo, G., 2005. Is chlamydial heat shock protein 60 a risk factor for oncogenesis? *Cellular and Molecular Life Sciences CMLS* 62(1), 4–9. doi:10.1007/s00018-004-4367-6

Ding, W., Baumdicker, F., Neher, R.A., 2018. panX: pan-genome analysis and

exploration. *Nucleic Acids Research* 46(1), e5. doi:10.1093/nar/gkx977

Dominguez-Bello, M.G., Blaser, M.J., 2011. The human microbiota as a marker for migrations of individuals and populations. *Annual Review of Anthropology* 40(1), 451–474. doi:10.1146/annurev-anthro-081309-145711

Draper, J.L., Hansen, L.M., Bernick, D.L., Abedrabbo, S., Underwood, J.G., Kong, N., Huang, B.C., Weis, A.M., Weimer, B.C., van Vliet, A.H.M., Pourmand, N., Solnick, J. V., Karplus, K., Ottemann, K.M., 2017. Fallacy of the unique genome: Sequence diversity within single *Helicobacter pylori* strains. *mBio* 8(1), e02321-16. doi:10.1128/mBio.02321-16

Dunn, B.E., Cohen, H., Blaser, M.J., 1997. *Helicobacter pylori*. *Clinical Microbiology Reviews* 10(4), 720–741.

Edgar, R.C., 2007. PILER-CR: Fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* 8(1), 18. doi:10.1186/1471-2105-8-18

Ershova, A.S., Rusinov, I.S., Spirin, S.A., Karyagina, A.S., Alexeevski, A. V., 2015. Role of restriction-modification systems in prokaryotic evolution and ecology. *Biochemistry (Moscow)* 80(10), 1373–1386. doi:10.1134/S0006297915100193

Eusebi, L.H., Zagari, R.M., Bazzoli, F., 2014. Epidemiology of *Helicobacter pylori* infection. *Helicobacter* 19(s1), 1–5. doi:10.1111/hel.12165

Falush, D., Kraft, C., Taylor, N.S., Correa, P., Fox, J.G., Achtman, M., Suerbaum, S., 2001. Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: Estimates of clock rates, recombination size, and minimal age. *Proceedings of the National Academy of Sciences of the United States of America* 98(26), 15056–15061. doi:10.1073/pnas.251396098

Falush, D., Wirth, T., Linz, B., Pritchard, J.K., Stephens, M., Kidd, M., Blaser, M.J., Graha, D.Y., Vacher, S., Perez-Perez, G.I., Yamaoka, Y., Mégraud, F., Otto, K., Reichard, U., Katzowitsch, E., Wang, X., Achtman, M., Suerbaum, S., 2003. Traces of human migrations in *Helicobacter pylori* populations. *Science* 299(5612), 1582–1585. doi:10.1126/science.1080857

Fischer, W., Buhrdorf, R., Gerland, E., Haas, R., 2001. Outer membrane targeting of passenger proteins by the Vacuolating cytotoxin autotransporter of *Helicobacter pylori*. *Infection and Immunity* 69(11), 6769–6775. doi:10.1128/IAI.69.11.6769-6775.2001

Forman, D., Newell, D.G., Fullerton, F., Yarnell, J.W., Stacey, A.R., Wald, N., Sitas, F., 1991. Association between infection with *Helicobacter pylori* and risk of gastric cancer: evidence from a prospective investigation. *The BMJ* 302, 1302–1305. doi:10.1136/bmj.302.6788.1302

Franceschi, F., Gasbarrini, A., Polyzos, S.A., Kountouras, J., 2015. Extragastric diseases and *Helicobacter pylori*. *Helicobacter* 20(s1), 40–46. doi:10.1111/hel.12256

Furuta, T., Soya, Y., Sugimoto, M., Shirai, N., Nakamura, A., Kodaira, C., Nishino, M., Okuda, M., Okimoto, T., Murakami, K., Fujioka, T., Hishida, A., 2007. Modified allele-specific primer–polymerase chain reaction method for analysis of susceptibility of *Helicobacter pylori* strains to clarithromycin. *Journal of Gastroenterology and Hepatology* 22(11), 1810–1815. doi:10.1111/j.1440-1746.2007.04919.x

García-Zea, J.A., Herrán, R. de la, Rodríguez, F.R., Navajas-Pérez, R., Rejón, C.R., 2019. Detection and variability analyses of CRISPR-like loci in the *H. pylori* genome. *PeerJ* 2019(1), e6221. doi:10.7717/peerj.6221

Geis, G., Suerbaum, S., Forsthoff, B., Leying, H., Opferkuch, W., 1993. Ultrastructure and biochemical studies of the flagellar sheath of *Helicobacter pylori*. *Journal of Medical Microbiology* 38(5), 371–377. doi:10.1099/00222615-38-5-371

Gerrits, M.M., Godoy, A.P., Kuipers, E.J., Ribeiro, M.L., Stoof, J., Mendonça, S., Van Vliet, A.H., Pedrazzoli, J., Kusters, J.G., 2006. Multiple mutations in or adjacent to the conserved penicillin-binding protein motifs of the penicillin-binding protein 1A confer amoxicillin resistance to *Helicobacter pylori*. *Helicobacter* 11(3), 181–187. doi:10.1111/j.1523-5378.2006.00398.x

Gevers, D., Cohan, F., Lawrence, J., Spratt, B., Coenye, T., Feil, E., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, F., Swings, J., 2005. Opinion: Re-evaluating prokaryotic species. *Nature Reviews Microbiology* 3(9), 733–739. doi:10.1038/nrmicro1236

Ghose, C., Pérez-Pérez, G.I., Van Doorn, L.J., Domínguez-Bello, M.G., Blaser, M.J., 2005. High frequency of gastric colonization with multiple *Helicobacter pylori* strains in Venezuelan subjects. *Journal of Clinical Microbiology* 43(6), 2635–2641. doi:10.1128/JCM.43.6.2635-2641.2005

Glocker, E., Berning, M., Gerrits, M.M., Kusters, J.G., Kist, M., 2005. Real-Time PCR screening for 16S rRNA mutations associated with resistance to tetracycline in *Helicobacter pylori*. *Antimicrobial Agents and Chemotherapy* 49(8), 3166–3170. doi:10.1128/AAC.49.8.3166

Go, M.F., Kapur, V., Graham, D.Y., Musser, J.M., 1996. Population genetic analysis of *Helicobacter pylori* by multilocus enzyme electrophoresis: extensive allelic diversity and recombinational population structure. *Journal of Bacteriology* 178(13), 3934–3938. doi:10.1128/jb.178.13.3934-3938.1996

Goderska, K., Agudo Pena, S., Alarcón, T., 2018. *Helicobacter pylori* treatment: antibiotics or probiotics. *Applied Microbiology and Biotechnology* 102(1), 1–7. doi:10.1007/s00253-017-8535-7

González, C.A., Figueiredo, C., Lic, B.C., Ferreira, R.M., Pardo, M.L., Liso, R.J.M., Alonso, P., Sala, N., Capella, G., Sanz-Anquela, J.M., 2011. *Helicobacter pylori cagA* and *vacA* genotypes as predictors of progression of gastric preneoplastic lesions: A long-term follow-up in a high-risk area in Spain. *American Journal of Gastroenterology* 106(5), 867–874. doi:10.1038/ajg.2011.1

González, C.A., Jakszyn, P., Pera, G., Agudo, A., Bingham, S., Palli, D., Ferrari, P., Boeing, H., del Giudice, G., Plebani, M., Carneiro, F., Nesi, G., Berrino, F., Sacerdote, C., Tumino, R., Panico, S., Berglund, G., Simán, H., Nyrén, O., Hallmans, G., Martinez, C., Dorronsoro, M., Barricarte, A., Navarro, C., Quirós,

J.R., Allen, N., Key, T.J., Day, N.E., Linseisen, J., Nagel, G., Bergmann, M.M., Overvad, K., Jensen, M.K., Tjonneland, A., Olsen, A., Bueno-de-Mesquita, H.B., Ocke, M., Peeters, P.H.M., Numans, M.E., Clavel-Chapelon, F., Boutron-Ruault, M.C., Trichopoulou, A., Psaltopoulou, T., Roukos, D., Lund, E., Hemon, B., Kaaks, R., Norat, T., Riboli, E., 2006. Meat intake and risk of stomach and esophageal adenocarcinoma within the European Prospective Investigation Into Cancer and Nutrition (EPIC). *Journal of the National Cancer Institute* 98(5), 345–354. doi:10.1093/jnci/djj071

Goodwin, C.S., Armstrong, J.A., Chilvers, T., Peters, M., Collins, M.D., SLY, L., McConnell, W., Harper, W.E.S., 1989. Transfer of *Campylobacter pylori* and *Campylobacter mustelae* to *Helicobacter* gen. nov. as *Helicobacter pylori* comb. nov. and *Helicobacter mustelae* comb. nov., respectively. *International Journal of Systematic Bacteriology* 39(4), 397–405. doi:10.1099/00207713-39-4-397

Goodwin, C.S., McCulloch, R.K., Armstrong, J.A., Wee, S.H., 1985. Unusual cellular fatty acids and distinctive ultrastructure in a new spiral bacterium (*Campylobacter pyloridis*) from the human gastric mucosa. *Journal of Medical Microbiology* 19(2), 257–267. doi:10.1099/00222615-19-2-257

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J.M., 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology* 57(1), 81–91. doi:10.1099/ijs.0.64483-0

Gravina, A.G., Zagari, R.M., De Musis, C., Romano, L., Loguercio, C., Romano, M., 2018. *Helicobacter pylori* and extragastric diseases: A review. *World Journal of Gastroenterology* 24(29), 3204–3221. doi:10.3748/wjg.v24.i29.3204

Gressmann, H., Linz, B., Ghai, R., Pleissner, K.-P., Schlapbach, R., Yamaoka, Y., Kraft, C., Suerbaum, S., Meyer, T.F., Achtman, M., 2005. Gain and loss of multiple genes during the evolution of *Helicobacter pylori*. *PLoS Genetics* 1(4), e43. doi:10.1371/journal.pgen.0010043

Gu, H., 2017. Role of flagella in the pathogenesis of *Helicobacter pylori*. *Current Microbiology* 74(7), 863–869. doi:10.1007/s00284-017-1256-4

Hashi, K., Imai, C., Yahara, K., Tahmina, K., Hayashi, T., Azuma, T., Miyabe-Nishiwaki, T., Sato, H., Matsuoka, M., Niimi, S., Okamoto, M., Hatakeyama, M., 2018. Evaluating the origin and virulence of a *Helicobacter pylori cagA*-positive strain isolated from a non-human primate. *Scientific Reports* 8(1), 1–13. doi:10.1038/s41598-018-34425-4

Hashinaga, M., Suzuki, R., Akada, J., Matsumoto, T., Kido, Y., Okimoto, T., Kodama, M., Murakami, K., Yamaoka, Y., 2016. Differences in amino acid frequency in CagA and VacA sequences of *Helicobacter pylori* distinguish gastric cancer from gastric MALT lymphoma. *Gut Pathogens* 8(1), 1–10. doi:10.1186/s13099-016-0137-x

Hatakeyama, M., 2017. Structure and function of *Helicobacter pylori* CagA, the first-identified bacterial protein involved in human cancer. *Proceedings of the Japan Academy, Series B* 93(4), 196–219. doi:10.2183/pjab.93.013

Hill, J.E., Paccagnella, A., Law, K., Melito, P.L., Woodward, D.L., Price, L., Leung, A.H., Ng, L.K., Hemmingsen, S.M., Goh, S.H., 2006. Identification of *Campylobacter* spp. and discrimination from *Helicobacter* and *Arcobacter* spp. by direct sequencing of PCR-amplified *cpn60* sequences and comparison to cpnDB, a chaperonin reference sequence database. *Journal of Medical Microbiology* 55(4), 393–399. doi:10.1099/jmm.0.46282-0

Hu, Y., Zhang, M., Lu, B., Dai, J., 2016. *Helicobacter pylori* and antibiotic resistance, a continuing and intractable problem. *Helicobacter* 21(5), 349–363. doi:10.1111/hel.12299

Huang, X., Deng, Z., Zhang, Q., Li, W., Wang, B., Li, M., 2016. Relationship between the *iceA* gene of *Helicobacter pylori* and clinical outcomes. *Therapeutics and Clinical Risk Management* 12, 1085–1092. doi:10.2147/TCRM.S107991

IARC, 1994. Schistosomes, liver flukes and *Helicobacter pylori*. *IARC Monographs on*

the Evaluation of Carcinogenic Risks to Humans 61(177), 1–241.

Israel, D.A., Peek, R.M., 2010. Surreptitious manipulation of the human host by *Helicobacter pylori*. *Gut Microbes* 1(2), 119–127. doi:10.4161/gmic.1.2.11991

Jones, K.R., Whitmire, J.M., Merrell, D.S., 2010. A tale of two toxins: *Helicobacter pylori* CagA and VacA modulate host pathways that impact disease. *Frontiers in Microbiology* 1(115), 1–17. doi:10.3389/fmicb.2010.00115

Junaid, M., Linn, A.K., Javadi, M.B., Al-Gubare, S., Ali, N., Katzenmeier, G., 2016. Vacuolating cytotoxin A (VacA) - A multi-talented pore-forming toxin from *Helicobacter pylori*. *Toxicon* 118, 27–35. doi:10.1016/j.toxicon.2016.04.037

Kabamba, E.T., Tuan, V.P., Yamaoka, Y., 2018. Genetic populations and virulence factors of *Helicobacter pylori*. *Infection, Genetics and Evolution* 60, 109–116. doi:10.1016/j.meegid.2018.02.022

Kelly, D.J., 1998. The physiology and metabolism of the human gastric pathogen *Helicobacter pylori*. In: Advances in Microbial Physiology. pp. 137–189. doi:10.1016/S0065-2911(08)60131-9

Kennemann, L., Didelot, X., Aebischer, T., Kuhn, S., Drescher, B., Droege, M., Reinhardt, R., Correa, P., Meyer, T.F., Josenhans, C., Falush, D., Suerbaum, S., 2011. *Helicobacter pylori* genome evolution during human infection. *Proceedings of the National Academy of Sciences of the United States of America* 108(12), 5033–5038. doi:10.1073/pnas.1018444108

Khalifa, M.M., Sharaf, R.R., Aziz, R.K., 2010. *Helicobacter pylori*: a poor man's gut pathogen? *Gut Pathogens* 2(2), 1–12. doi:10.1186/1757-4749-2-2

Kibria, K.M.K., Hossain, M.E., Sultana, J., Sarker, S.A., Bardhan, P.K., Rahman, M., Nahar, S., 2015. The prevalence of mixed *Helicobacter pylori* infections in symptomatic and asymptomatic subjects in Dhaka, Bangladesh. *Helicobacter* 20(5), 397–404. doi:10.1111/hel.12213

Kim, I.-J., Blanke, S.R., 2012. Remodeling the host environment: modulation of the gastric epithelium by the *Helicobacter pylori* vacuolating toxin (VacA). *Frontiers in*

Cellular and Infection Microbiology 2(37), 1–18. doi:10.3389/fcimb.2012.00037

Koch, R., 1882. Die Ätiologie der Tuberkulose. *Berliner Klinische Wochenschrift* 15, 221–230.

Kostrzynska, M., Betts, J.D., Austin, J.W., Trust, T.J., 1991. Identification, characterization, and spatial localization of two flagellin species in *Helicobacter pylori* flagella. *Journal of Bacteriology* 173(3), 937–946. doi:10.1128/jb.173.3.937-946.1991

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33(7), 1870–1874. doi:10.1093/molbev/msw054

Kusters, J.G., van Vliet, A.H.M., Kuipers, E.J., 2006. Pathogenesis of *Helicobacter pylori* infection. *Clinical Microbiology Reviews* 19(3), 449–90. doi:10.1128/CMR.00054-05

Kwon, D., El-zaatari, F.A.K., Kato, M., Osato, M.S., Reddy, R., Yamaoka, Y., Graham, D.Y., 2000. Analysis of *rdxA* and involvement of additional genes encoding NAD(P)H flavin oxidoreductase (FrxA) and ferredoxin-like protein (FdxB) in metronidazole resistance of *Helicobacter pylori*. *Antimicrobial Agents and Chemotherapy* 44(8), 2133–2142. doi:10.1128/aac.44.8.2133-2142.2000

Kyle, R.A., Steensma, D.P., Shampo, M.A., 2016. Barry James Marshall—Discovery of *Helicobacter pylori* as a cause of peptic ulcer. *Mayo Clinic Proceedings* 91(5), e67–e68. doi:10.1016/j.mayocp.2016.01.025

Labenz, J., Stolte, M., Bo, G., Blum, A.L., Bayerdorffer, E., Meining, A., Borsch, G., 1997. Curing *Helicobacter pylori* infection in patients with duodenal ulcer may provoke reflux esophagitis. *Gastroenterology* 112(5), 1442–1447. doi:10.1016/s0016-5085(97)70024-6

Lai, C.-H., Huang, J.-C., Chiang-Ni, C., Li, J.-P., Wu, L.-T., Wu, H.-S., Sun, Y.-C., Lin, M.-L., Lee, J.-F., Lin, H.-J., 2016. Mixed infections of *Helicobacter pylori* isolated from patients with gastrointestinal diseases in Taiwan. *Gastroenterology*

Research and Practice 2016, 1–5. doi:10.1155/2016/7521913

Langenberg, W., Rauws, E.A.J., Widjojokusumo, A., Tytgat, G.N., Zanen, H.C., 1986. Identification of *Campylobacter pyloridis* isolates by restriction endonuclease DNA analysis. *Journal of Clinical Microbiology* 24(3), 414–417. doi:10.1016/s0016-5085(97)70024-6

Lario, S., Ramírez-Lázaro, M.J., Aransay, A.M., Lozano, J.J., Montserrat, A., Casalots, Á., Junquera, F., Álvarez, J., Segura, F., Campo, R., Calvet, X., 2012. MicroRNA profiling in duodenal ulcer disease caused by *Helicobacter pylori* infection in a Western population. *Clinical Microbiology and Infection* 18(8), E273–282. doi:10.1111/j.1469-0691.2012.03849.x

Lario, S., Ramírez-Lázaro, M.J., Sanjuan-Herráez, D., Brunet-Vega, A., Pericay, C., Gombau, L., Junquera, F., Quintás, G., Calvet, X., 2017. Plasma sample based analysis of gastric cancer progression using targeted metabolomics. *Scientific Reports* 7(1), 17774. doi:10.1038/s41598-017-17921-x

Lauwers, G.Y., Riddell, R.H., 1999. Gastric epithelial dysplasia. *Gut* 45(5), 784–90. doi:10.1136/gut.45.5.784

Lawson, A.J., Elviss, N.C., Owen, R.J., 2005. Real-time PCR detection and frequency of 16S rDNA mutations associated with resistance and reduced susceptibility to tetracycline in *Helicobacter pylori* from England and Wales. *Journal of Antimicrobial Chemotherapy* 56, 282–286. doi:10.1093/jac/dki199

Lee, I., Kim, Y.O., Park, S.C., Chun, J., 2016. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *International Journal of Systematic and Evolutionary Microbiology* 66(2), 1100–1103. doi:10.1099/ijsem.0.000760

Lee, S.M., Kim, N., Kwon, Y.H., Nam, R.H., Kim, J.M., Park, J.Y., Lee, Y.S., Lee, D.H., 2018. *rdxA*, *frxA*, and efflux pump in metronidazole-resistant *Helicobacter pylori*: Their relation to clinical outcomes. *Journal of Gastroenterology and Hepatology* 33(3), 681–688. doi:10.1111/jgh.13906

Leja, M., Axon, A., Brenner, H., 2016. Epidemiology of *Helicobacter pylori* infection. *Helicobacter* 21(s1), 3–7. doi:10.1111/hel.12332

Leker, K., Lozano-Pope, I., Bandyopadhyay, K., Choudhury, B.P., Obonyo, M., 2017. Comparison of lipopolysaccharides composition of two different strains of *Helicobacter pylori*. *BMC Microbiology* 17(1), 226. doi:10.1186/s12866-017-1135-y

Lertsethtakarn, P., Ottemann, K.M., Hendrixson, D.R., 2011. Motility and chemotaxis in *Campylobacter* and *Helicobacter*. *Annual Review of Microbiology* 65(1), 389–410. doi:10.1146/annurev-micro-090110-102908

Li, H., Liao, T., Debowski, A.W., Tang, H., Nilsson, H.O., Stubbs, K.A., Marshall, B.J., Benghezal, M., 2016. Lipopolysaccharide structure and biosynthesis in *Helicobacter pylori*. *Helicobacter* 21(6), 445–461. doi:10.1111/hel.12301

Li, H., Tang, H., Debowski, A.W., Stubbs, K.A., Marshall, B.J., Benghezal, M., 2018. Lipopolysaccharide structural differences between western and asian *Helicobacter pylori* strains. *Toxins* 10(9), 1–10. doi:10.3390/toxins10090364

Li, H., Yang, T., Liao, T., Debowski, A.W., Nilsson, H.O., Fulurija, A., Haslam, S.M., Mulloy, B., Dell, A., Stubbs, K.A., Marshall, B.J., Benghezal, M., 2017. The redefinition of *Helicobacter pylori* lipopolysaccharide O-antigen and core-oligosaccharide domains. *PLoS Pathogens* 13(3), 1–21. doi:10.1371/journal.ppat.1006280

Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Roumagnac, P., Falush, D., Stamer, C., Prugnolle, F., Van Der Merwe, S.W., Yamaoka, Y., Graham, D.Y., Perez-Trallero, E., Wadstrom, T., Suerbaum, S., Achtman, M., 2007. An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* 445(7130), 915–918. doi:10.1038/nature05562

Linz, B., Windsor, H.M., Gajewski, J.P., Hake, C.M., Drautz, D.I., Schuster, S.C., Marshall, B.J., 2013. *Helicobacter pylori* genomic microevolution during naturally occurring transmission between adults. *PLoS ONE* 8(12), e82187.

doi:10.1371/journal.pone.0082187

Linz, B., Windsor, H.M., McGraw, J.J., Hansen, L.M., Gajewski, J.P., Tomsho, L.P., Hake, C.M., Solnick, J. V., Schuster, S.C., Marshall, B.J., 2014. A mutation burst during the acute phase of *Helicobacter pylori* infection in humans and rhesus macaques. *Nature Communications* 5(4165), 1–8. doi:10.1038/ncomms5165

Mahachai, V., Sirimontaporn, N., Tumwasorn, S., Thong-ngam, D., 2011. Sequential therapy in clarithromycin-sensitive and -resistant *Helicobacter pylori* based on polymerase chain reaction. *Journal of Gastroenterology and Hepatology* 26, 825–828. doi:10.1111/j.1440-1746.2011.06660.x

Mahdavi, J., Sondén, B., Hurtig, M., Olfad, F.O., Forsberg, L., Roche, N., Angström, J., Larsson, T., Teneberg, S., Karlsson, K.-A., Altraja, S., Wadström, T., Kersulyte, D., Berg, D.E., Dubois, A., Petersson, C., Magnusson, K.-E., Norberg, T., Lindh, F., Lundskog, B.B., Arnqvist, A., Hammarström, L., Borén, T., 2002. *Helicobacter pylori* SabA adhesin in persistent infection and chronic inflammation. *Science* 297(5581), 573–578. doi:10.1126/science.1069076

Maiden, M.C.J., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., Spratt, B.G., 1998. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America* 95(6), 3140–3145. doi:10.1073/pnas.95.6.3140

Malaty, H.M., Kumagai, T., Tanaka, E., Ota, H., Kiyosawa, K., Graham, D.Y., Katsuyama, T., 2000. Evidence from a nine-year birth cohort study in Japan of transmission pathways of *Helicobacter pylori* infection. *Journal of Clinical Microbiology* 38(5), 1971–1973.

Mansour, K. Ben, Fendri, C., Battikh, H., Garnier, M., Zribi, M., Jlizi, A., Burucoa, C., 2016. Multiple and mixed *Helicobacter pylori* infections: Comparison of two epidemiological situations in Tunisia and France. *Infection, Genetics and*

Evolution 37, 43–48. doi:10.1016/j.meegid.2015.10.028

Marshall, B.J., Warren, J.R., 1984. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *The Lancet* 323(8390), 1311–1315. doi:10.1016/S0140-6736(84)91816-6

Mascellino, M.T., Margani, M., Oliva, A., 2009. *Helicobacter pylori*: Determinant and markers of virulence. *Disease Markers* 27, 137–156. doi:10.3233/DMA-2009-0658

McGee, D.J., George, A.E., Trainor, E.A., Horton, K.E., Hildebrandt, E., Testerman, T.L., 2011. Cholesterol enhances *Helicobacter pylori* resistance to antibiotics and LL-37. *Antimicrobial Agents and Chemotherapy* 55(6), 2897–2904. doi:10.1128/AAC.00016-11

McInerney, J.O., McNally, A., O'Connell, M.J., 2017. Why prokaryotes have pangenomes. *Nature Microbiology* 2(4), 17040. doi:10.1038/nmicrobiol.2017.40

Megraud, F., Kist, M., López Brea, M., Hirschl, A., Andersen, L.P., Glupczynski, Y., Hirschl, A., 2011. Surveillance of *Helicobacter pylori* resistance to antibiotics in Europe 2008-2009. *Gastroenterology* 140(5), S-312. doi:10.1016/S0016-5085(11)61257-2

Meliț, L.E., Mărginean, C.O., Mărginean, C.D., Mărginean, M.O., 2019. The relationship between Toll-like receptors and *Helicobacter pylori* -related gastropathies: Still a controversial topic. *Journal of Immunology Research* 2019, 1–10. doi:10.1155/2019/8197048

Mendz, G.L., Hazell, S.L., 1995. Aminoacid utilization by *Helicobacter pylori*. *The International Journal of Biochemistry & Cell Biology* 27(10), 1085–1093. doi:10.1016/1357-2725(95)00069-2

Miñana-Galbis, D., Farfán, M., Lorén, J.G., Fusté, M.C., 2010. The reference strain *Aeromonas hydrophila* CIP 57.50 should be reclassified as *Aeromonas salmonicida* CIP 57.50. *International Journal of Systematic and Evolutionary Microbiology* 60(3), 715–717. doi:10.1099/ijs.0.017939-0

Mitchell, H., Katelaris, P., 2016. Epidemiology, clinical impacts and current clinical management of *Helicobacter pylori* infection. *Medical Journal of Australia* 204(10), 376–380. doi:10.5694/mja16.00104

Mobley, H., Mendz, G., Hazell, S., 2001. *Helicobacter pylori*: Physiology and genetics., ASM Press (Washington DC). Washington (DC).

Mohebi, N., Mamarabadi, M., Moghaddasi, M., 2013. Relation of *Helicobacter pylori* infection and multiple sclerosis in Iranian patients. *Neurology International* 5(2), 31–33. doi:10.4081/ni.2013.e10

Moodley, Y., Linz, B., Bond, R.P., Nieuwoudt, M., Soodyall, H., Schlebusch, C.M., Bernhöft, S., Hale, J., Suerbaum, S., Mugisha, L., van der Merwe, S.W., Achtman, M., 2012. Age of the association between *Helicobacter pylori* and man. *PLoS Pathogens* 8(5), e1002693. doi:10.1371/journal.ppat.1002693

Moodley, Y., Linz, B., Yamaoka, Y., Windsor, H.M., Breurec, S., Wu, J.Y., Maady, A., Bernhöft, S., Thiberge, J.M., Phuanukoonnon, S., Jobb, G., Siba, P., Graham, D.Y., Marshall, B.J., Achtman, M., Bernhoft, S., Thiberge, J.M., Phuanukoonnon, S., Jobb, G., Siba, P., Graham, D.Y., Marshall, B.J., Achtman, M., 2009. The peopling of the pacific from a bacterial perspective. *Science* 323(5913), 527–530. doi:10.1126/science.1166083

Morelli, G., Didelot, X., Kusecek, B., Schwarz, S., Bahlawane, C., Falush, D., Suerbaum, S., Achtman, M., 2010. Microevolution of *Helicobacter pylori* during prolonged infection of single hosts and within families. *PLoS Genetics* 6(7), 1–12. doi:10.1371/journal.pgen.1001036

Moss, S.F., 2017. The clinical evidence linking *Helicobacter pylori* to gastric cancer. *Cellular and Molecular Gastroenterology and Hepatology* 3(2), 183–191. doi:10.1016/j.jcmgh.2016.12.001

Ndip, R.N., MacKay, W.G., Farthing, M.J.G., Weaver, L.T., 2003. Culturing *Helicobacter pylori* from clinical specimens: Review of microbiologic methods. *Journal of Pediatric Gastroenterology and Nutrition* 36(5), 616–622. doi:10.1097/00005176-

200305000-00005

Nishizawa, T., Suzuki, H., Tsugawa, H., Muraoka, H., Matsuzaki, J., Hirata, K., Ikeda, F., Takahashi, M., Hibi, T., 2011. Enhancement of amoxicillin resistance after unsuccessful *Helicobacter pylori* eradication. *Antimicrobial Agents and Chemotherapy* 55(6), 3012–3014. doi:10.1128/AAC.00188-11

NobelPrize.org, 2005a. Press release: The Nobel Prize in Physiology or Medicine 2005 [WWW Document]. *Nobel Media AB*. URL https://www.nobelprize.org/prizes/medicine/2005/press-release/ (accessed 1.30.19).

NobelPrize.org, 2005b. Barry J. Marshall - Facts [WWW Document]. *Nobel Media AB*. URL https://www.nobelprize.org/prizes/medicine/2005/marshall/facts/ (accessed 1.30.19).

NobelPrize.org, 2005c. The Nobel Prize in Physiology or Medicine 2005 [WWW Document]. *Nobel Media AB*. URL http://nobelprize.org/medicine/laureates/2005/illpres/3_chronic.html (accessed 1.30.19).

Nomura, A., Stemmermann, G.N., Chyou, P.-H., Kato, I., Pérez-Pérez, G.I., Blaser, M.J., 1991. *Helicobacter pylori* infection and gastric carcinoma among japanese americans in Hawaii. *New England Journal of Medicine* 325(16), 1132–1136. doi:10.1056/NEJM199110173251604

Oliveira, J., Reygaert, W.C., 2019. Gram Negative Bacteria, StatPearls.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., Stevens, R., 2014. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Research* 42(D1), D206–D214. doi:10.1093/nar/gkt1226

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A., Parkhill, J., 2015. Roary: Rapid large-scale prokaryote

pan genome analysis. *Bioinformatics* 31(22), 3691–3693. doi:10.1093/bioinformatics/btv421

Palau, M., Kulmann, M., Ramírez-Lázaro, M.J., Lario, S., Quílez, M.E., Campo, R., Piqué, N., Calvet, X., Miñana-Galbis, D., 2016. Usefulness of housekeeping genes for the diagnosis of *Helicobacter pylori* infection, strain discrimination and detection of multiple infection. *Helicobacter* 21(6), 481–487. doi:10.1111/hel.12304

Palframan, S.L., Kwok, T., Gabriel, K., 2012. Vacuolating cytotoxin A (VacA), a key toxin for *Helicobacter pylori* pathogenesis. *Frontiers in Cellular and Infection Microbiology* 2(92), 1–9. doi:10.3389/fcimb.2012.00092

Parsonnet, J., Friedman, G.D., Vandersteen, D.P., Chang, Y., Vogelman, J.H., Orentreich, N., Sibley, R.K., 1991. *Helicobacter pylori* infection and the risk of gastric carcinoma. *New England Journal of Medicine* 325(16), 1127–1131. doi:10.1056/NEJM199110173251603

Patel, S., Mathivanan, N., Goyal, A., 2017. Bacterial adhesins, the pathogenic weapons to trick host defense arsenal. *Biomedicine & Pharmacotherapy* 93, 763–771. doi:10.1016/j.biopha.2017.06.102

Peek, R.M., Thompson, S.A., Atherton, J.C., Blaser, M.J., Miller, G.G., 1996. Expression of a novel ulcer-associated *H. pylori* gene, *iceA*, following adherence to gastric epithelial cells. *Gastroenterology* 110(Suppl), A225.

Pinto-Ribeiro, I., Ferreira, R., Batalha, S., Hlaing, T., Wong, S., Carneiro, F., Figueiredo, C., 2016. *Helicobacter pylori vacA* genotypes in chronic gastritis and gastric carcinoma patients from Macau, China. *Toxins* 8(142), 1–10. doi:10.3390/toxins8050142

Piqué, N., Palau, M., Berlanga, M., Miñana-Galbis, D., 2016. Advances in the research of new genetic markers for the detection of *Helicobacter pylori* infection. In: *Recent Advances in Pharmaceutical Sciences*. pp. 165–188.

Pormohammad, A., Ghotaslo, R., Leylabadlo, H.E., Nasiri, M.J., Dabiri, H., Hashemi, A.,

2018. Risk of gastric cancer in association with *Helicobacter pylori* different virulence factors: A systematic review and meta-analysis. *Microbial Pathogenesis* 118, 214–219. doi:10.1016/j.micpath.2018.03.004

Quaglia, N., Dambrosio, A., 2018. *Helicobacter pylori*: A foodborne pathogen? *World Journal of Gastroenterology* 24(31), 3472–3487. doi:10.3748/wjg.v24.i31.3472

Raei, N., Behrouz, B., Zahri, S., Latifi-Navid, S., 2016. *Helicobacter pylori* infection and dietary factors act synergistically to promote gastric cancer. *Asian Pacific Journal of Cancer Prevention* 17(3), 917–921. doi:10.7314/APJCP.2016.17.3.917

Ramírez-Lázaro, M.J., Lario, S., Casalots, A., Sanfeliu, E., Boix, L., García-Iglesias, P., Sánchez-Delgado, J., Montserrat, A., Bella-Cueto, M.R., Gallach, M., Sanfeliu, I., Segura, F., Calvet, X., 2011. Real-time PCR improves *Helicobacter pylori* detection in patients with peptic ulcer bleeding. *PLoS ONE* 6(5), e20009. doi:10.1371/journal.pone.0020009

Raymond, J., Thiberge, J.M., Chevalier, C., Kalach, N., Bergeret, M., Labigne, A., Dauga, C., 2004. Genetic and transmission analysis of *Helicobacter pylori* strains within a family. *Emerging Infectious Diseases* 10(10), 1816–1821. doi:10.3201/eid1010.040042

Redondo, J.J., Keller, P.M., Zbinden, R., Wagner, K., 2018. A novel RT-PCR for the detection of *Helicobacter pylori* and identification of clarithromycin resistance mediated by mutations in the 23S rRNA gene. *Diagnostic Microbiology and Infectious Disease* 90(1), 1–6. doi:10.1016/j.diagmicrobio.2017.09.014

Reshetnyak, V., Reshetnyak, T., 2017. Significance of dormant forms of *Helicobacter pylori* in ulcerogenesis. *World Journal of Gastroenterology* 23(27), 4867–4878. doi:10.3748/wjg.v23.i27.4867

Reynolds, D.J., Penn, C.W., 1994. Characteristics of *Helicobacter pylori* growth in a defined medium and determination of its amino acid requirements. *Microbiology* 140(10), 2649–2656. doi:10.1099/00221287-140-10-2649

Rhead, J.L., Letley, D.P., Mohammadi, M., Hussein, N., Mohagheghi, M.A., Eshagh

Hosseini, M., Atherton, J.C., 2007. A new *Helicobacter pylori* vacuolating cytotoxin determinant, the intermediate region, is associated with gastric cancer. *Gastroenterology* 133(3), 926–936. doi:10.1053/j.gastro.2007.06.056

Richter, J.E., Falk, G., Vaezi, M.F., 1998. *Helicobacter pylori* and gastroesophageal reflux disease: the bug may not be all bad. *American Journal of Gastroenterology* 93(10), 1800–1802. doi:10.1111/j.1572-0241.1998.00523.x

Rimbara, E., Noguchi, N., Kawai, T., Sasatsu, M., 2008. Mutations in penicillin-binding proteins 1, 2 and 3 are responsible for amoxicillin resistance in *Helicobacter pylori*. *Journal of Antimicrobial Chemotherapy* 61(5), 995–998. doi:10.1093/jac/dkn051

Roche, 2014. 454 Sequencing System Guidelines for Amplicon Experimental Design.

Rothenbacher, D., Blaser, M.J., Bode, G., Brenner, H., 2000. Inverse relationship between gastric colonization of *Helicobacter pylori* and diarrheal illnesses in children: Results of a population-based cross-sectional study. *The Journal of Infectious Diseases* 182(5), 1446–1449. doi:10.1086/315887

Rudi, J., Kolb, C., Maiwald, M., Kuck, D., Sieg, A., Galle, P.R., Stremmel, W., 1998. Diversity of *Helicobacter pylori vacA* and *cagA* genes and relationship to VacA and CagA protein expression, cytotoxin production, and associated disease. *Journal of Clinical Microbiology* 36(4), 944–948.

Sabbagh, P., Mohammadnia-Afrouzi, M., Javanian, M., Babazadeh, A., Koppolu, V., Vasigala, V.K.R., Nouri, H.R., Ebrahimpour, S., 2019. Diagnostic methods for *Helicobacter pylori* infection: ideals, options, and limitations. *European Journal of Clinical Microbiology and Infectious Diseases* 38(1), 55–66. doi:10.1007/s10096-018-3414-4

Salama, N.R., Hartung, M.L., Müller, A., 2013. Life in the human stomach: Persistence strategies of the bacterial pathogen *Helicobacter pylori*. *Nature Reviews Microbiology* 11(6), 385–399. doi:10.1038/nrmicro3016

Sánchez Delgado, J., García-Iglesias, P., Titó, L., Puig, I., Planella, M., Gené, E., Saló, J., Martínez-Cerezo, F., Molina-Infante, J., Gisbert, J.P., Calvet, X., 2018.

Actualización en el manejo de la infección por *Helicobacter pylori*. Documento de posicionamiento de la Societat Catalana de Digestologia. *Gastroenterología y Hepatología* 41(4), 272–280. doi:10.1016/j.gastrohep.2017.12.009

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F., 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* 75(23), 7537–7541. doi:10.1128/AEM.01541-09

Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27(6), 863–864. doi:10.1093/bioinformatics/btr026

Schulte, R.D., Makus, C., Schulenburg, H., 2013. Host-parasite coevolution favours parasite genetic diversity and horizontal gene transfer. *Journal of Evolutionary Biology* 26(8), 1836–1840. doi:10.1111/jeb.12174

Seemann, T., 2015. Snippy: Fast bacterial variant calling from NGS reads [WWW Document]. URL https://github.com/tseemann/snippy

Sheu, S.M., Sheu, B.S., Lu, C.C., Yang, H.B., Wu, J.J., 2009. Mixed infections of *Helicobacter pylori*: Tissue tropism and histological significance. *Clinical Microbiology and Infection* 15(3), 253–259. doi:10.1111/j.1469-0691.2008.02666.x

Stone, G.G., Shortridge, D.E.E., Versalovic, J., Beyer, J., Flamm, R.K., Graham, D.Y., Ghoneim, A.T., Tanaka, S.K.E.N., 1997. A PCR-Oligonucleotide ligation assay to determine the prevalence of 23S rRNA gene mutations in clarithromycin-resistant *Helicobacter pylori*. *Antimicrobial Agents and Chemotherapy* 41(3), 712–714. doi:10.1128/AAC.41.3.712

Stothard, P., Grant, J.R., Van Domselaar, G., 2017. Visualizing and comparing circular genomes using the CGView family of tools. *Briefings in Bioinformatics* bbx081. doi:10.1093/bib/bbx081

Su, Y.-L., Huang, H.-L., Huang, B.-S., Chen, P.-C., Chen, C.-S., Wang, H.-L., Lin, P.-H., Chieh, M.-S., Wu, J.-J., Yang, J.-C., Chow, L.-P., 2016. Combination of OipA, BabA, and SabA as candidate biomarkers for predicting *Helicobacter pylori*-related gastric cancer. *Scientific Reports* 6(1), 36442. doi:10.1038/srep36442

Suerbaum, S., Achtman, M., 2004. *Helicobacter pylori*: Recombination, population structure and human migrations. *International Journal of Medical Microbiology* 294(2–3), 133–139. doi:10.1016/j.ijmm.2004.06.014

Suerbaum, S., Michetti, P., 2002. *Helicobacter pylori* infection. *The New England Journal of Medicine* 347(15), 1175–1186. doi:10.1056/NEJMra020542

Suzuki, R., Shiota, S., Yamaoka, Y., 2012. Molecular epidemiology, population genetics, and pathogenic role of *Helicobacter pylori*. *Infection, Genetics and Evolution* 12(2), 203–213. doi:10.1016/j.meegid.2011.12.002

Sycuro, L.K., Pincus, Z., Gutierrez, K.D., Biboy, J., Stern, C.A., Vollmer, W., Salama, N.R., 2010. Peptidoglycan crosslinking relaxation promotes *Helicobacter pylori*'s helical shape and stomach colonization. *Cell* 141(5), 822–833. doi:10.1016/j.cell.2010.03.046

Sycuro, L.K., Wyckoff, T.J., Biboy, J., Born, P., Pincus, Z., Vollmer, W., Salama, N.R., 2012. Multiple peptidoglycan modification networks modulate *Helicobacter pylori*'s cell shape, motility, and colonization potential. *PLoS Pathogens* 8(3), e1002603. doi:10.1371/journal.ppat.1002603

Sze, M., Chen, Y., Tam, S., Tashkin, D., Wise, R., Connett, J., Man, S., Sin, D., 2015. The relationship between *Helicobacter pylori* seropositivity and COPD. *Thorax* 70, 923–929. doi:10.1136/thoraxjnl-2015-207059

Talarico, S., Gold, B.D., Fero, J., Thompson, D.T., Guarner, J., Czinn, S., Salama, N.R., 2009. Pediatric *Helicobacter pylori* isolates display distinct gene coding capacities and virulence gene marker profiles. *Journal of Clinical Microbiology* 47(6), 1680–1688. doi:10.1128/JCM.00273-09

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular

evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30(12), 2725–2729. doi:10.1093/molbev/mst197

Tatusova, T., Dicuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E.P., Zaslavsky, L., Lomsadze, A., Pruitt, K.D., Borodovsky, M., Ostell, J., 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research* 44(14), 6614–6624. doi:10.1093/nar/gkw569

Testerman, T., Conn, P., Mobley, H., McGee, D., 2006. Nutritional requirements and antibiotic resistance patterns of *Helicobacter* species in chemically defined media. *Journal of Clinical Microbiology* 44(5), 1650–1658. doi:10.1128/JCM.44.5.1650-1658.2006

Testerman, T., McGee, D., Mobley, H., 2001. *Helicobacter pylori* growth and urease detection in the chemically defined medium Ham's F-12 nutrient mixture. *Journal of Clinical Microbiology* 39(11), 3842–3850. doi:10.1128/JCM.39.11.3842-3850.2001

Teymournejad, O., Mobarez, A.M., Hassan, Z.M., Abadi, A.T.B., 2017. Binding of the *Helicobacter pylori* OipA causes apoptosis of host cells via modulation of Bax/Bcl-2 levels. *Scientific Reports* 7(1), 8036. doi:10.1038/s41598-017-08176-7

Tindall, B.J., Rosselló-Móra, R., Busse, H.J., Ludwig, W., Kämpfer, P., 2010. Notes on the characterization of prokaryote strains for taxonomic purposes. *International Journal of Systematic and Evolutionary Microbiology* 60(1), 249–266. doi:10.1099/ijs.0.016949-0

Toita, N., Yokota, S., Fujii, N., Konno, M., 2013. Clonality analysis of *Helicobacter pylori* in patients isolated from several biopsy specimens and gastric juice in a japanese urban population by random amplified polymorphic DNA fingerprinting. *Gastroenterology Research and Practice* 2013, 6. doi:10.1155/2013/721306

Trang, T.T.H., Binh, T.T., Yamaoka, Y., 2016. Relationship between *vacA* types and development of gastroduodenal diseases. *Toxins* 182(6), 10. doi:10.3390/toxins8060182

Trespalacios-Rangél, A.A., Otero, W., Arévalo-Galvis, A., Poutou-Piñales, R.A., Rimbara, E., Graham, D.Y., 2016. Surveillance of levofloxacin resistance in *Helicobacter pylori* isolates in Bogotá-Colombia (2009-2014). *PLoS ONE* 11(7), 1–10. doi:10.1371/journal.pone.0160007

Vandamme, P., Falsen, E., Rossau, R., Hoste, B., Segers, P., Tytgat, R., De Ley, J., 1991. Revision of *Campylobacter*, *Helicobacter*, and *Wolinella* taxonomy: Emendation of generic descriptions and proposal of *Arcobacter* gen. nov. *International Journal of Systematic Bacteriology* 41(1), 88–103. doi:10.1099/00207713-41-1-88

Villesen, P., 2007. FaBox: An online toolbox for FASTA sequences. *Molecular Ecology Notes* 7(6), 965–968. doi:10.1111/j.1471-8286.2007.01821.x

Wang, Y.K., Kuo, F.C., Liu, C.J., Wu, M.C., Shih, H.Y., Wang, S.S.W., Wu, J.Y., Kuo, C.H., Huang, Y.K., Wu, D.C., 2015. Diagnosis of *Helicobacter pylori* infection: Current options and developments. *World Journal of Gastroenterology* 21(40), 11221–11235. doi:10.3748/wjg.v21.i40.11221

Waskito, L.A., Salama, N.R., Yamaoka, Y., 2018. Pathogenesis of *Helicobacter pylori* infection CagA, gastric infection, motility, regulation of virulence. *Helicobacter* 23(s1), e12516. doi:10.1111/hel.12516

Weeks, D.L., Eskandari, S., Scott, D.R., Sachs, G., 2000. A H^+-gated urea channel: The link between *Helicobacter pylori* urease and gastric colonization. *Science* 287(5452), 482–485. doi:10.1126/science.287.5452.482

Weinert, L.A., Welch, J.J., 2017. Why might bacterial pathogens have small genomes? *Trends in Ecology & Evolution* 32(12), 936–947. doi:10.1016/j.tree.2017.09.006

Weng, M.-T., Chiu, Y.-T., Wei, P.-Y., Chiang, C.-W., Fang, H.-L., Wei, S.-C., 2019. Microbiota and gastrointestinal cancer. *Journal of the Formosan Medical Association* 118, S32–S41. doi:10.1016/j.jfma.2019.01.002

Worku, M.L., Sidebotham, R.L., Baron, J.H., Misiewicz, J.J., Logan, R.P.H., Keshavarz, T., Karim, Q.N., 1999. Motility of *Helicobacter pylori* in a viscous environment.

European Journal of Gastroenterology & Hepatology 11(10), 1143–1150. doi:10.1097/00042737-199910000-00012

World Health Organization, 2017. Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics, World Health Organization. doi:10.1590/S0100-15742013000100018

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., Zhang, G., Gu, Y.Q., Coleman-Derr, D., Xia, Q., Wang, Y., 2019. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research* 47(W1), W52–W58. doi:10.1093/nar/gkz333

Yamaoka, Y., 2009. *Helicobacter pylori* typing as a tool for tracking human migration. *Clinical Microbiology and Infection* 15(9), 829–834. doi:10.1111/j.1469-0691.2009.02967.x

Yamaoka, Y., Graham, D.Y., 2014. *Helicobacter pylori* virulence and cancer pathogenesis. *Future Oncology* 10(8), 1487–1500. doi:10.2217/fon.14.29

Yang, J.C., Lu, C.W., Lin, C.J., 2014. Treatment of *Helicobacter pylori* infection: Current status and future concepts. *World Journal of Gastroenterology* 20(18), 5283–5293. doi:10.3748/wjg.v20.i18.5283

You, Y., He, L., Zhang, M., Zhang, J., 2015. Comparative genomics of a *Helicobacter pylori* isolate from a chinese Yunnan Naxi ethnic aborigine suggests high genetic divergence and phage insertion. *PLoS ONE* 10(3), e0120659. doi:10.1371/journal.pone.0120659

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M., Larsen, M.V., 2012. Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy* 67(11), 2640–2644. doi:10.1093/jac/dks261

ANNEX I. Supplementary material to Chapter 3

H. pylori strain	Reference sequence number
51	CP000012
52	CP001680
83	CP002605
908	CP002184
2017	CP002571
2018	CP002572
26695	CP003904
26695	NC_000915
35A	CP002096
Aklavik117	CP003483
Aklavik86	CP003476
B38	NC_012973
B8	NC_014256
Cuz20	CP002076
ELS37	CP002953
F16	AP011940
F30	AP011941
F32	AP011943
F57	AP011945
G27	NC_011333
Gambia94/24	CP002332
HPAG1	NC_008086
HUP-B14	NC_017733
India7	CP002331
J99	NC_000921
Lithuania75	CP002334
P12	NC_011498
PeCan18	NC_017742
PeCan4	NC_014555
Puno120	CP002980
Puno135	CP002982
Rif1	CP003905
Rif2	CP003906
Sat464	CP002071

Shi112	NC_017741
Shi169	NC_017740
Shi417	NC_017739
Shi470	NC_010698
SJM180	NC_014560
SNT49	CP002983
SouthAfrica7	CP002336
v225d	CP001582
XZ274	CP003419

Table AI.1 *H. pylori* complete genome sequences used for primer design.

ANNEX II. Supplementary material to Chapter 4

H. pylori strain	*cgt* sequence	*luxS* sequence	Observations
7C	CP012905	CP012905	
29CaP	CP012907	CP012907	
35A	CP002096	CP002096	
51	CP000012	CP000012	
52	CP001680	CP001680	
83	CP002605	CP002605	
908	CP002184	CP002184	*cgt* and *luxS* sequences identical to these from the strain 2017
2017	CP002571	CP002571	
2018	CP002572	CP002572	*cgt* and *luxS* sequences identical to these from the strain 2017
26695	CP003904	CP003904	
26695 (2)	NC_000915	NC_000915	*cgt* and *luxS* sequences identical to these from the strain 26695
26695-1	AP013354	AP013354	*cgt* and *luxS* sequences identical to these from the strain 26695
26695-1MET	CP010436	CP010436	*cgt* and *luxS* sequences identical to these from the strain 26695
Aklavik86	CP003476	CP003476	
Aklavik117	CP003483	CP003483	
APP134	KU053359	KU053431	
ATCC 49503	KU053360	KU053432	
ATCC 51932	KU053361	KU053433	
B8	NC_014256	NC_014256	
B38	NC_012973	NC_012973	
B247	KU053362	KU053434	
B271	KU053363	KU053435	
B319	KU053364	KU053436	
B355	KU053365	KU053437	
B373	MG950173	MG950172	
B491	KU053366	KU053438	
B508S	KU053367	KU053439	
B657A-1	KU053370	KU053442	
B657C	KU053372	KU053444	
B659A	KU053373	KU053445	
BM012A	CP006888	CP006888	

BM012B	CP007605	CP007605	*luxS* sequence identical to that from the strain BM012A
BM012S	CP006889	CP006889	*cgt* and *luxS* sequences identical to these from the strain BM012A
BM013A	CP007604	CP007604	
BM013B	CP007606	CP007606	*cgt* and *luxS* sequences identical to these from the strain BM013A
CC33C	CP011484	CP011484	
CRL122	KU053375	KU053447	
Cuz20	CP002076	CP002076	
ELS37	CP002953	CP002953	
F16	AP011940	AP011940	
F30	AP011941	AP011941	
F32	AP011943	AP011943	
F57	AP011945	AP011945	
G27	NC_011333	NC_011333	
Gambia94/24	CP002332	CP002332	
Hp238	CP010013	CP010013	
HPAG1	NC_008086	NC_008086	
HUP-B14	NC_017733	NC_017733	
India7	CP002331	CP002331	
J99	NC_000921	NC_000921	
J166	CP007603	CP007603	*cgt* sequence identical to that from the strain B8
L7	CP011482	CP011482	
Lithuania75	CP002334	CP002334	
ML1	AP014710	AP014710	
ML2	AP014711	AP014711	
ML3	AP014712	AP014712	
OK113	AP012600	AP012600	*cgt* sequence identical to that from the strain F32
OK310	AP012601	AP012601	
oki102	CP006820	CP006820	
oki112	CP006821	CP006821	*cgt* sequence identical to that from the strain oki102
oki128	CP006822	CP006822	
oki154	CP006823	CP006823	
oki422	CP006824	CP006824	*luxS* sequence identical to that from the strain oki112
oki673	CP006825	CP006825	*cgt* and *luxS* sequences identical to these from the strain oki128
oki828	CP006826	CP006826	*luxS* sequence identical to that from the strain oki128

oki898	CP006827	CP006827	*cgt* and *luxS* sequences identical to these from the strain oki112
P12	NC_011498	NC_011498	
PeCan4	NC_014555	NC_014555	
PeCan18	NC_017742	NC_017742	
Puno120	CP002980	CP002980	
Puno135	CP002982	CP002982	*cgt* sequence identical to that from the strain F32
Rif1	CP003905	CP003905	*cgt* and *luxS* sequences identical to these from the strain 26695
Rif2	CP003906	CP003906	*cgt* and *luxS* sequences identical to these from the strain 26695
Sat464	CP002071	CP002071	
Shi112	NC_017741	NC_017741	*cgt* sequence identical to that from the strain Cuz20
Shi169	NC_017740	NC_017740	*cgt* sequence identical to that from the strain Sat464
Shi417	NC_017739	NC_017739	*cgt* sequence identical to that from the strain Sat464
Shi470	NC_010698	NC_010698	
SJM180	NC_014560	NC_014560	
SNT49	CP002983	CP002983	
SouthAfrica7	CP002336	CP002336	
SouthAfrica20	CP006691	CP006691	
SVC135	KU053376	KU053448	
UM032	CP005490	CP005490	*cgt* sequence identical to that from the strain ML2
UM037	CP005492	CP005492	
UM066	CP005493	CP005493	*cgt* sequence identical to that from the strain ML3
UM298	CP006610	CP006610	*cgt* and *luxS* sequences identical to these from the strain UM032
UM299	CP005491	CP005491	*cgt* and *luxS* sequences identical to these from the strain UM032
v225d	CP001582	CP001582	
XZ274	CP003419	CP003419	

Table AII.1 Reference sequences used in this study.

ANNEX III. Supplementary material to Chapter 5

B508A-S1 s2m2i2 scaffold_15 (s= 286bp, m= 642bp, i= 432bp)

1 TGCTAAATTT AGCGTTTTGT GCGCTTAATT CTAAATTTTT ACTAAAATCG CTTATTGAGT
61 GGCTTCGCTC TAACGCTCTT TTTAGGTGAT ACTTTAAAGC CGCGCCTGCG GTGGCTTCGT
121 TTAAAGCGTA GCCGCTAAAC TAAAACTTTT TTTAAGAGCG CAGTTGTTGT AAGGCTTCTT
...
5641 ACTTTTTAAT CCTATGGGTT TTATATTCAT TTATCTTAAC TTAATAAAAA TTGAACATTG
5701 GTTGTAGATA CTGCATATTT ATAACCTTAA TCGTAAATGC AACAGAAATT TTCTAGTCTA
5761 AAGTCGCACC CTTTGTGCAA AAATCGTTTT ACAAAAAGAA AGGAAAAAAA TGGAAATACA
5821 ACAAACACAC CGCAAAATCA ATCGCCCCAT TATCTCTCTC GCTCTAGTGG GGGTGTTAAT
5881 GGGCACCGAA CTAGGGGCTA ATACGCCAAA TGATCCCATA CACAGCGAGA GTCGCGCCTT
5941 TTTTACAACC GTGATCATTC CAGCCATTGT TGGGGGTATC GCTACAGGCG CTGCTGTAGG
6001 AACGGTCTCA GGGCTTCTTA GCTGGGGGCT CAAACAAGCC GAACAAGCCA ATAAAGCCCC
6061 AGACAAACCC GATAAAGTTT GGCGCATTCA AGCAGGAAGA GGCTTTGATA ATTTCCCCCA
6121 CAAGCAATAC GACTTATACA AATCCCTACT ATCTAGCAAA ATTGATGGAG GTTGGGA<u>C</u>TG
6181 GGGGAATGCC GCTAGGCATT ATTGGGTCAA AGACGGGCAG TGGAACAAAC TTGAAGTGGA
6241 TATGCAAAAC GCTGTAGGGA CTTATAACCT TTCAGGCCTT ATCAACTTTA CTGGTGGGGA
6301 TTTGGACATC AATATGCAAA AAGCCACTTT GCGTTTGGGC CAATTCAATG GCAATTCTTT
6361 CACAAGCTAT AAGGATAGCG CCGATCGCAC CACGAGGGTG AATTTTGACG CTAAAAATAT
6421 CTTAATTGAT AATTTTGTAG AAATCAACAA TCGTGTGGGT TCTGGAGCCG GAAGAAAAGC
6481 CAGCTCTACG GTTTTGACCT TGAAAAGTTC AGAAAAAATT ACAAGCCGTG AAAACGCGGA
6541 AATCTCTCTT TATGATGGCG CCACGCTCAA TTTGGTTTCA AGCTCAAATC AGAGCGTTGA
6601 TCTATATGGG AAAGTGTGGA TGGGCCGTTT GCAATACGTG GGAGCGTATT TAGCCCCTTC
6661 ATACAGCACA ATAGACACTT CAAAAGTGCA AGGGGAAATG AATTTTCGCC ATCTCGCTGT
6721 GGGTGATCAA AACGCCGCTC AAGCGGGCAT TATCGCTAAT AAAAAGACTA ATATTGGCAC
6781 ACTGGATTTG TGGCAAAGTG CGGGGTTAAG CATCATCACC CCTCCCGAAG GCGGTTATGA
6841 GAGTAAAACT AAAGATAACC CTTCTCAAAA CAACCCTAAA AATGACACGC AAAAAACAGA
6901 AATTCAACCC ACGCAAGTCA TTGACGGGCC TTTTGCAGGC GGCAAAGACA CGGTTGTGAA
6961 TATTTTCCAC TTAAACACTA AAGCCGATGG CACGCTTAAA GCGGGAGGGT TTAAAGCTTC
7021 TCTTAGCACG AATGCGGCTC ATTTGAATAT CGGCGAAGGC GGTGTCAATC TGTCCAATCA
7081 AGCGAGCGGG CGCTCTCTTT TAGTGGAAAA CCTAACCGGG AATATCACCG TTGAGGGGAC
7141 TTTAAGAGTG AATAATCAAG TGGGCGGTGC TGCTGTGGCA GGCTCAAGCG CGAATTTTGA
7201 GTTTAAGGCT GGTGAAGACA CCAACAACGC CACAGCCACT TTTAATAACG ATATCCATCT
7261 AGGAAAAGCG GTGAATTTAA GAGTGGATGC TCATACAGCT TATTTTAATG GCAATATTTA
7321 TCTGGGAAAA TCCACGAATT TAAGAGTGAA TGGCCATAGC GCTCATTTTA AAAATATTGA
7381 TGCCAGTAAG AGCGATAACG GGCTAAACAC TAGCGCTTTG GATTTTAGCG GCGTTACAGA
7441 CAAAGTCAAT ATCAACAAGC TCACTACATC TGCCACTAAT GTGAACGTTA AAAACTTTGA
7501 CATTAAGGAA TTGGTGGTTA CAACCCGAGT TCAAAGTTTT GGGCAATACA CTATTTTTGG
7561 CGAAAATATA GGCGATAAGT CTCGCATTGG TGTCGTGAGT TTGCAAACGG GATATAGCCC
7621 GGCTTATTCT GGGGGCGTTA CTTTTAAAAG CGGTAAGAAA CTCGTTATAG ATGAAATTTA
7681 CCATGCCCCT TGGAATTATT TTGACGCTAG GAATGTTACC GATGTTGAAA TCAACAAAAG

B508A-T2A s1m2i2 scaffold_17 (s = 259 bp, m= 642bp, i = 432bp)

1 TGCTAAATTT AGCGTTTTGT GCGCTTAATT CTAAATTTTT ACTAAAATCG CTTATTGAGT
61 GGCTTCGCTC TAACGCTCTT TTTAGGTGAT ACTTTAAAGC CGCGCCTGCG GTGGCTTCGT
121 TTAAAGCGTA GCCGCTAAAC TAAAACTTTT TTTAAGAGCG CAGTTGTTGT AAGGCTTCTT
...
5461 TTTACAACAA AAAATCGCTT TGATGGACAC CCCACAAGGC ACGATTTGGG AGAAGCTTTT
5521 TTAAACGCCT CCAATTTTAC CTTTTTACAC GCTCTAGCCA CAAATTCTAG CAATATTGCT
5581 TTTTAATCTT GTTGAGTTTT ATGTTCATTT ACCTTAATTT GATAAAAATT GAATATTGGT
5641 TGTAGATACT ATATATTTAT AGCCTTAATC GTAAATGCAA CAGAAATTTT CTAGTCTAAA
5701 GTCGTACCCT TTGTGCAAAA ATCGTTTTAC AAGAAAAGAA GAAAGGAAAG AAATGGAAAT
5761 ACAACAAACA CACCGCAAAA TGAATCGCCC TTTAGTTTCT CTTGTTTTAG CAGGAGCGTT
5821 GATTAGCTCC ATACCGCAAG AGAGTCATGC TGCCTTTTTC ACAACCGTGA TCATTCCAGC
5881 CATTGTTGGG GGTATCGCTA CAGGCACCGC TGTAGGAACG GTATCAGGGC TTCTTAGCTG
5941 GGGACTCAAA CAAGCCGAAG AAGCGAATAA AACCCCAGAT AAACCCGATA AAGTTTGGCG
6001 CATTCAAGCA GGAAGAGGCT TTGATAATTT CCCCCACAAG CAATACGACT TATACAAATC
6061 CCTACTATCT AGCAAAATTG ATGGAGGTTG GGA**C**TGGGGG AATGCCGCTA GGCATTATTG
6121 GGTCAAAGAC GGGCAGTGGA ACAAACTTGA AGTGGATATG CAAAACGCTG TAGGGACTTA
6181 TAACCTTTCA GGCCTTATCA ACTTTACTGG TGGGGATTTG GACATCAATA TGCAAAAAGC
6241 CACTTTGCGT TTGGGCCAAT TCAATGGCAA TTCTTTCACA AGCTATAAGG ATAGCGCCGA
6301 TCGCACCACG AGGGTGAATT TTGACGCTAA AAATATCTTA ATTGATAATT TTGTAGAAAT
6361 CAACAATCGT GTGGGTTCTG GAGCCGGAAG AAAAGCCAGC TCTACGGTTT TGACCTTGAA
6421 AAGTTCAGAA AAAATTACAA GCCGTGAAAA CGCGGAAATC TCTCTTTATG ATGGCGCCAC
6481 GCTCAATTTG GTTTCAAGCT CAAATCAGAG CGTTGATCTA TATGGGAAAG TGTGGATGGG
6541 CCGTTTGCAA TACGTGGGAG CGTATTTAGC CCCTTCATAC AGCACAATAG ACACTTCAAA
6601 AGTGCAAGGG GAAATGAATT TTCGCCATCT CGCTGTGGGT GATCAAAACG CCGCTCAAGC
6661 GGGCATTATC GCTAATAAAA AGACTAATAT TGGCACACTG GATTTGTGGC AAAGTGCGGG
6721 GTTAAGCATC ATCACCCCTC CCGAAGGCGG TTATGAGAGT AAAACTAAAG ATAACCCTTC
6781 TCAAAACAAC CCTAAAAATG ACACGCAAAA AACAGAAATT CAACCCACGC AAGTCATTGA
6841 CGGGCCTTTT GCAGGCGGCA AAGACACGGT TGTGAATATT TTCCACTTAA ACACTAAAGC
6901 CGATGGCACG CTTAAAGCGG GAGGGTTTAA AGCTTCTCTT AGCACGAATG CGGCTCATTT
6961 GAATATCGGC GAAGGCGGTG TCAATCTGTC CAATCAAGCG AGCGGGCGCT CTCTTTTAGT
7021 GGAAAACCTA ACCGGGAATA TCACCGTTGA GGGGACTTTA AGAGTGAATA ATCAAGTGGG
7081 CGGTGCTGCT GTGGCAGGCT CAAGCGCGAA TTTTGAGTTT AAGGCTGGTG AAGACACCAA
7141 CAACGCCACA GCCACTTTTA ATAACGATAT CCATCTAGGA AAAGCGGTGA ATTTAAGAGT
7201 GGATGCTCAT ACAGCTTATT TTAATGGCAA TATTTATCTG GGAAAATCCA CGAATTTAAG
7261 AGTGAATGGC CATAGCGCTC ATTTTAAAAA TATTGATGCC AGTAAGAGCG ATAACGGGCT
7321 AAACACTAGC GCTTTGGATT TTAGCGGCGT TACAGACAAA GTCAATATCA ACAAGCTCAC
7381 TACATCTGCC ACTAATGTGA ACGTTAAAAA CTTTGACATT AAGGAATTGG TGGTTACAAC
7441 CCGAGTTCAA AGTTTTGGGC AATACACTAT TTTTGGCGAA AATATAGGCG ATAAGTCTCG
7501 CATTGGTGTC GTGAGTTTGC AAACGGGATA TAGCCCGGCT TATTCTGGGG GCGTTACTTT
7561 TAAAAGCGGT AAGAAACTCG TTATAGATGA AATTTACCAT GCCCCTTGGA ATTATTTTGA
7621 CGCTAGGAAT GTTACCGATG TTGAAATCAA CAAAAGAATT CTTTTTGGAG CCCCAGGAAA

B508A-T4 s2m2i2 scaffold_15 (s= 286 bp, m= 642 bp, i= 432 bp)

```
   1 TTTTCGGCGT TGGCTTGTTC TAATGGGCTT AAATTGGTAG CGTTTTGTGG GGTTTCGTTT
  61 AAATTTTCTG TATTTTTTAA TAAATCCTCT TGATCTATTA AATCTTTTTG TGTAGTATTC
 121 TCTGTAGCAC CGCTATAGGA ACTCAACCTA TGGTTCACCT TTGCATCAGC ATTGAGTGTA
 ...
6481 TTAACTTAAT AAAAATTGAA CATTGGTTGT AGATACTGCA TATTTATAAC CTTAATCGTA
6541 AATGCAACAG AAATTTTCTA GTCTAAAGTC GCACCCTTTG TGCAAAAATC GTTTTACAAA
6601 AAGAAAGGAA AAAAATGGAA ATACAACAAA CACACCGCAA AATCAATCGC CCCATTATCT
6661 CTCTCGCTCT AGTGGGGGTG TTAATGGGCA CCGAACTAGG GGCTAATACG CCAAATGATC
6721 CCATACACAG CGAGAGTCGC GCCTTTTTTA CAACCGTGAT CATTCCAGCC ATTGTTGGGG
6781 GTATCGCTAC AGGCGCTGCT GTAGGAACGG TCTCAGGGCT TCTTAGCTGG GGGCTCAAAC
6841 AAGCCGAACA AGCCAATAAA GCCCCAGACA AACCCGATAA AGTTTGGCGC ATTCAAGCAG
6901 GAAGAGGCTT TGATAATTTC CCCCACAAGC AATACGACTT ATACAAATCC CTACTATCTA
6961 GCAAAATTGA TGGAGGTTGG GACTGGGGGA ATGCCGCTAG GCATTATTGG GTCAAAGACG
7021 GGCAGTGGAA CAAACTTGAA GTGGATATGC AAAACGCTGT AGGGACTTAT AACCTTTCAG
7081 GCCTTATCAA CTTTACTGGT GGGGATTTGG ACATCAATAT GCAAAAAGCC ACTTTGCGTT
7141 TGGGCCAATT CAATGGCAAT TCTTTCACAA GCTATAAGGA TAGCGCCGAT CGCACCACGA
7201 GGGTGAATTT TGACGCTAAA AATATCTTAA TTGATAATTT TGTAGAAATC AACAATCGTG
7261 TGGGTTCTGG AGCCGGAAGA AAAGCCAGCT CTACGGTTTT GACCTTGAAA AGTTCAGAAA
7321 AAATTACAAG CCGTGAAAAC GCGGAAATCT CTCTTTATGA TGGCGCCACG CTCAATTTGG
7381 TTTCAAGCTC AAATCAGAGC GTTGATCTAT ATGGGAAAGT GTGGATGGGC CGTTTGCAAT
7441 ACGTGGGAGC GTATTTAGCC CCTTCATACA GCACAATAGA CACTTCAAAA GTGCAAGGGG
7501 AAATGAATTT TCGCCATCTC GCTGTGGGTG ATCAAAACGC CGCTCAAGCG GGCATTATCG
7561 CTAATAAAAA GACTAATATT GGCACACTGG ATTTGTGGCA AAGTGCGGGG TTAAGCATCA
7621 TCACCCCTCC CGAAGGCGGT TATGAGAGTA AAACTAAAGA TAACCCTTCT CAAAACAACC
7681 CTAAAAATGA CACGCAAAAA ACAGAAATTC AACCCACGCA AGTCATTGAC GGGCCTTTTG
7741 CAGGCGGCAA AGACACGGTT GTGAATATTT TCCACTTAAA CACTAAAGCC GATGGCACGC
7801 TTAAAGCGGG AGGGTTTAAA GCTTCTCTTA GCACGAATGC GGCTCATTTG AATATCGGCG
7861 AAGGCGGTGT CAATCTGTCC AATCAAGCGA GCGGGCGCTC TCTTTTAGTG GAAAACCTAA
7921 CCGGGAATAT CACCGTTGAG GGGACTTTAA GAGTGAATAA TCAAGTGGGC GGTGCTGCTG
7981 TGGCAGGCTC AAGCGCGAAT TTTGAGTTTA AGGCTGGTGA AGACACCAAC AACGCCACAG
8041 CCACTTTTAA TAACGATATC CATCTAGGAA AAGCGGTGAA TTTAAGAGTG GATGCTCATA
8101 CAGCTTATTT TAATGGCAAT ATTTATCTGG GAAAATCCAC GAATTTAAGA GTGAATGGCC
8161 ATAGCGCTCA TTTTAAAAAT ATTGATGCCA GTAAGAGCGA TAACGGGCTA AACACTAGCG
8221 CTTTGGATTT TAGCGGCGTT ACAGACAAAG TCAATATCAA CAAGCTCACT ACATCTGCCA
8281 CTAATGTGAA CGTTAAAAAC TTTGACATTA AGGAATTGGT GGTTACAACC CGAGTTCAAA
8341 GTTTTGGGCA ATACACTATT TTTGGCGAAA ATATAGGCGA TAAGTCTCGC ATTGGTGTCG
8401 TGAGTTTGCA AACGGGATAT AGCCCGGCTT ATTCTGGGGG CGTTACTTTT AAAAGCGGTA
8461 AGAAACTCGT TATAGATGAA ATTTACCATG CCCCTTGGAA TTATTTTGAC GCTAGGAATG
```

Figure AIII.1. *vacA* genomic sequences extracted from the whole genomes of strains B508A-S1, B508A-T2A and B508A-T4. Primers used for the detection of *s, i* and *m* regions are highlighted.

clusterID	Number of proteins	Swiss-Prot Hit
cluster 6	3	DUF874 family protein [BLAST]
cluster1409	2	hypothetical protein

(a)

clusterID	Number of proteins	Swiss-Prot Hit
cluster3	4	DUF874 family protein [BLAST]
cluster4	4	Outer membrane beta-barrel protein
cluster1405	2	Glycosiltransferase family 8 protein
cluster1406	2	hypothetical protein
cluster1407	2	Glycosiltransferase family 8 protein
cluster1408	2	DUF3519 domain-containing protein
cluster1410	2	DUF3519 domain-containing protein
cluster1411	2	Outer membrane protein family
cluster1412	2	Acetyl-coenzyme A synthetase
cluster1413	2	Alpha-1,2-fucosultransferase
cluster1414	2	Flagellar biosynthesis protein FlhF
cluster1415	2	Modification methylase HinfI

(b)

clusterID	Number of proteins	Swiss-Prot Hit
cluster5	4	DUF874 family protein [BLAST]
cluster1416	2	Modification methylase VspI
cluster1417	2	hypothetical protein
cluster1418	2	DUF697 domain-containing protein
cluster1419	2	Uncharacterized chromosomal cassette SCCmec type IVc protein CR006

cluster1420	2	hypothetical protein
cluster1421	2	outer membrane protein OipA
cluster1422	2	LPS heptosyltransferase family protein
cluster1423	2	Lipooligosaccharide biosynthesis protein lic2B
cluster1424	2	hypothetical protein
cluster1425	2	Flagellar biosynthesis protein FlhA
cluster1426	2	Type III restriction-modification system
cluster1427	2	hypothetical protein
cluster1428	2	ATP-binding protein
cluster1429	2	SAM-dependent methyltransferase
cluster1430	2	Uncharacterized protein HI_1472
cluster1431	2	Inner membrane ABC transporter permease protein YejE
cluster1432	2	Proline/betaine transporter
cluster1433	2	ATP-binding protein
cluster1434	2	cysteine desulfurase
cluster1435	2	Trigger factor

(c)

Table AIII.1. Shared clusters between strains found using OrthoVenn2. (a) Strains B508A-T2A and B508A-T4. (b) Strains B508A-S1 and B508A-T2A. (c). Strains B508A-S1 and B508A-T4.

*vacA*_4	Vacuolating cytotoxin autotransporter (*vacA*_4)
flhA	Flagellar biosynthesis protein FlhA (flhA_1)
fucT_1	Alpha-(1,3)-fucosyltransferase FucT (fucT_1)
fucT_2	Alpha-(1,3)-fucosyltransferase FucT (fucT_2)
type-1 restriction enzyme	Putative type-1 restriction enzyme specificity protein MG438
hsdM	Type I restriction enzyme EcoKI M protein (hsdM_2)
rocC	Amino-acid permease RocC (rocC_1)
xerH	Tyrosine recombinase XerH (xerH_3)
hydB	Quinone-reactive Ni/Fe-hydrogenase large chain (hydB)
era	GTPase Era (era_3)
cpdA	3',5'-cyclic adenosine monophosphate phosphodiesterase CpdA
era	GTPase Era (era_2)
der	GTPase Der (der_3)
msrAB	Peptide methionine sulfoxide reductase msrA/msrB (msrAB)
yejE	Inner membrane ABC transporter permease protein YejE (yejE_1)

Table AIII.2 Description of the unique genes found with Roary.

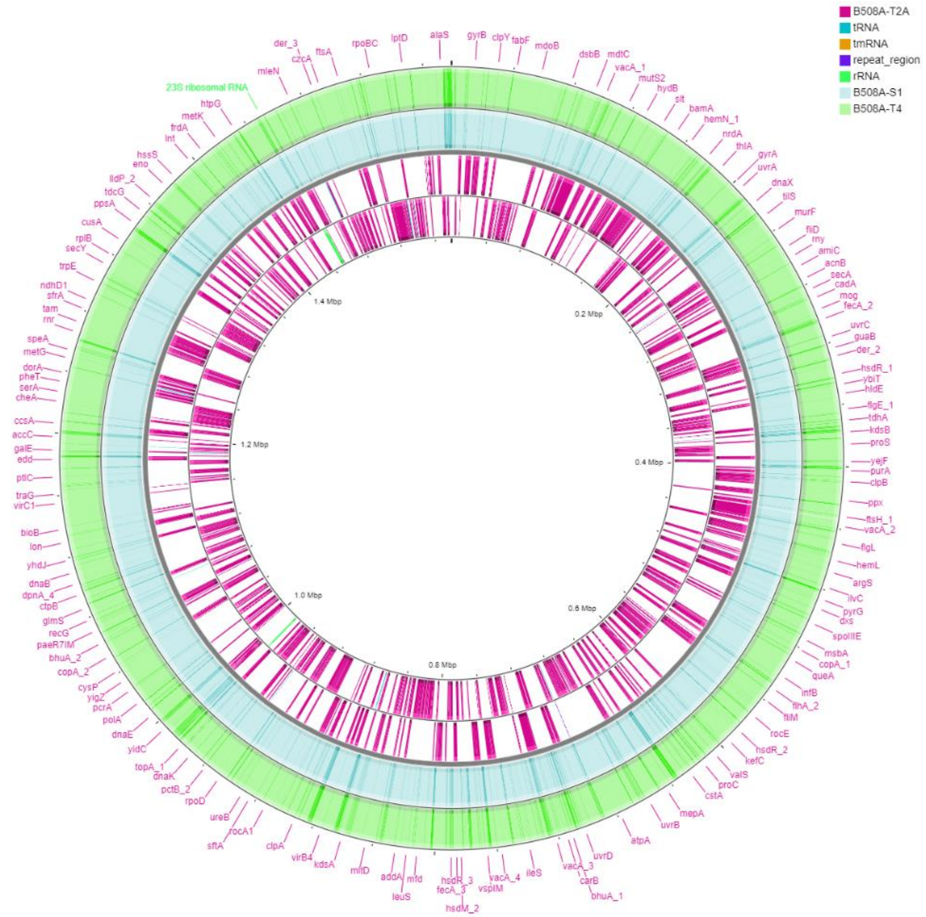
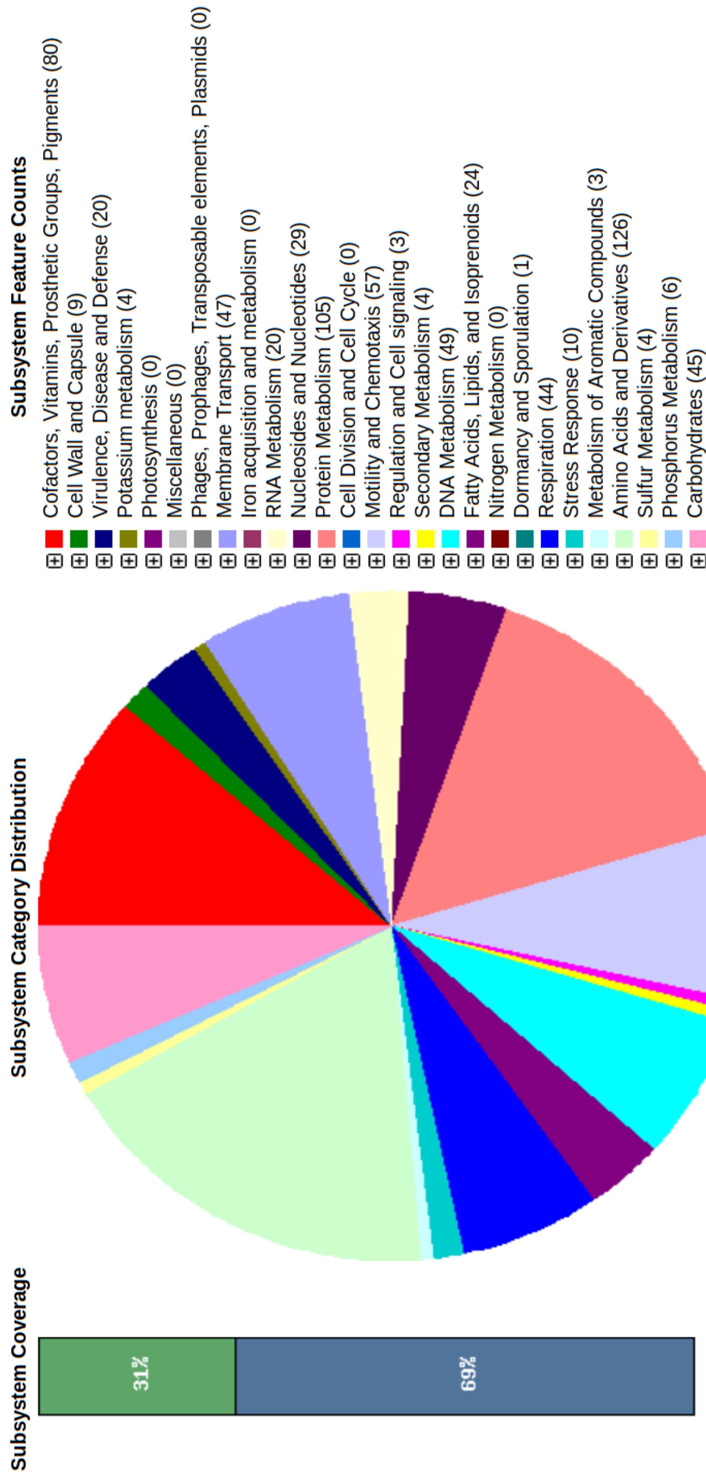
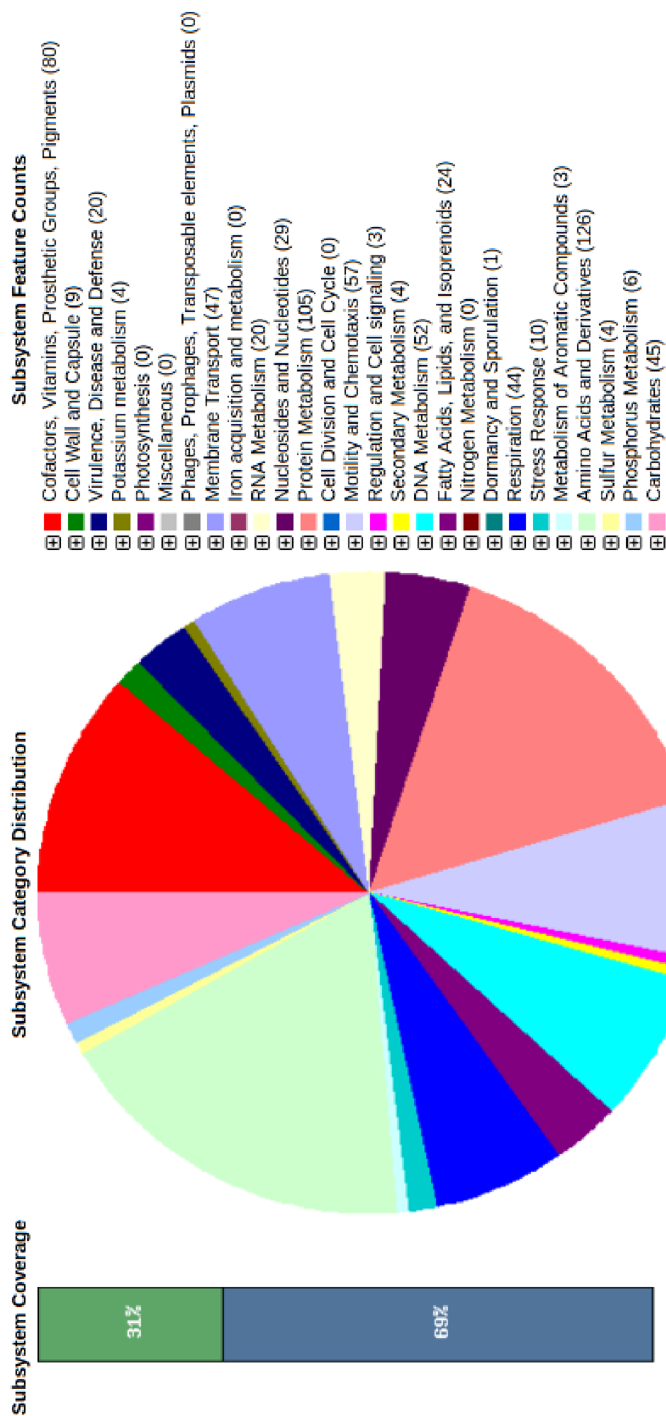


Figure AIII.2 Circular comparative representation of the whole genomes using CGView. Inner circles represent the forward and reverse sequences of B508A-T2A.

(**a**)



(**b**)



(**c**)

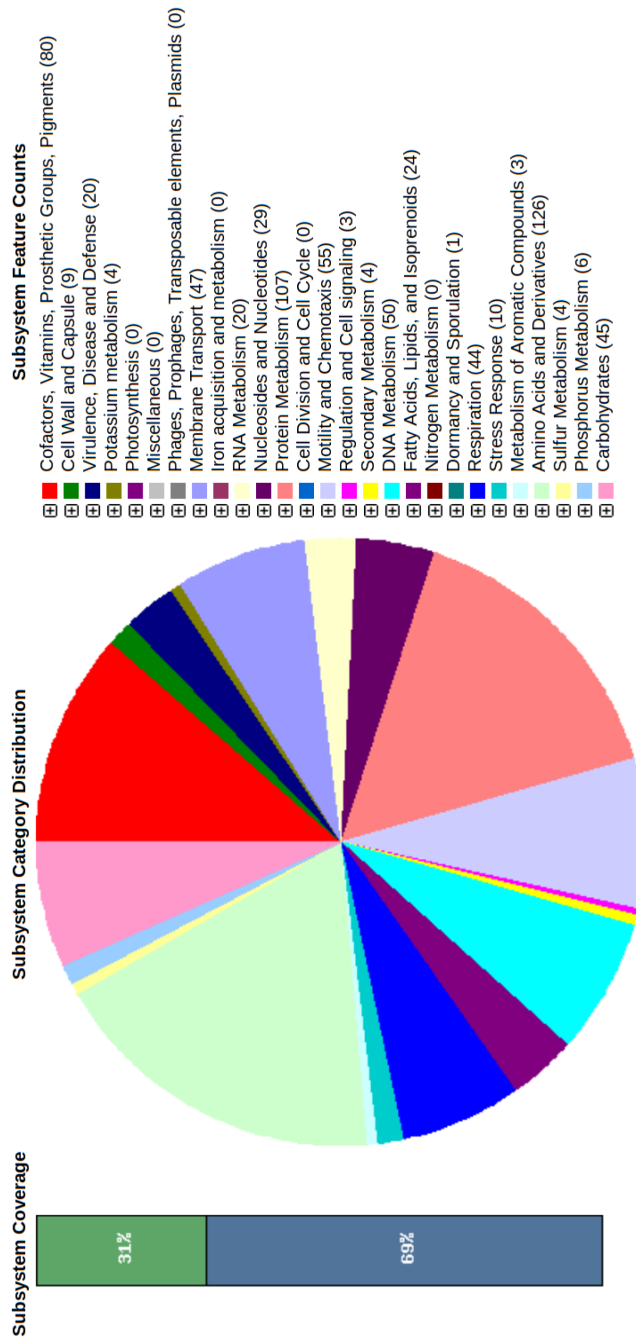


Figure AIII.3 Pie charts representing RAST functional subsystems in the B508A-S1 (**a**), B508A-T2A (**b**), and B508A-T4 (**c**) genomes. The pie chart shows the % distribution of 19 most abundant subsystems on the "category" level, each of which is represented by a particular colour indicated in the right column showing the counts of features.

ANNEX IV. Publications during the PhD period

Albarral, V., Sanglas, A., Palau, M., Miñana-Galbis, D., Fusté, M.C., 2016. Potential pathogenicity of *Aeromonas hydrophila* complex strains isolated from clinical, food, and environmental sources. *Canadian Journal of Microbiology* 62(4), 296–306. doi: 10.1139/cjm-2015-0466

Berlanga, M., Palau, M., Guerrero, R., 2017. Functional stability and community dynamics during spring and autumn seasons over 3 years in Camargue microbial mats. *Frontiers in Microbiology* 8, 2619. doi: 10.3389/fmicb.2017.02619

Berlanga, M., Palau, M., Guerrero, R., 2018. Gut microbiota dynamics and functionality in *Reticulitermes grassei* after a 7-day dietary shift and ciprofloxacin treatment. *PLoS ONE* 13(12), e0209789. doi: 10.1371/journal.pone.0209789

Boujida, N., Palau, M., Charfi, S., El Moussaoui, N., Manresa, A., Miñana-Galbis, D., Skali Senhaji, N., Abrini, J., 2018. Isolation and characterization of halophilic bacteria producing exopolymers with emulsifying and antioxidant activities. *Biocatalysis and Agricultural Biotechnology* 16, 631–637. doi: 10.1016/j.bcab.2018.10.015

Boujida, N., Palau, M., Charfi, S., Manresa, À., Senhaji, N.S., Abrini, J., Miñana-Galbis, D., 2019. *Marinobacter maroccanus* sp. nov., a moderately halophilic bacterium isolated from a saline soil. *International Journal of Systematic and Evolutionary Microbiology* 69(1), 227–234. doi: 10.1099/ijsem.0.003134

Palau, M., Boujida, N., Manresa, A., Miñana-Galbis, D., 2018. Complete genome sequence of *Marinobacter flavimaris* LMG 23834T, which is potentially useful in bioremediation. *Genome Announcements* 6(16), e00273-18. doi: 10.1128/genomeA.00273-18

Palau, M., Kulmann, M., Ramírez-Lázaro, M.J., Lario, S., Quilez, M.E., Campo, R., Piqué, N., Calvet, X., Miñana-Galbis, D., 2016. Usefulness of housekeeping genes for the diagnosis of *Helicobacter pylori* infection, strain discrimination and detection of multiple infection. *Helicobacter* 21(6), 481–487. doi: 10.1111/hel.12304

Piqué, N., Palau, M., Berlanga, M., Miñana-Galbis, D., 2016. Advances in the research of new genetic markers for the detection of *Helicobacter pylori* infection. In: *Recent Advances in Pharmaceutical Sciences*. pp. 165–188

ANNEX V. Participation in projects, Consolidated Research Groups and research stays

1. Projects

Project:	**Noves tecnologies com a eines no invasives per a la prognosi/diagnosi del càncer gàstric**
Funding institution:	Fundació La Marató de TV3
Reference:	1007/C/2013
Principal investigator:	Dr. David Miñana i Galbis
Duration:	2014–2017

Project:	**Estudio metagenómico para evaluar el papel de la microbiota y las variantes alélicas de *Helicobacter pylori* en la progresión a cáncer gástrico**
Funding institution:	Roche Diagnostics SL
Reference:	METAGC14
Principal investigator:	Dr. David Miñana i Galbis
Duration:	2014–2016

2. Consolidated Research Groups

Project:	**Grup de recerca consolidat 'Grup de Recerca en Inflamació Gastrointestinal i Malalties Hepàtiques'**
Funding institution:	Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR)
Reference:	2017SGR853
Principal investigador:	Dr. Xavier Calvet Calvo (Fundació Parc Taulí)
Duration:	01/05/2017–31/12/2019

Project:	**Grup de recerca consolidat 'Genòmica i proteòmica dels factors de virulència bacterians'**
Funding institution:	Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR)
Reference:	2014SGR106
Principal investigador:	Dr. Juan Tomás Magaña
Duration:	01/01/2014–30/04/2017

3. Research stays

Project:	**Genomic and metagenomic tools to find novel *H. pylori* virulence markers**
Programme:	ERASMUS-MUNDUS Action 2, Strand 2, Lot 1
Reference:	545700-EM-1-2013-1-ES-ERAMUNDUS-EMA22
Place of stay:	Langille lab. Dalhousie University, Halifax, Nova Scotia, Canada
Duration:	01/09/2016-31/07/2017