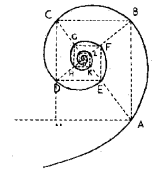
UNIVERSITÀ DEGLI STUDI DI MILANO



DOTTORATO IN MEDICINA MOLECOLARE E TRASLAZIONALE

CICLO XXIX

Anno Accademico 2015/2016

SETTORE BIO/10

Development of a webtool for interactome-sequencing data analysis and identification of *H. pylori* epitopes responsible for host immuno-response modulation.

Dottorando : Simone PUCCIO

Matricola N° R10708

TUTORE : Prof.ssa Cristina Battaglia
CO-TUTORE: Dott.ssa Clelia PEANO

COORDINATORE DEL DOTTORATO: Prof. Mario Clerici

ABSTRACT

To elucidate the molecular mechanisms involved in persistency/latency of the *H. pylori* infection or in its progression towards serious diseases, it is necessary to analyse the host pathogen interaction *in vivo*. The circulating antibody repertoire represents an important source of diagnostic information, serving as biomarker to provide a "disease signature".

The aim of this work is the identification of *H. pylori* epitopes responsible for host immuno-response modulation through: a discovery-driven approach that couples "phage display" and deep sequencing (interactome-sequencing) and the development of a specific webtool for interactome-sequencing data analysis.

We used this approach to identify novel antigens by screening gDNA libraries created from the pathogen's genome, directly with sera from infected patients. Two genomic phage display libraries from 26695 and B128 *H. pylori* strains have been constructed by using ß-lactamase ORF selection vector. Genomic DNA was sonicated, fragments cloned into the filtering vector, after transformation libraries of $1x10^6$ clones were obtained and sequenced by Illumina technology. More than 93% of Hp CDSs were represented in the phage genomic libraries therefore being representative of the whole *H. pylori* antigenic ORFeome.

A webtool for interactome-sequencing data analysis was developed and used to identify the *H. pylori* antigens/epitopes which could be considered specific for infection progression towards three different pathological outcomes.

Putative antigens were selected from libraries using sera from patients affected by: i) gastric adenocarcinoma; ii) autoimmune gastritis; iii) MALT lymphoma. The results, obtained thanks to the new interactome sequencing pipeline developed, show that the diversity of the libraries after selection is significantly reduced. Furthermore, individual ranks, for each infection condition, have been compared highlighting the pattern of putative antigens, shared by all the conditions, and some that can distinguish the different stages of infection.

One of this new antigens, that seems to be specific for infection progression towards more serious diseases, has been successfully validated through ELISA assay on a wide number of sera from patients. Other more specific antigens identified by our approach and by the application of the new data analysis pipeline here described are in validation.

SOMMARIO

Per comprendere al meglio i meccanismi molecolari coinvolti nella persistenza/latenza dell'infezione da *H. pylori* o nella sua progressione verso patologie più gravi, è necessario analizzare l'interazione ospite patogeno *in vivo*. Il repertorio di anticorpi circolanti rappresenta un importante risorsa di informazioni diagnostiche, utili per la definizione di una "disease signature".

Lo scopo di questo lavoro è l'identificazione degli epitopi di *H. pylori* responsabili per la modulazione della risposta immunitare attraverso un approccio guidato che accoppia la tecnologia del “phage display” e del sequenziamento ultra-massivo (interactome-sequencing) e lo svilippo di una specifica applicazione Web per l'analisi di questa tipologie di dati.

Questo approccio è stato applicato per identificare nuovi antigeni attravero lo screening di librerie fagiche costruite a partire dal DNA genomico di *H. pylori* direttamente con i sieri di pazienti affetti e con diversi outcome patologici. Sono state costruite due librerie fagiche genomiche dai ceppi *H. pylori* 26695 e B128 utilizzando un vettore di selezione delle ORF che sfrutta come folding reporter la ß-lattamasi. Il DNA genomico è stato sonicato e i frammenti sono stati clonati nel un vettore di filtraggio utilizzato poi per la trasformazione di *E. coli;* sono state così ottenute librerie di $1x10^6$ cloni e sequenziate con la tecnologia Illumina. Più del 93% delle sequenze codificanti (CDS) di *H. pylori* sono rappresentate nelle librerie genomiche che posso quindi essere considerate rappresentative dell'intero ORFe-oma/Domain-oma antigenico del patogeno.

Successivamente ho sviluppato un webtool per l'analisi dei dati provenienti da esperiemnti di interctome-sequencing e lo strumento di analisis sviluppato in questo lavoro di tesi è stato utilizzato nell'identificazione di antigeni di *H. pylori* specifici per le tre patologie prese in considerazione: i) adenocarcinoma gastrico ii) gastrite autoimmune iii) linfoma MALT. . I risultati prodotti grazie al nuovo strumento di analisis sviluppato, dimostrano che la diversità delle librerie dopo la selezione è significativamente ridotta. Inoltre le liste ottenute per ognuna delle tre patologie sono state confrontate evidenziando un

pannello di antigeni comuni e di antigeni specifici per i diversi outcome patologici.

Uno degli antigeni identificati, che sembra essere specifico per la progressione dell'infezione verso patologie più gravi, è stato già validato con successo tramite saggi ELISA su un numero ampio di sieri di pazienti. Altri antigeni identificati con questo approccio e grazie alla nuova strategia di analisi da me sviluppata e qui descritta sono attualmente in fase di validazione.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ABBREVIATIONS

AIG: Autoimmune gastritis

AP-MS: gel electrophoresis and mass spectrometry

ChIP-seq : Chromatin immunoprecipitation sequencing CDS: Coding sequence

DEG : Differentially Expressed Gene

FDR : False discovery rate

GA: Gastric Adenocarcinoma

HP : Helicobacter pylori

Interactome-Seq : Interactome-Sequencing

MALT: Mucosa-associated lymphoid tissue

NGS : Next-generation sequencing

ORF: Open reading frame

RUT : rapid urease test

Y2H: yeast two-hybrid

WHO: World Health Organization

1. INTRODUCTION

1.1. Helicobacter pylori

The pathogen, *Helicobacter pylori* is a gram-negative, neutralophilic organism, that colonizes half of the world's human population. *H. pylori* was discovered in 1984 by Barry Marshall and Robin Warren and for demonstrating its role in inducing gastritis, peptide ulcer formation and gastric cancer [1], they were awarded the Nobel Prize for medicine in 2005. This finding broke the dogma of the stomach as a sterile organ. In 1994 the International Agency for Research on Cancer classified *H. pylori* as a class I carcinogen [2]. *H. pylori* has evolved specific mechanisms to colonize and persist within the stomach in spite of the harsh acidic conditions encountered in this environment. Different genomic analysis has revealed symbiotic relationship between human host and *H. pylori*. Perturbation of this relationship leads to the dysregulation of this host-pathogen interaction, resulting in the development of severe gastroduodenal disease, including adenocarcinoma, autoimmune gastritis and gastric MALT lymphoma [3]. *H. pylori* is an epsilonproteobacterium and a member of the Helicobacteraceae family [4]. Helicobacters are classified into two types according to their customary niche: gastric and enteric. The Helicobacter species adapted to humans is *H. pylori* and is a gastric Helicobacter.

1.2. Epidemiology

The World Health Organization (WHO) has estimated that close to half of the world's population is positively infected with *H. pylori*. As human populations transmigrated across the world their endemic *H. pylori* strains diverged aboard them leading to phylogeographic differentiation of this infective agent that can be grouped into Amerindian, European,Asian and Africans subgroups. In various nonindustrial states, more than 81% of the population is positive affected by *H. pylori*, and infection occurs even among young people [5]. The *H. pylori* prevalence in developed countries is under 41% and is substantially higher in adults or in elderly individuals than young people [6]. In certain

geographical areas, the proportion of individuals affected bt *H. pylori* infection is reciprocally correlates with social and economic status, in particular there is a specific correlation to life conditions during prepuberty [7]. Meanwhile the prevalence of *H. pylori* infection in underdeveloped countries is relatively constant, in the industrialized countries is rapidly declining [8]. The causes of this decrease is due to the improvement of hygiene and sanitation via antimicrobial treatment [9].

1.3. Transmission

The mechanisms of transmission whereby *H. pylori* is acquired are not well known. *H. pylori* infection is believed to occur as a effect of direct person-to-person transmission, via either an oral-oral or fecal-oral route or both. *H. pylori* has been found in different biological specimens like saliva, feces, vomitus and also in gastric refluxate, but there is no clear evidence for prevalent transmission through any of these specimens [10,11]. Various reports have shown the contamination of *H. pylori* DNA in environmental water sources, but this is due to the contamination with dead *H. pylori* organisms [12]. An increased risk of infection was observed among hospitalized adolescent people during outbreaks of gastroenteritis [13]. Conjugated with the extreme sensitivity of the organism to atmospheric oxygen pressure, lack of nutrients, and temperatures outside the range of 34°C-40°C, direct human-to-human transmission is the most probable transmission route [14].

1.4. Genomic diversity and flexibility

The *H. pylori* genome consists of 1.65 million bp and codes for about 1,500 proteins and it has an average GC content of 38.9% [12]. *H. pylori* expresses an uncommon genetic variability that could potentially account for the bacterium's ability to adapt within the host gastric stomach [13]. These adaptations comprehend both temporary and permanent changes to the genomic conformation such as regulatory mechanisms that modulate gene expression [14]. Analyses of numerous whole genome sequences indicate that

H. pylori has evolved clusters of genes within genomic islands that create distinct areas of variability, called “plasticity zones”. These plasticity zones are are prevalent and widely distributed among almost all *H. pylori* strains and are also implicated in lateral gene transfer, they allow fitness advantages to the organism by promoting genetic recombination events, which are involved in immune evasion and host colonization mechanisms [15]. Research recent published supports this hypothesis and indicates that genetic modifications occur ten times faster during host infection, when *H. pylori* is initially encountering the host’s immune response. A study published by Linz B. et al have shown that *H. pylori* mutation rates of genes encoding for outer membrane proteins (OMPs) are higher than any other bacteria tested [16]. Brodsky IE et al. have shown that OMPs are involeved in host-pathogen interaction and are targets for the adaptive immune response [17]. *H. pylori* mutating these exposed antigens promotes the chronicization of infection and the genes encoding OMPs under high selective pressure [18].

2. *HELICOBACTER PYLORI* ASSOCIATED DISEASES

H. pylori typically is localized in the mucus away from the epithelial surface and does not directly adhere with gastric epithelial cells. Interaction of *H. pylori* with host cells activates many signaling pathways, the release of toxins or other signaling molecules [19]. In all subjects persistently colonized by *H. pylori* occurs a humoural immune response together with tissue infiltration by mononuclear and polymorphonuclear leukocytes [20]. The host inflammatory response is comparatively weak when compared with the response that occurs against others transient bacterial pathogens. As mentioned above, *H. pylori* persistently colonizes the human stomach for many years, causing alterations in the gastric niche and induces the host immune response, but its own peculiarity is to be a symbiotic organism usually and in many cases does not cause adverse effects [21]. What factors contribute to the stability of the *H. pylori*-host equilibrium? One important element is the presence of the bacterium within the gastric mucus layer, without any substantial adhesion of

host tissue. Another important element is the synthesis of *H. pylori* factors that are highly adapted to decrease the intensity of the host defense system [22].
H. pylori lipopolysaccharide (LPS) is distinguished by modifications of the lipid A component that make it less proinflammatory than lipoglycans found in the outer membrane of other Gram-negative bacteria [23]. Although *H. pylori* typically colonizes the stomach for many years without specific symptoms, its presence is correlated with an increased risk of several pathological outcomes [24]. The risk of these outcomes is related to different factors like host genetic and environmental factors host responses and interactions between host and microbe. The progression steps from normal stomach to inflammation and precancerous conditions were described by Correa in 1992 [25]. In Correa's cascade, host responses and interactions between host and microbe modulate the transition from one step to next. The main gastric pathological outcomes related to *H. pylori* infection are (Figure 2):

- Gastric adenocarcinoma (GA), which develops in gastric cancer in 1-3% of infected individuals;
- Autoimmune gastritis (AIG), developed by the great majority of patients but that mostly remains asymptomatic;
- Gastric mucosa-associated lymphoid tissue lymphoma (MALT lymphoma), which develops in 0.1% of infected subjects;

2.1. Gastric adenocarcinoma

Gastric adenocarcinoma (GA) is the second principal cause of cancer-related death worldwide and between the common cancer is the fourth, every year causes about 700,000 deaths. The incidence rates of GA is not uniform and varies depending on the country, with the highest rates found in Asia (mainly in the east), Central America, some regions of South America, and Eastern Europe [26] (Figure 1).

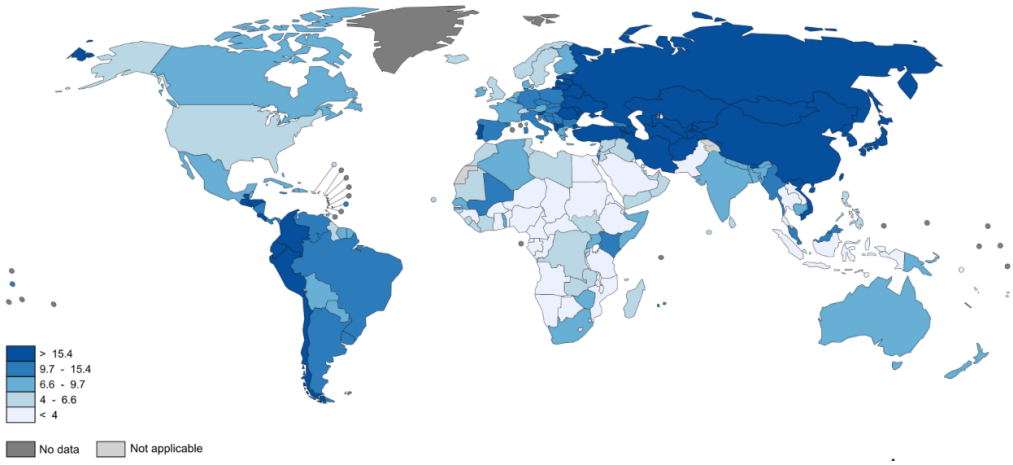


Figure 1 - Map of global distribution of Incidence from gastric cancer. Image reprinted by [26].

As shown in Figure 2 GA is initiated by the transition from normal mucosa to chronic superficial gastritis; this is followed by the development of atrophic gastritis and metaplasia, finally leading to dysplasia and adenocarcinoma [27]. GA affects male twice as frequently as female; it is thought that estrogen may confer some protection in women. Over the past century, the incidence rate of GA in developed countries is remarkably decreased [2]. This decline can primarily be attributed to a decline in intestinal-type adenocarcinomas in the distal stomach [26]. Currently, in the USA, distal GA is diagnosed most commonly in elderly people and occurs more often in African-Americans, Hispanic-Americans, and Native Americans than among other ethnicities. While the incidence of GA of the distal stomach has been declining, the incidence rates of proximal GA as well as those originating in the gastroesophageal junction have been increasing in both the United States and Europe [28].

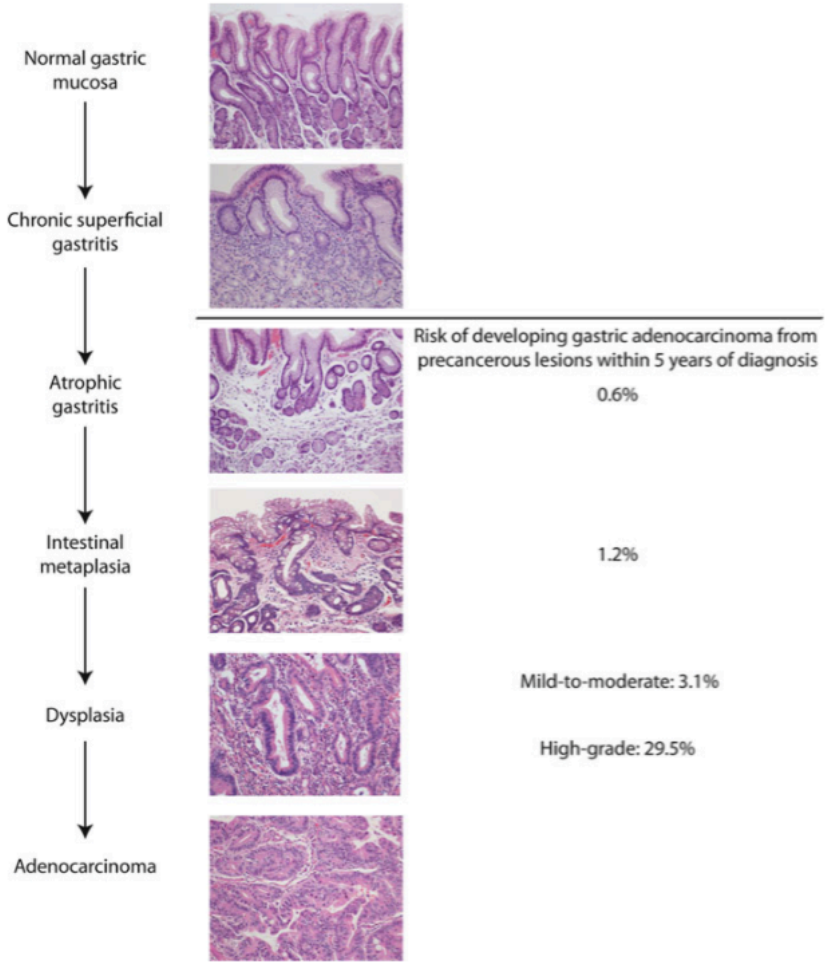


Figure 2 - Correa's cascade proposed in 1992. Image reprinted from [25]

2.1.1. Signs and symptoms

It is usually difficult to diagnose GA at an early onset stage of because it often does not cause particular symptoms. When symptoms do occur, they may be not specific and these symptoms can also be caused by many other illnesses, such as a stomach virus or an ulcer. The signs and symptoms associated with GA are similar to other gastric disease and they include indigestion, discomfort in the abdomen, nausea and vomiting, diarrhea or constipation, loss of appetite, sensation of food getting stuck in the throat while eating, unexplained weight loss [29].

2.1.2. Cause

Besides *H. pylori*, the pathogenesis of GA is correlated with diet and cigarette smoking. High dietary salt intake was found to be associated with an increased risk of GA [30]. Two studies, from South Korea and Japan stated that *H. pylori*-infected subjects taking high-salt diet had a greater risk of GA than subject who underwent a low-salt diet [31], even if the mechanisms that link a high-salt diet with the increase of GA development remains unclear. The risk of developing GA in *H. pylori* infected subjects is also associated with cigarette smoking. By several studies, it emerges that exists a relationship between *H. pylori* infection and sigarette smoking with enhanced risk of developing GA [32,33].

2.1.3. Pathophysiology

The pathogenesis of *H. pylori* associated to GA is a multi-factorial and multi-step process, and as mentioned in the previous chapter its development depends on a combination of host, bacterial and environmental factor as shown in Figure 3. Many studies focused on the effects of *H. pylori* infection on the host epithelium. The gastric epithelium has the function to protect the underlying tissue and has also the responsibility for digestive processes. *H. pylori* has developed sophisticated mechanisms to avoid host defence and to adapt to the immune system response in order to colonize the gastric niche, such, for example, disruption of epithelial junctions, stimulation of cytokine production, overproliferation, DNA damage, apoptosis and DNA transformation [34,35]. *H. pylori* disrupts intercellular apical junctions, these are critical in keeping the integrity of mucus layer, furthermore these intercellular junctions are indispensable for cellular functions because through the binding to specific cellular receptors they stimulate the signaling pathways [36]. The *H. pylori* virulence factor CagA is translocated into epithelial cells through the Type IV secretion system (T4SS) and interacts with membrane proteins like E-cadherin and ZO-1 and disrupts the tight of adherence junctions. It has been confirmed that E-cadherin interacts with β-catenin, a component of a cell adhesion

complex [37]. The complex E-cadherin/β-catenin play pivotal role in maintaining epithelial integrity and in the stabilization of cellular architecture [34,38]. During *H. pylori* infection, the structure of this complex is destabilized by the virulence factor CagA in a phosphorylation independent manner [39]. The evidences support that also the virulence factor VacA interacts with epithelial cells inducing autophagy [40]. Other studies demonstrated that to facilitate the intracellular survival and for promoting the carcinogenesis *H. pylori* suppress autophagy [41]. Saberi S. et al have shown that the fast turnover of epithelial cells is an important element to protect the gastric epithelium from infection, but *H. pylori* blocks this turnover inducing apoptosis in order to favour its survival [37]. A deleterious consequence of apoptosis iduction is an imbalance in normal gastric mucosal homeostasis leading to dysregulated tissue growth [42]. The mechanism behind this process is not clear, enzymes such as urease and virulence factors such as VacA lead to cellular apoptosis by damaging the gastric epithelium and activating the immune response [43]. The high level of free radicals generated by neutrophils and cytokines like IFN-γ in inflammatory response can damage DNA and stimulate apoptosis of gastric epithelial cells. Proinflammatory cytokines including IL-8, IL-6, MCP-1, TNF-α, MIF, IL-1α, TGF-β and IL-1β are secreted by epithelial cells and these play a fundamental role in pathogenesis of GA [44]. Another important aspect involved in pathogenesis of GA is the polymorphisms of cytokine genes. As reported by Amieva et al. the risk of GA in many populations was affected by the polymorphism of the genes encoding IL-1β, TNFα, IL-8, IL-17 and IL-10 [45]. These polymorphisms are important in gastric carcinogenesis because they affect the oxidative stress process and DNA mutagenesis, this phenomenon could also explain the occurrence of GA only in a small proportion of *H. pylori*-infected individuals [46].

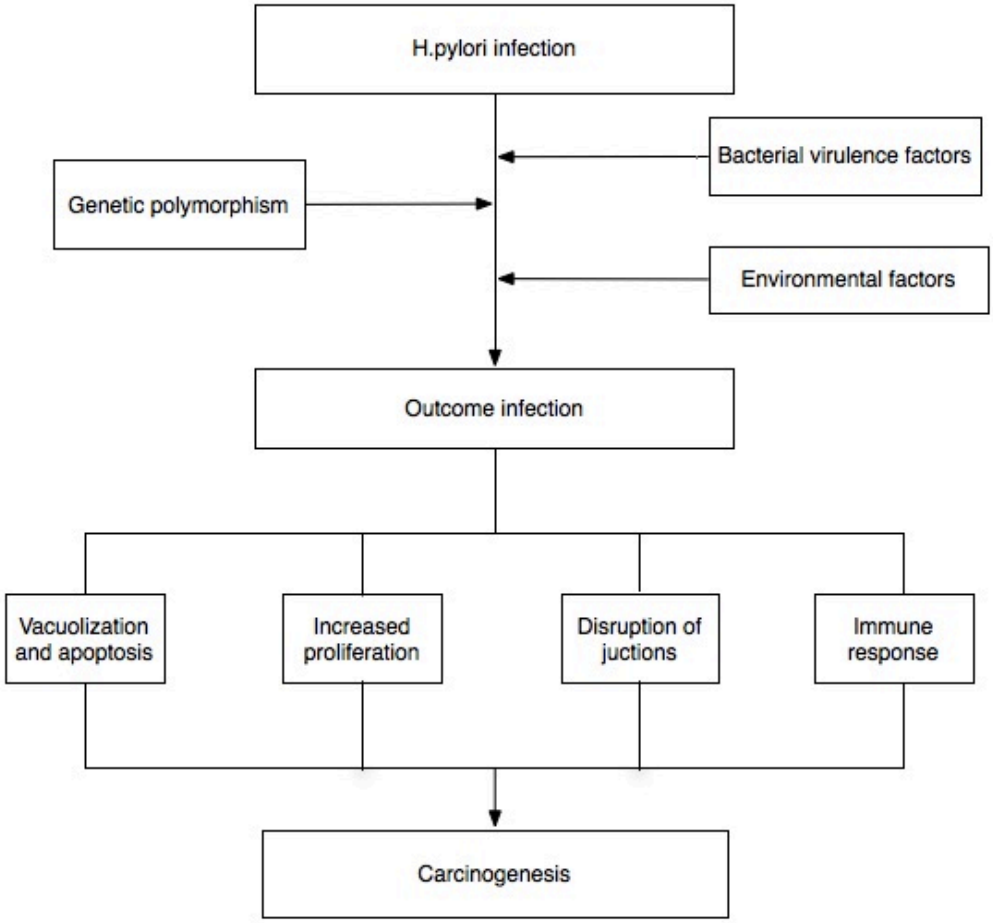


Figure 3 - Role of *H. pylori* in the gastric carcinogenesis. Host genetics, environmental factors, and bacterial strain differences in virulence properties contribute to disease progression and increased risk of negative outcomes. Image reprinted from [36]

2.1.4. Diagnosis

The diagnosis of GA is based on endoscopy and biopsy. Endoscopic ultrasonography and computed tomography (CT) of the chest and abdomen are currently the primary means of staging for the diagnostic of gastric cancer. Endoscopic ultrasonography and computed tomography are uncomfortable, painful and expensive, emphasizing the need for rapid and non-invasive tests for screening and monitoring the patients with dyspeptic symptoms. Recently has been developed a panel which combines a serum pepsinogen-I (PGI) and -II (PGII), gastrin-17 (G-17) and HP IgG antibodies (IgG-HP) [47,48]. These

biomarkers were tested in a cohort of 91 patients revealing a low sensitivity and specificity, for these reasons gastroscopic exams and CT remain the most commonly prescribed diagnostic test.

2.2. Autoimmune Gastritis

Autoimmune gastritis (AIG) is a chronic inflammation of the stomach caused by an indirect action of the *H. pylori* on the immune system. The term AIG includes a variety of definitions such as atrophic body gastritis and pernicious anemia [49].

2.2.1. Signs and symptoms

Patients with AIG are asymptomatic for many years, until parietal cell mass reduced capacity to secrete HCL acid causing hypochlorhydria or achlorhydria. Abnormal gastrin production (hypergastrinemia) occurs in achlorhydric patients, the clinical features of AIG are largely due to the pernicious anemia, weight loss occurs in about 50% of patients and AIG are occasionally associated also with sore tongue and neurological problems including numbness, weakness, and ataxia. Patients with AIG may have other autoimmune diseases such as Graves' disease, a condition in which the thyroid gland produces excessive hormones and idiopathic adrenocortical insufficiency [49].

2.2.2. Cause

H. pylori is the main aetiologic factor for AIG, the degree of mucosal inflammation results from the interplay of *H. pylori* virulence factor, genetic host susceptibility and environmental factors. AIG associated with *H. pylori* infection is limited to the antrum of the stomach (Type B gastritis). Other causes are the excessive and long term assumption of Nonsteroidal anti-inflammatory drug (NSAID); less common causes include alcohol, cocaine and Crohn disease [50].

2.2.3. Pathophysiology

AIG is an immune-based, non-infectious inflammation of the stomach, as a result of molecular mimicry between *H. pylori* antigens and H+/K+-ATPase. The hypothesis is based on the observation that *H. pylori*-infected patients with AIG harbour gastric CD4+ T-cells that recognize cross-reactive epitopes in common between gastric H+/K+-ATPase and *H. pylori* proteins. Cross-reaction antigens shared by H. pylori induce T-cell proliferation, production of Th1 cytokine and H+/K+-ATPase antibodies were identified in about 22-30% of *H. pylori* infected patients [51]. The clones deriving from T-cell proliferation reacted to epitopes different to the proton pump alone. Consequently, this autoimmune response induces the destruction of parietal cells and dysfunction of the oxyntic gastric mucosa [50].

2.2.4. Diagnosis

Despite the significant progress in the understanding of pathogenesis of chronic inflammation in AIG, diagnosis is based on histology and detection of the presence of Parietal Cell Antibody (PCA). Literature data on association of AIG with PCA is weak; in a clinical study of 2012 on a cohort of 108 patients with AIG 82% were positive for PCA while in a similara study a cohort of 140 patients with AIG 75% had PCA, which correlated positively with the grade of atrophy [50]. Histology coupled with endoscopy can provide a unequivocal diagnosis of AIG. Through endoscopy multiple gastric biopsy sampling is performed, diagnosis of AIG requires a minimum of 4 biopsies from antrum and corpus of stomach [48].

2.3. Gastric mucosa-associated lymphoid tissue lymphoma (MALT lymphoma)

Extranodal marginal zone B cell lymphoma of mucosa-associated lymphoid tissue (MALT lymphoma) is the third most common subtype of non-Hodgkin's lymphoma [52]. The microbial pathogens that underlie such chronic inflammatory diseases, also play a pivotal role in both malignant transformation and successive clonal expansion of the transformed clone. This is well exemplified by the causative role of *H. pylori* infection in the development of gastric MALT lymphoma, the definition of this causative relationship led to successful innovative treatment of lymphoma with antibiotics [53].

2.3.1. Signs and symptoms

Gastric MALT lymphoma is an indolent disease and the clinical presentation is poorly specific. Symptoms ranging from fatigue, low-grade fever and nausea to alarm symptoms such as gastrointestinal bleeding or persistent vomiting [54]. Others symptoms like weight loss or night sweats are extremely rare in stomach MALT lymphoma. A systematic review published in 2010 including data of 2000 patients found that gastric MALT lymphoma occurs in a wide age range of patients and sex ratio incidence is essentially equal, and only in 42% of patients alarm symptoms were present [55].

2.3.2. Cause

Different infections have been involved in the progression of gastric MALT lymphoma. There is a clear correspondence between gastric H. pylori infection and MALT lymphoma diagnosed in the stomach (90% of cases) [56]. In fact, eradication of *H. pylori* leads to the disappearance of intratumoral *H. pylori* specific T-cells and thus to the removal of the growth support of neoplastic B-cells inducing the regression of lymphoma [57]. More or less, 5-10% of gastric MALT lymphomas are negative for *H. pylori* and can be associated to *H.heilmannii* infection [58].

2.3.3. Pathophysiology

The importance of *H. pylori* in the pathogenesis of gastric MALT lymphoma are confirmed by direct evidence obtained from in vitro studies. These studies show that lymphoma growth is stimulated in culture when exposed to the bacterium, through tumour-infiltrating lymphocytes (TILs) involving two costimulatory molecules: CD40 and CD40L. The infiltrated B cells are activated by T-cells specifically recognising *H. pylori* and undergo towards malignant transformation due to the acquisition of genetic abnormalities [59]. Epithelial cells are activated by chronic infectious stimuli, expressing on their surface high levels of HLA-DR (Human Leukocyte Antigen - antigen D Related) and costimulatory molecules, including CD80. Epithelial cells present antigens along with HLA molecules to T cells, CD40 ligand molecules expressed on the activated T cells can react with the CD40 molecule on B cells, consequently the expression of B cell CD80 is upregulated. CD80 is a protein found on activated B cells that provides a costimulatory signal necessary for T cell activation by binding CD28 [60]. Activated CD4 T-cells together with the action of various cytokines and chemokines can stimulate B-cells through CD40L-CD40 interaction. Activated epithelial cells, T cells and B cells allow to survive cooperatively in lymphoepithelial lesions (LELs), that are thought to be at the origin of lymphomas [61].

2.3.4. Diagnosis

Gastric MALT lymphoma associated with *H. pylori* infection presenting nonspecific symptoms, for this reason diagnosis is based on invasive approach and requires endoscopic biopsy [62]. Recent advancement in endoscopic techniques allow an accurate diagnosis in 91% of case [63]. Endoscopic biopsy appearances of MALT lymphoma are an atypical lymphoid infiltrate, lymphoepithelial lesions and moderate cytological atypia of lymphoid cells. Many biopsy specimens lack some of these characteristics [64]. In 1993, Dr A.C Wotherspoon proposed a useful scoring system, shown in the Table 1, currently in use, to better defining and differentiating the gastric lymphoid

infiltrates [65]. The diagnosis should also consider *H. pylori* infection, for this reason final diagnosis should include histology, rapid urease testing, urea breath testing, stool antigen testing, or serologic studies [66].

Score	Diagnosis	Histological features
0	Normal	Scattered plasma cells in lamina propria. No lymphoid follicles
1	Chronic active gastritis	Small clusters lymphocites in lamina propria. No lymphoid follicles. No lymphoepithelial lesions.
2	Chronic active gastritis with florid lymphoid follicle formation	Prominent lymphoid follicles with a surrounding manntle zone and plasma cells. No lymphoepithelial lesions.
3	Suspicious lymphoid infiltrate, probably reactive	Lymphoid follicles surrounded by small lymphocytes that infiltrate diffusely in lamina propria and occasionally into the epithelium
4	Suspicious lymphoid infiltrate, probably lymphoma	Lymphoid follicles surrounded by marginal-zone cells that infiltrate diffusely in lamina propria and into the epithelium in small groups
5	MALT lymphoma	Presence of dense diffuse infiltrate of marginal-zone cells in lamina propria with prominent lymphoepithelial lesions.

Table 1 - Histological criteria established by Wotherspoon. Table reprinted by [65].

3. MOLECULAR MECHANISM OF H. PYLORI INFECTION AND PATHOGENESIS

To promote chronic infection *H. pylori* has evolved a variety of mechanisms to survive in the acidic environment of gastric mucosa. To neutralize the hostile acid condition at beginning of infection *H. pylori* has developed an “acid acclimation mechanism” that promotes adjustment of periplasmic pH in the stomach by regulating activity of urease and carbonic anhydrase [67]. Once the bacterium has created a favourable environment in term of pH, the next step is the mobility through the protective layer of gastric mucus in the host stomach. Motility of *H. pylori* depends on the presence of up to 6 functional unipolar flagella [68]. *H. pylori* regulates cell motility by responding chemotactic cues: it has been demonstrated that the operon Tlp A B C D, and in particular TlpD is necessary for bacterial persistence in the murine stomach and growing in the infected and inflamed antrum [69]. Flagella-mediated motility is followed by specific interactions between bacterial adhesins with host cell receptors,

which thus leads to successful colonization and persistent infection [70]. Finally, *H. pylori* releases several effector proteins/toxins causing host tissue damage.

The molecular mechanism by which *H. pylori* causes disease in humans can be described as a multi-step process. In summary, four steps are critical for H. pylori colonization and pathogenesis (Figure 4):

- Survival under acid stomach conditions;
- Movement toward epithelium cells through flagella-mediated motility;
- Attaching to host receptors by adhesins;
- Causing tissue damage by toxin release;

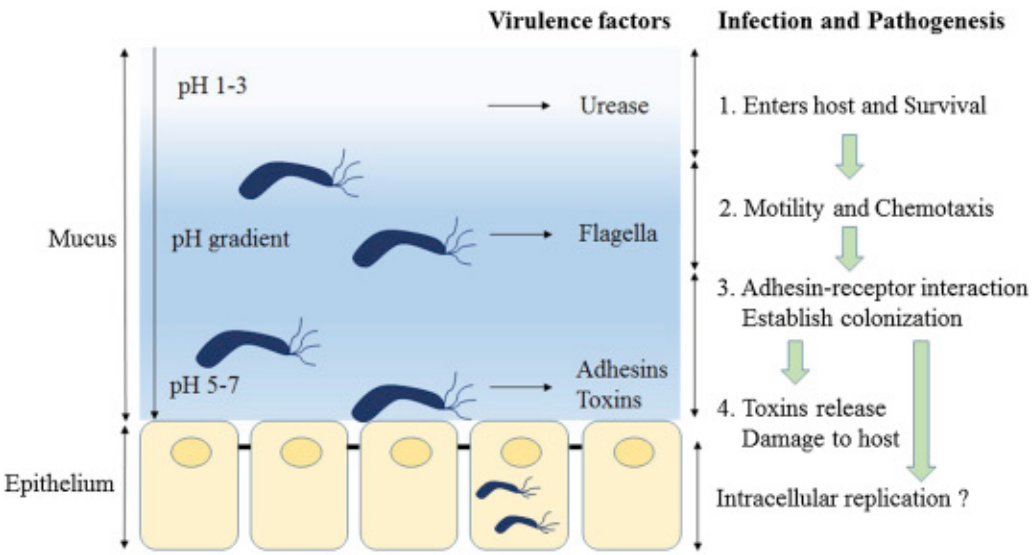


Figure 4 - Schematic representation of *H. pylori* infection and pathogenesis mechanism. Image reprinted from [71].

3.1. Urease and survival under acidic stomach conditions

H. pylori has adapted an original mechanisms to colonize and persist within the gastric environment, for acid resistance the organism uses urea to produce ammonia through the metalloenzyme urease, that catalyzes the conversion of urea into ammonia and CO_2 .The NH_3 produced by this reaction increases the local pH surrounding the bacterium thus allowing its survival in the hazardous acidic environment of the human stomach [72]

The constitutive production of the urease accounts up to 10% of the proteins synthesized by the organism at pH 7.0. In several bacterial species ureases are entirely cytoplasmic, instead *H. pylori* urease is found both inside the cytoplasmic and outside adhering to the surface of the organism. For many years it has been believed that external urease provides the acid resistance of *H. pylori* but there is no evidence supporting this hypothesis, furthermore the extreme acidity encountered in the stomach irreversibly inactivates this enzyme [22].

Instead urease localized in cytoplasmic protects the organism by producing uncharged ammonia inside this compartment, after synthesis ammonia is carried out to periplasmic space where it can bind protons raising the periplasmic pH and preventing a high concentration of protons from migrating into the cytoplasm and at the same time avoiding the pH decreasing to toxic levels [66]. The massive ammonia synthesis by urease activity constitutes the principal protection against proton overload, acid acclimation depends both on urea degradation and ammonia synthesis but also on urea uptake across the bacterial inner membrane [73]. Urea is transported from the periplasmic space into the cytoplasm where it is hydrolysed by urease for the production of ammonia through the channel transporter UreI [74]. The control of the periplasmic pH on acid exposure is unique to *H. pylori*, and make *H. pylori* colonization durable and persistent [75].

3.2 Bacterial motility toward epithelium cells

Flagella-mediated motility is essential for the *H. pylori* colonization of gastric mucosa, flagella can be considered as an early stage colonization/virulence factor [70]. *H. pylori* flagellum is mainly composed of the basal body hook and flagellar filament, that consists of two flagellins (FlaA and FlaB) [76]. The hook is composed of FlgE, and it links the basal body and flagellar filament. The basal body is composed of several structures, and it plays a role in providing the energy source for motility [77]. Several study have indicated *flaA* and *flaB* as necessary genes for the complete motility of *H. pylori* [76]. Recently, the

role of flagellins in bacterial adherence to mammalian hosts has been demonstrated by constructing flagellins (*flaA* and *flaB*) and flagellar regulator (*fibA*) mutants and it was shown that all mutants are still able to adher to gastric cells, supporting the hypothesis that flagella do not play a direct role in adhesion of *H. pylori*. Although a lower adhesion rate in a *flbA* mutant suggested that in addition to regulating flagella, FlaA may regulate some *H. pylori* adhesins [76] (Figure 5).

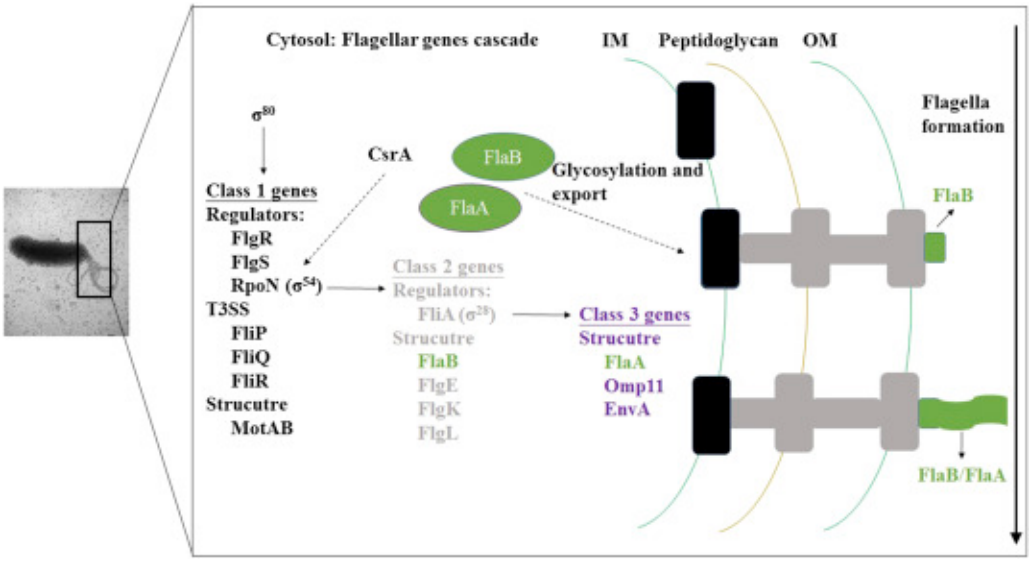


Figure 5- Schematic model of the flagellar transcriptional regulatory cascade for *H. pylori* flagellar biosynthesis. Image reprinted by [71].

3.3. Adhesins

Adhesins are a large group of bacterial cell-surface proteins that enable a strong interaction with epithelial cells. The adherence of pathogens to mucosal epithelial cells is considered to be the first important step in pathogenesis onset in the stomach. Bacterial adherence to the gastric mucosa is crucially important for protection from acidic pH, mucus, and exfoliation [78].

H. pylori adhesins are involved in numerous processes during early and chronic phases of infection. They also contribute to the differential outcome in infected patients by triggering disease development [79]. *H. pylori* adhesive factors belong to the largest outer membrane protein (OMP) family of the

bacterium, namely, the Hop family [80]. The Hop family contains the most well-known adhesins of *H. pylori* like BabA, BabB, SabA. Additionally, there are other proteins, such as NAP and Hsp, which have been implicated in cell adhesion and mediate the tropism of H. pylori to the gastric tissue [81].

3.3.1. Blood group binding adhesin (BabA and BabB)

Three *H. pylori* bab allelic types have been identified, including babA1, babA2 and babB. The molecular mass of the BabA protein is nearly 78 kDa, encoded by babA2. The babA1 and babA2 coding sequences are highly similar, but the translational starting codon is lacking in babA1 [71]. *H. pylori* employees BabA to bind to fucosylated Lewis B blood-group antigen (Leb) expressed on host gastric epithelium cells, when *H. pylori* initially infects the human stomach [82]. BabA and BabB are nearly identical in their 5′ and 3′ regions, with most of their sequence divergence being in their mid-regions. Importantly, the middle region of the BabA sequence determines the adhesion ability of BabA. The function of BabB is still unclear, but the expression of BabB was associated with increased gastric histologic lesions in patients [72]. To study the dynamics of Leb adherence during infection paired H. pylori isolates obtained sequentially from chronically infected patients were analysed [73]. The results showed that a complete loss or significant reduction of Leb binding was observed in strains from 5 out of 23 individuals, indicating that the BabA-Leb binding phenotype is quite stable during chronic human infection. Sequence comparisons revealed that most amino acid changes were found in the putative N-terminal extracellular adhesion domain [73]. This recombination mediates dynamic changes in adherence properties, which suggests that it contributes to the persistence and adaption of H. pylori in ever-changing gastric environments.

3.3.2. Sialic acid-binding adhesin (SabA)

SabA binds to sialylated structures such as the sialyl-Lewis X/A (sLex/a) antigens also found on mucins and epithelial cells [74]. This suggests that SabA adhesin plays a critical role to assist H. pylori to adhere to and colonize

the gastric epithelium cells of patients with gastritis [75]. Especially when lacking gastric Leb expression, Lex and Lea were closely related to H. pylori colonization. The sabB gene is homologous to sabA, but appears not to be involved in sLex binding. Therefore, the function of sabB in bacterial adhesion and pathogenesis is worth investigating. SabA is transcriptionally repressed by the acid sensitive ArsRS regulonand is therefore down regulated during acidic conditions [76]. In addition to this direct regulation by transcription factors, expression of the sabA gene is regulated through slipped strand mispairing through length variation in the poly(T) repeat tract that lies upstream of the promoter element of sabA (Berg et al., 2014). Additionally, the gene contains cysteine-thymidine (CT) repeats in the 5' part of the coding sequence, causing on/off phase variation. These dynamic adaptations may allow H. pylori to specialise for individual host variation in mucosal glycoprotein sialylation during persistent infection. Recently, the crystal structure of the extracellular parts of the SabA protein was resolved. The structure is dominated by alpha helices and resembles a club with a handle and a head. The head part contains the ligand-binding cavity, which is constrained by two highly conserved disulphide bridges. One of these pairs of conserved cysteines is also present in BabA, but the major part of the ligand-binding region is not homologous between the two proteins [78].

3.3.3. Neutrophil activating protein (NAP)

Neutrophil activating protein (NAP or HP-NAP) was first identified to stimulate high production of oxygen radicals from neutrophils, leading to damage of local tissues, and promote neutrophil adhesion to endothelial cells during *H. pylori* infection [83]. NAP upregulates the expression of β2 integrins in neutrophils and monocytes [80], and also is highly associated with the hallmark of chronic gastritis, and infiltration of neutrophils and mononuclear cells into the gastric mucosal barrier, caused by *H. pylori* infection. The glycosphingolipids expressed on the neutrophil surface serve as a major receptor to interact with the NAP expressed on bacterial surface [81]. Moreover, NAP is supposed to

facilitate SabA-mediated binding of sialyl-Lewis antigens on the membrane of the host cell [82]. To study *H. pylori* colonization in vivo Wang et al. [83] infected mice with both *H. pylori* napA mutant strain and wild-type strain, compare to the mice infected with wild-type strain the degree of survival was lower than the mice infected with napA mutant. However, for pathological outcomes related to *H. pylori* infection only one report showed that in sera from patients with GA the level of NAP-specific Ab was remarkably higher than that from infected patients with AIG [84]. No other studies have shown the direct involvement of NAP with gastric inflammation among patients *H. pylori* positive.

3.3.4. Heat shock protein 60 (Hsp)

Heat shock proteins (Hsp) are a highly conserved protein family found in prokaryotes and eukaryotes. Hsp are induced by several factors like a variety temperature, pH change, ischemia, and microbial infection [84]. *H. pylori* produces mainly two heat shock proteins, GroES-like HspA (Hsp10), and GroEL-like HspB (Hsp60). The high expression of Hsp60 at low pH, which interacts with the receptor-like sulfatide (sulfoglycolipid), indicates that the acid stress may change the specificity of *H. pylori* to receptors [85]. Hsp60 induces activation of NF-κB via TLR2 and the mitogen-activated protein kinase pathway, and thereby induces human monocytes to secrete IL-8 [86]. Moreover, anti-Hsp60 antibodies are consistently detected in *H. pylori*-infected patients, and these antibodies are associated with the progression of gastritis or gastric cancer [85].

3.3.5. Other adhesins

There are several others know adhesins used by *H. pylori* to adapt to different hosts/tissues, including AlpA, AlpB, HopZ, OipA and HopQ, all of them belong to the major Hop family of OMPs. The adherence-associated lipoproteins AlpA (HopC) and AlpB (HopB) are highly related and can be found in basically all *H. pylori* isolates. They are transcribed from the same operon, and their loss

impairs bacterial binding to the apical side of human gastric tissue sections, and in animal models. Unlike other OMPs, they are not subjected to phase variation but are expressed in virtually all clinical isolates [87]. The ligand of AlpA and AlpB has not yet been determined, although extracellular matrix laminins have been proposed as possible candidates, and the understanding of their role in human infection is still incomplete. The outer inflammatory protein A OipA (HopH) has been proposed to amplify IL-8 secretion and activate β-catenin, in parallel to cagPAI. OipA is phase-variable and can be switched "on" and "off" by slipped strand mispairing during chromosomal replication [88]. OipA expression status is associated with the presence of cagPAI and VacA s1m1 alleles in western isolates, and it has therefore been difficult to assess the separate influence of OipA on clinical manifestations [19]. Like AlpA/AlpB, the host surface receptor/interaction partner of OipA has not yet been identified. Other Hop proteins that have been implicated in adhesion are HopZ and HopQ. HopZ has been shown to be involved in the early phase of colonisation and is regulated by phase variation of CT repeats in the region encoding for the signal sequence. HopQ has also been implicated in binding to epithelial cells, but the binding partner has not yet been identified [89].

3.4. Virulence factors

Several *Helicobacter pylori* genes have been shown to modulate virulence of bacterial isolates. *H. pylori* yields various virulence factors that may dysregulate host intracellular signaling pathways and decrease the threshold for neoplastic transformation. Among all the virulence factors, *cagA* (cytotoxin-associated gene A) and its pathogenicity island (*cag* PAI) *vacA* (vacuolating cytotoxin A) are the most relevant.

3.4.1. cag Pathogenicity Island (*Cag*PAI) and CagA

Cag pathogenicity island (*cag*PAI) is a genetic element of exogenous origin that has been inserted by horizontal transfer into the glutamate racemase gene and is made up of 27-31 genes, with a total size of approximately 37 kb [90].

These genes encode structural components of a type four secretion system (T4SS) and an effector protein, the cytotoxicity associated virulence factor CagA. The T4SS pilus is used to inject CagA into the host cells, where it interferes with a series of host proteins and signalling pathways. The presence of this pathogenicity island highly varies between different strains: almost all East Asian isolates but only approximately 60-70% of Western isolates are cagPAI positive [91]. The actions of CagA can be divided into phosphorylation dependent inside the host cell, and phosphorylation-independent. The crucial region for CagA phosphorylation is represented by EPIYA (glutamic acid-proline-isoleucine-tyrosine-alanine) motifs in the C-terminal region. The tyrosines of these motifs can be phosphorylated by host Src-family kinases, and show a variability that has been linked to strain virulence and geographical origin [90]. On the basis of sequences flanking the EPIYA motifs, 4 different segments, termed EPIYA-A, -B, -C and -D, have been described (Figure 6).

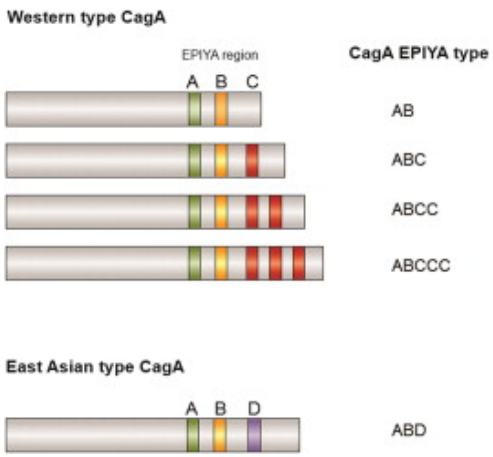


Figure 6 - Schematic Representation of the Western type CagA and the East Asian type CagA. Image reprinted by [92].

Most CagA positive strains have the A and B segments, while EPIYA-C is characteristic only of European strains and it is thus termed Western CagA. EPIYA-D is specific of East Asian strains CagA and consequently it is known as East Asian CagA. The East Asian CagA has been shown to have more

oncogenic potential than Western CagA, but Western CagA can also vary in the number of C regions, where more repeats are correlated to greater oncogenic potential. The effects of CagA on host cells include dysregulation of cell-cell adhesion and loss of polarisation in the epithelial cell, cellular elongation that resembles epithelial to mesenchymal transition (EMT), and the activation of the Ras-Erk cascade. The latter increases interleukin 8 (IL-8) release and the activation of NFκB, which leads to an increase in tumour necrosis factor alpha (TNF-α) and cyclooxygenase 2 (COX-2) pathway, among other effects [27]. Within the cell, phosphorylated CagA activates the SHP-2 phosphatase and the MAPK-ERK signalling cascade, causing effects resembling those induced by growth factor signalling. Reported phosphorylation-independent effects include the interaction with E-cadherin, the c-met receptor, and phospholipase C, which disrupts cell-cell junctions and induces loss of cell polarity, and pro-inflammatory and mitogenic effects [19].

3.4.2. The vacuolating cytotoxin A (VacA)

Along with CagA, the vacuolating cytotoxin (VacA) is the best-studied virulence factor. VacA is a toxin that forms hexameric, anion-selective pores through lipid bilayers such as the cytoplasmic and organelle membranes and, as the name implies, induces vacuolisation in the host epithelial cells [92]. There is significant variation among strains in their capacity to induce cell vacuolization [93]. This variation is attributed to the genetic structural diversity of the *vacA* gene that can assume different polymorphic rearrangements (Figure 7) [94].

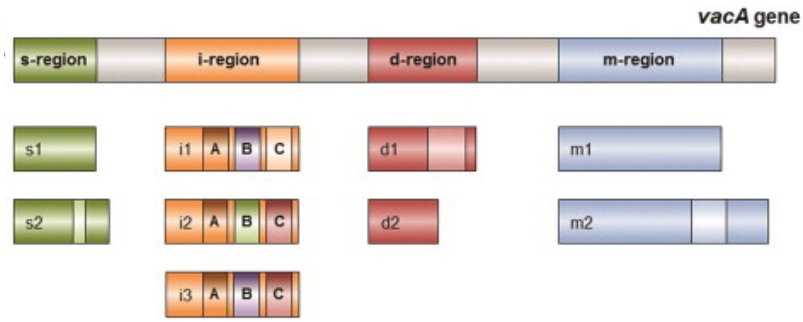


Figure 7 - Schematic representation of *H. pylori* vacA gene showing the s-, i-, d- and m-polymorphic regions. Image reprinted from [92].

The initial studies on *vacA* detected two main polymorphic regions, the signal (s)- and the middle (m)- regions. The s-region encodes a different signal peptide and assumes two forms, s1 or s2, s1 form is active while s2 form encodes a different signal peptide cleavage site which results in a short N-terminal part that display attenuated vacuolating activity. The m-region encodes the VacA binding site to host cells, and like s-region it can have two isoforms: the m1 and the m2 forms; m1 is the more effective of the two [92]. More recently, new regions of variations were found in *vacA*, the intermediate (i)- and deletion (d)-regions [92]. Like signal (s)- and middle (m)- regions they have two main polymorphic forms (i1,i2 and d1,d2) [92]. the mosaic combination of these forms discriminate among *H.pylori* strains vacuolating and strains non-vacuolating, for example the s2 type of VacA has been shown to be virtually nontoxic [95], while strains with the s1/m1/i1 genotype have been shown to be associated with a higher risk of advanced disease and are found in patients with higher gastric cancer risk. However, this genotype commonly coincides with cagA positivity, making the distinct contribution of the two virulence factors in these strains hard to be evaluated [92]. VacA can be transferred to host epithelial cells either by secretion or by contact-dependent transfer and is cleaved during its transport through the bacterial outer membrane [92]. The precise mechanism of VacA secretion and entry to the host cell is still controversial but both epithelial derived growth factor receptor (EGFR), the receptor tyrosine phosphatase alpha and beta (RPTPa/RPTBb), and sphingomyelin in lipid rafts have been shown to act as interaction partners [87]. Apart from its capacity to induce vacuolisation by forming pores in endosomes, VacA can also induce apoptosis in epithelial cells and lymphocytes by interfering with membrane trafficking, leading to loss of mitochondrial membrane potential, mitochondrial instability and the subsequent release of cytochrome C [96]. VacA interacts with epithelial cells as well as with immune cells, including B- cells, T-cells and phagocytes. In phagocytes, VacA can inhibit processing and presentation of antigen peptides

to T-cells; however, VacA can also interfere with T-cell function directly by blocking antigen-dependent proliferation or by inhibiting the activation of nuclear factor of activated T-cells (NFAT) [69].

3.4.3. Other virulence factors

Aside from VacA and the Cag pathogenicity island, several other factors have been proven to play a role in the virulence and severity of *H. pylori* infection. High temperature requirement A (HtrA) is a serine protease that is well conserved in gram-negative bacteria and has been described to increase the viability of the organism under stressful conditions. H. pylori HtrA is secreted into the extracellular space in its active form where it can cleave E-cadherin, a tumour suppressor commonly lost in several cancer settings, that is involved in cell-cell adhesion [94]. The extent to which the cleavage of E- cadherin affects E-cadherin signalling and function is not yet known, but this cleavage has also been shown to allow for bacterial entry into the intercellular space by disrupting the adherence junctions [72]. Gamma-glutamyl transpeptidase (GGT) is a virulence factor that has gained increased attention in recent years. GGT is an enzyme that converts glutamine into glutamate and ammonia, and glutathione into glutamate and cysteinylglycine. This has been shown to lead to glutamine and glutathione consumption in the host cells, which interferes with the oxidative capacity of the cell, resulting in the production of ammonia and ROS. These products affect many central cell functions, inducing cell-cycle arrest, apoptosis, and necrosis in gastric epithelial cells. GGT can also induce immune tolerance by influencing dendritic cell differentiation and inhibition of T cell-mediated immunity, affecting the efficiency of the immune response towards *H. pylori* [97]. Duodenal ulcer promoting gene A (DupA) is another virulence factor that was identified based on its positive relationship to duodenal ulcer and inverse relationship to gastric cancer [98]. DupA has been suggested to form a Type IV secretion systems (T4SS) with *vir* genes that make up the *dupA* gene cluster. The presence of a complete *dupA* cluster was

found to significantly increase the risk of duodenal ulcer compared to *H. pylori* infection with an incomplete *dupA* cluster or without the *dupA* gene [99].

3.6. Diagnosis of *H. pylori* infection

A number of tests with varying accuracy and sensitivity have been developed to diagnose *H.pylori* infection. These tests could be separated into invasive and non-invasive approaches depending on whether an endoscopy is needed. The choice of the test depends on factors such as age, clinical situations, availability, cost and prevalence of infection in the population. The numerous tests used to diagnose *H. pylori* infection in adults may also be used in children but some non-invasive tests like urea breath test (UBT) and serology have limitations because at age below 6 years, the UBT can lead to nonspecific results, and serology may not be enough sensitive [100]. Non-invasive tests can identify *H. pylori* in a sample, but do not reveal the amount of tissue damage, on the contrary invasive approaches not only diagnoses the amount of damage in the stomach but also allows the clinician to directly sample the tissue to test it for the presence of *H. pylori* infection [101].

3.6.1. Invasive tests

Invasive procedures available for detection of *H. pylori* are endoscopy, histology, culture, rapid urease test (RUT) and polymerase chain reaction (PCR) [102]. The procedures are rather expensive and laborious and scarcely applicable to wide samplesets. Histopathology staining remains one of the best methods for detecting *H. pylori* infection and together with endoscopy, it provides important information on the pathological condition of the stomach. Endoscopy is expensive, unpleasant for the patients and requires highly specialized operator to be used as a routine diagnostic test. Bacterial culture is the gold standard for diagnosis of *H. pylori* infection, because at the same time it is possible to test the antibiotic susceptibility [101]. The clinical application of bacterial culture is limited because *H. pylori* is difficult to cultivate on culture media, costly, and compared to other methods time consuming, for this reason

the use of this procedure is limited to research and epidemiologic purpose [103]. Measuring urease production from biopsy specimen can be accomplished by rapid urease test (RUT). A small biopsy sample is placed in a vial containing the urease test solution and incubated overnight, the sample is observed for a change in a color, pink indicates positive result. This test has the advantages of being more rapid than histology or culture, and has a high sensitivity, but false negative urease tests may occur if patient has taken antibiotics or compounds containing bismuth [102]. Molecular methods like PCR have been used successfully to detect *H. pylori* DNA in gastric tissues by amplifying genes such us urease gene, adhesion genes and 16S rRNA gene. The 16S rRNA is one of the specific targets for PCR diagnosis of *H. pylori* infection [104]. Urease is another specific gene for detecting *H. pylori* infection and positive amplification can be considered as a direct evidence of the presence of the pathogen [105]. The disadvantages of PCR as a routine test are that it is a technically demanding and expensive test compared to culture and the rapid urease tests, and it requires special laboratory conditions. The technique is highly sensitive, but it is subject to false-positive results due to possible environmental contamination. A positive result obtained by PCR does not indicate active infection but could also detect the presence of DNA of dead bacteria.

3.6.2. Non-invasive tests

Non-invasive procedures available for detection of *H. pylori* are urea breath test, fecal antigen test and serological tests. The first two are considered active tests because can detect the presence of *H. pylori* and provide an evidence of an active infection, while serological tests are considered passive tests because provide the evidence of exposure to *H. pylori* and are usually not indicated in cases of eradication [106]. These tests are based on the detection of antibodies to *H. pylori*. The urea breath test is the preferred non invasive choice for *H. pylori* diagnosis before and after treatment. This test exploits the fact that *H. pylori* metabolizes urease (see above), so for this test patients

ingest a small amount of urea that is radioactively labeled (^{13}C), bacterial urease split off radio-labeled carbon dioxide, which is detect in the breath. In patients not infected by *H. pylori* radio-labeled carbon dioxide will not be detected in their breath and the urea ingested will be completely eliminated within urine. This test is considered the "gold standard" for in vivo detection of *H. pylori* infection and is also a good measure for the follow-up of the eradication therapy [107]. The disadvantage of this technique is that the detection of ^{13}C requires a mass spectrometer, which may not be easily available in clinics. Fecal antigen test is a relatively new diagnostic procedure that uses an enzyme linked immunosorbent assay to detect the presence of *H. pylori* stool antigen (HpSA) in the stool with the use of polyclonal anti-*H. pylori* antibody. This test has a high sensitivity and specificity but its disadvantage is the increasing of false-negative results for the stool antigen test during proton pump inhibitor (PPI) treatment [108]. Serological tests measure circulating IgG, IgM and IgA antibodies in a patient's serum and have sensitivity and specificity of 94% and compared to invasive techniques are extremely cheap. The heterogeneity of *H. pylori* strains has been well documented and is characterized by a considerable variation in the prevalence of specific strains, especially from different geographical areas [109], thus the success of a serology test depends on the use of antigens that are present in *H. pylori* strains from a given population. Serological tests have several advantages, namely they are non-invasive and they do not produce false negative results in patients receiving treatment (proton pump inhibitors and antibiotics) or presenting acute bleeding [110], for this reason several *H. pylori* immunogenic proteins have been presented as candidates to detect infection, such as the FlidD protein, multiple recombinant (CagA, VacA, GroEL, gGT, HcpC and UreA) proteins, CagA or Omp18 [111]. All current diagnostic procedures have its own advantages, disadvantages and limitations, between non-invasive methods and especially between serology tests there's the need to improve the diagnostic yield of *H. pylori* infection detection in specific clinical conditions [100]. At the same time, for wider applications, it would be

fundamental to identify novel biomarkers that could allow the diagnosis and/or prognosis of the progression of infection towards specific different pathological outcomes such ad Gastric Cancer, Autoimmune Gastritis and MALT lymphoma [19].

4. PHAGE-DISPLAY TECHNOLOGY

Biomarkers are indispensable tools for diagnostics and play roles in molecular medicine in the identification, prevention, and diagnosis being also very important for practical clinical disease management. Nowdays there is an increasing awareness of the need for new biomarkers that could contribute to diagnosis and prognosis, various protein detection techniques have been applied in biomarker discovery. SDS-PAGE and Western blotting are the most basic and traditional techniques, these techniques have been combined with high-throughput techniques, indeed different experimental methods, such as two-dimensional electrophoresis (2-DE), protein or peptide microarray, high-performance liquid chromatography (HPLC) and mass spectrometry (MS) have been employed for the discovery of immunogenic proteins [112]. However, the identification of the antigens that trigger specific systemic antibody responses is still difficult and limits the applications of these technologies. To serve this purpose phage display technology was proposed in 1985 by George P. Smith to identify polypeptides with specific binding activity and subsequently evolved with many versatile applications [113]. The phage-display technology was recently used to identify novel potential biomarkers from phage genomic libraries produced from different Mycoplasma species [114], Salmonella Typhimurium [115] and Neisseria gonorrhoeae [116]. However, the ability of phage display libraries to be enriched for epitopes of monoclonal antibodies has long been demonstrated [117]. Phages contain the genetic information of the proteins displayed on their capsids, thus genotype and phenotype are linked together, in this way this technology allows the direct determination of the genetic information (ORF/ORF fragment) from which the protein/antigen derives [118].

4.1 ORF PHAGE-DISPLAY

A link between the genotype and phenotype of the phage is provided by inserting DNA inside of the phage that encode for peptides which are displayed on the phage surface [105]. Each phage expresses on his capsid a unique peptide but the whole library includes a large number of different displayed peptides that afterwards will be used by affinity selection in order to identify specific ligands for different target [106]. The affinity selection technique, called biopanning, is the procedure of selecting specific binding partners from phage display libraries [107]. One interesting advantage of ORF Phage-Display is that this technique does not need prior knowledge of the identity and attributes of the target, furthermore once generated, libraries can reuse for an unlimited number of screenings [108]. This technology has applications in drug discovery [109], antibody engineering and epitope mapping [110], gene/drug delivery [111], enzyme technology, organ targeting [112], bioimaging and biosensing [113], study of interactions protein–DNA and protein–protein [114], antiviral research [115].

4.1.1 Phage display and ORF selection

Phage display technology application was expanded to include the display of antibodies and many other proteins [116]. As mentioned in the previous paragraph cDNA phage display has been successfully and widely applied both to identify antibody epitopes or binding partners [108,117] but the large number of non-functional clones present made the cDNA libraries inefficient and difficult to use. Although this high rate of non-functional clones may be tolerable when starting to work with DNA from a single gene or even a small genome but become unfunctional if employing more complex DNA sources.

To solve this problem the first strategy applied was to use λ-based vectors for cDNA display, despite this vector C-terminal intracellular vectors increase the likelihood that ORFs will be displayed, they do not provide any selective pressure for ORFs suggesting then the need for a further selective step to filter DNA fragments encoding ORFs [118]. Different attempts have been made to

filter DNA fragments, the first strategy was proposed in the 1992 by Seehaus [119], the selection was made with an antibiotic resistant gene to remove deletion mutants from antibody library with a plasmid in which antibody library was cloned upstream of a β–lactamase gene, thus only those antibody genes in frame were capable to allow ampicillin resistance by the creation of an antibody–lactamase fusion protein, whereas those that contained deletions or frameshifts were not conferring ampicillin resistance. Zacchi et al in 2003 [120] applied a similar strategy with a phagemid, they inserted the cDNA fragment followed by the gene of β–lactamase delimited by two homologous lox sites. To obtain efficient display of foreign polypeptides at N-terminus, the selection step with ampicillin , β–lactamase selection was removed by Cre recombinase-mediated recombination [120]. Faix et al [104] further improved this approach, they first selected fragments with a β–lactamase gene in a plasmid, after this fragments were extracted from ampicillin-resistant plasmids and re-cloned into a phagemid and afterwards rescued by hyperphage system. D'Angelo e al. [121] in order to demonstrate the feasibility of filtering method with antibiotic resistant gene have applied β–lactamase filtering vector at whole genome level in Clostridium thermocellum for domain-based functional annotation purposes. They demonstrated that domainome libraries are easy to generated by applying β–lactamase filtering vector to randomly fragmented bacterial gDNA libraries and that once a library is generated, it can be used for an unlimited number of screenings.

4.1.2 Comparison of Phage display with other techniques for interactome analysis

To elucidate protein physical association and elucidate protein functions different technologies have been widely used such as protein affinity purification coupled with 1D or 2D gel electrophoresis and mass spectrometry (AP-MS) or yeast two-hybrid (Y2H to identify protein-protein interactions [119]. Both AP-MS and Y2H have been applied for the recognition of binding peptides for specific protein bait as well as for domain-scale interactome

mapping. AP-MS has been well demonstrated for efficient mapping of interactome for yeast, *E.coli* [120] and human [121] .

AP-MS technology is more flexible than Y2H but has several limitations, it has a low sensitivity, consequently, less abundant proteins my not be detected (Table 2). Although Y2H and AP-MS have been widely used to identify protein-protein interactions, they are limited to high cost, technical complexity, instruments requirements, Li and Caberoy. [122] estimated the error rate that is more or less 50% for AP-MS (15% with tandem affinity purification), more or less 48% for Y2H while for ORF phage display is 29% (Table 2). One of the advantages of ORF phage display is that can significantly reduce the time required to identify a peptide with binding characteristics. Phage has a robust growth rate and another advantage is the versatility, Y2H is applicable only to protein-protein interactions studies. Several studies have shown that phage display is applicable to proteins, antibodies, peptides, multi-molecular complexes and affinity selection may be performed in vivo and in vitro [123]. One disadvantage of bacterium-based ORF phage display is that proteins displayed on phage surface lack appropriate post-translational modifications but this approach facilitates the identification of binding compounds with high affinity in a large scale with an efficiency and sensitivity comparable to Y2H and AP-MS [119]. Other limiting factors for these experiments are the availability of expression constructs, data analysis and the downstream validations.

	ORF Phage Display	Y2H	MS-based approach
Capacity of the technologies			
Protein-protein interaction	Yes	Yes	Yes
Protein-polysaccharide interaction	Yes	No	Yes
Protein-lipid interaction	Yes	No	Yes
Protein-antibody interaction	Yes	No	Yes
Protein-DNA interaction	Yes	Yes	Yes
Protein-RNA interaction	Yes	Yes	Yes
Protein-virus interaction	Yes	No	No
Protein-cell interaction	Yes	No	No
Protein-tissue interaction	Yes	No	No
In vivo selection	Yes	No	No

Table 2 - Comparison of different technologies for functional proteomics. Table reprinted [124] with modifications.

4.2 Phage Display evolution: INTERACTOME-SEQUENCING

Nowadays the advent of high-throughput sequencing (HTS) techniques is transforming multiple research fields as well as the analysis of the intrinsic heterogeneity of phage libraries [125]. The complex high throughput data provided by deep sequencing enables the analysis of highly complex samples such as phage display libraries and holds the potential to circumvent the traditional laborious picking and testing of individual phage rescued clones [126]. In the recent times there have been significant attempts in coupling HTS with phage display, in 2009 Dias-Neto et al. [127] proposed the use of NGS coupled with RT-PCR to improve phage analysis of the library inserts encoding phage-displayed variants.

Pyrosequencing was employed for deep-sequencing amplicons obtained from phage ssDNA libraries, recovered directly from four human tissue biopsies, using primers flanking the library gragment within the fusion by Di Niro et al (2010) [129]. Thus combining phage display and NGS they increased of two orders the number of affinity-selected clones. Authors screened an ORF-

filtered cDNA phage library, obtained from several human cell lines mRNA , to identify proteins in contact with tissue transglutaminase 2 (tTG2), an enzyme involved in different clinical outcomes. This approach has driven the characterization of a “landscape” of binding variants from the phage display library. Another relevant application of NGS coupled with phage display was shown by Gourlay et al. [128] , these authors generated a ORF-filtering library using the *Burkholderia pseudomallei* genome in order to develop a high-throughput tool to identify soluble protein domains from the entire protein repertoire of the bacterium. With this approach they demonstrated that the extension of the filtering power of β–lactamase based filtering selection strategy to identify outer membrane protein domains is very useful in the context of host-pathogen interaction studies. Between the 1279 ORFs represented in the library the focus fell on a potential antigen, BPSL2063, which was recognized by positive antibodies from patients that tested positive for *B. pseudomallei*. To identify the domains of this antigen a second single-gene domain filtering library was constructed and after a specific bioinformatic analysis two domains were revealed at N- and C- termini, respectively; afterwards they were produced in soluble form and successfully crystallized.

In conclusion the ORF-filtering library approach allows the characterization of the whole ORF-eome and Domain-ome of any intronless genome, favoring the selection of ORFs encoding for proteins that are potentially exposed on the cell surface or involved in transport mechanisms. It allows the identification of outer membrane localized protein domains supporting the identification of antigens.

The Domain-filtering library approach applied to single gene/proteins (very long more than 1000 aa and transmembrane) allows the selection of protein domain boundaries, and can accelerate and improve the steps leading to soluble protein production for crystallogenesis.

4.3 ORF-filtering library analysis

The output generated by ORF-filtering libraries sequencing can't be analyzed with the existing bioinformatics resources, as a matter of facts the kind of data do not resemble other sequencing data like ChIP-seq or RNA-Seq for which there are many free tools and softwares that are constantly released or updated. Phage display technology coupled to sequencing was introduced only in 2010 by Di Niro et al [129] and from a bioinformatic point of view the analysis of this kind of data was always performed by adapting tools designed for other purposes. In two recent works from D'Angelo et al. [130] and Gourlay and et al. [128] the analysis for the identification of specific domains/antigens was performed through NGS-TreX developed by Boria et al. [131]. NGS-TreX is a free available web-tool designed for the analysis of RNA-Seq data and it is not specific for the analysis of Interactome-Seq data. In literature, up to date, there aren't dedicated pipelines able to reveal enriched domains from dataset generated by Interactome-Seq technology. Therefore, at the beginning of my thesis work, there was the urgent need to develop a robust workflow that should be able to manage Interactome-Seq data, both with fixed or user-defined parameters, and to perform specific testing for the identification of enriched domains/antigens. At the same time the visualization and the sharing of the features identified was not possible so I decided not only to develop a new pipeline of analysis but to implement it into a web-tool easy to use, user friendly and available to the whole scientific community interested in analysing and consulting Interactome-Seq data.

5. AIMS OF THESIS

The aim of this thesis was the identification of *H. pylori* epitopes responsible for host immuno-response modulation through: a discovery-driven approach that couples “phage display” and deep sequencing (interactome-sequencing) and the development of a specific webtool for interactome-sequencing data analysis.

To this purpose my thesis work was focused on the:

- Identification of novel biomarkers to provide a “disease signature” by screening genomic DNA libraries, created from *H. pylori* genome, directly with sera from infected patients with different pathological outcome.

- Development of a new pipeline for interactome-sequencing data analysis.

- Implementation of a Webtool for functional analysis and visualization of interactome-sequencing data.

6. MATERIALS AND METHODS

6.1. Bacterial strains

Bacterial strains HP-26695 and HP-B128 used in this study were isolated at the Carregi Hospital and gently provided by Dott. Mario Milco D'Elios.

6.2. H. pylori *genomic ORF-filtering library construction and sequencing*

To identify all the potential soluble domains on a genomic scale we applied our strategy, based on the construction of an ORFeome library from two *H. pylori* genomes. The procedure initially involves the fragmentation of a whole genome into DNA fragments of 200-800 bp (D'Angelo et al., 2011, Heger & Holm, 2003) with the purpose to create a library of fragments coding for potential domains (or parts of). DNA fragments encoding well-folded protein domains, fused upstream of ß-lactamase, allow the reporter enzyme to fold correctly and allow bacteria to survive the selective pressure posed by the antibiotic. Genomic DNA from *H. Pylori* strains HP-26695 and HP-B128 was fragmented by ultra-sonication (Covaris) (Duty Cycle 10%; Intensity 5.0; Cycles per burst 200; Duration 2X60 seconds, total 120 seconds; Mode frequency sweeping; temperature 6°C) to obtain fragments in a length range between 200 and 800 bp (about 65–250 aa). Fragments were collected, blunt-ended using the Quick Blunting Kit (New England Biosciences) and cloned into the pFilter vector using EcoRV cloning site between a pelB leader sequence and the mature β-lactamase gene [132]. After ligation the *H. pylori* genomic DNA library was electroporated into DH5a F' cells and plated on chloramphenicol plates (34 mg/ml), supplemented with ampicillin (25 µg/mL). A small dilution of the library was also plated and grown in parallel on both chloramphenicol and chloramphenicol plus ampicillin resistance in order to calculate the size of the library and the efficiency of the filtering.

10^7 bacterial cells were plated and after overnight growth approximately 1% cells survived at the highest ampicillin concentration, yielding one library with an estimated size of $1x10^6$ clones.

Bacteria were harvested, mixed, resuspended in 20% sterile glycerol and stored in small aliquots at –80°C. One aliquot was used to extract plasmid DNA used for the preparation of the phage-library and for sequencing analysis of the inserts. To this aim the HP genomic DNA inserts were recovered from the pFILTER vector, by amplification with specific primers. The primers used to rescue the inserts were linked at their 5' end to Illumina adaptors sequences, thus allowing the direct sequencing of the HP genomic DNA inserts by using the MiSeq Illumina sequencer (Figure 8). The libraries were indexed, pooled together and sequenced on a MiSeq Illumina sequencer; 250nt paired end reads were generated

Forward primer:

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG*GCAGCAAGCGGCGCG CATGCC*;

Reverse-primer:

GTCTCGTGGGCTCGGAGATGTGTATAAGAGACA*GGGGATTGGTTTGCC GCTAGC*;

In bold are indicated the Illumina adaptors,in italic are indicated the specific primers.

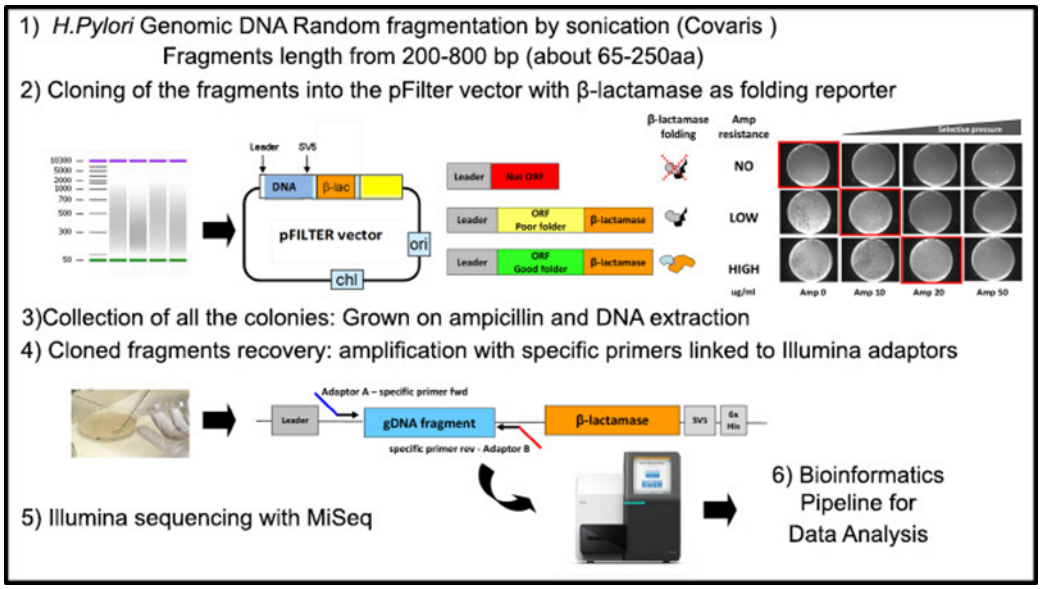


Figure 8- Schematic Overview of the main steps for the construction of the *H.pylori* genomic ORF-filtering library: 1) Random fragmentation of genomic DNA 2) gDNA fragment cloning into the pFILTER vector and filtering using β-lactamase as a folding reporter 3) Collection of all colonies and DNA extraction 4) gDNA fragment recovery by amplification using specific primers linked to adaptors for sequencing 5) deep sequencing 6) data analysis.

6.3. H. pylori Phage-library construction

In order to create a phage-display library, filtered DNA fragments were collect and cloned inside phagemid vector. Phagemid enables the expression of the foreign DNA, purposely introduced in the vector in such a way that it is expressed in conjunction with a phage protein, as a fusion protein, for display on the phage surface. This instrument is necessary to perform the selections with sera, it allows to couple genotype and phenotype in the same phage (Figure 9). For the *H. pylori* phage-library preparation the filtered HP ORF were recovered from the pFILTER3 vector and cloned in the pDAN5 [133] upstream to the cDNA codifying for the g3p protein. After ligation the library was electroporated in DH5aF' competent bacterial cells and grown for 18h at 37°C on 2XTY-Agar plates supplemented with 100 µg/ml of Ampicillin. Bacteria were harvested, mixed, resuspended in 20% sterile glycerol and stored in small aliquots at -80°C. One aliquot was used to extract plasmid DNA and subjected to sequencing analysis of the inserts, performed as described above.

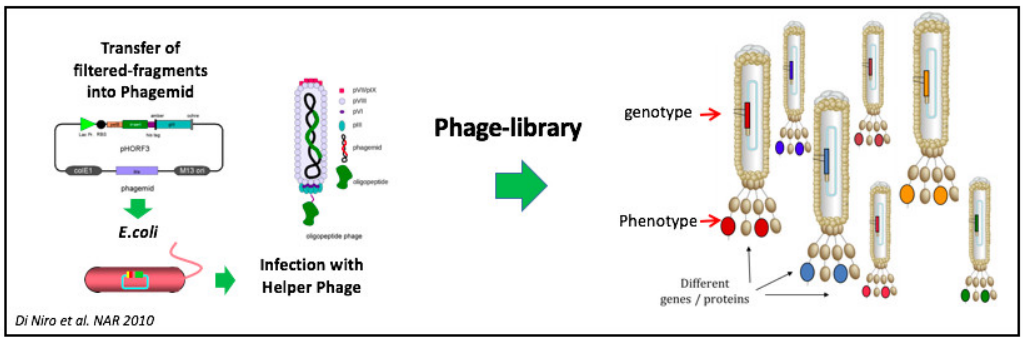


Figure 9 - Schematic Overview of the main steps for the construction of the phage library.

6.4. Sera selection

Once recloned into a phage display context, we directly used the ORF-filtering libraries to selected putative antigens by capturing the antibodies present in the sera from patients affected by *H. pylori* and presenting increasing degrees of infection: A) gastric adenocarcinoma (pool of 2 sera); B) autoimmune gastritis and pernicious anemia (pool of 3 sera); C) MALT lymphoma (pool of 2 sera); at the same time control sera from healthy patients both affected or not by *H. pylori* (D) were used to get a background control for successive normalization of the signals. The sera were grouped into different pool and each pool was independently used for three consecutive rounds of selection, in order to enrich the library for immunoreactive clones, with increasing washing and binding stringency Two cycles of selection and amplification were performed, and mini-libraries of selected phages were collected after each cycle (Table3). Test set antibodies were incubated with library phages, immune-complexes were recovered by protein A coated magnetics-beads and bound phages were eluted by standards procedures. Selected phages were used to infect *E. coli* and to produce mini-libraries from each cycle of selection. DNA inserts were recovered from each individual selected library, purified and sequenced by using the Illumina MiSeq sequencer (Figure 10)

Library description	Library name
H. pylori ORF-filtering library	Genomic
Healthy patients both affected or not by *H. pylori*	Control
Patients affected by gastric adenocarcinoma	SelectionA
Patients affected by autoimmune gastritis	SelectionB
Patients affected by MALT lymphoma	SelectionC

Table 3 - Library description and library name assigned**.**

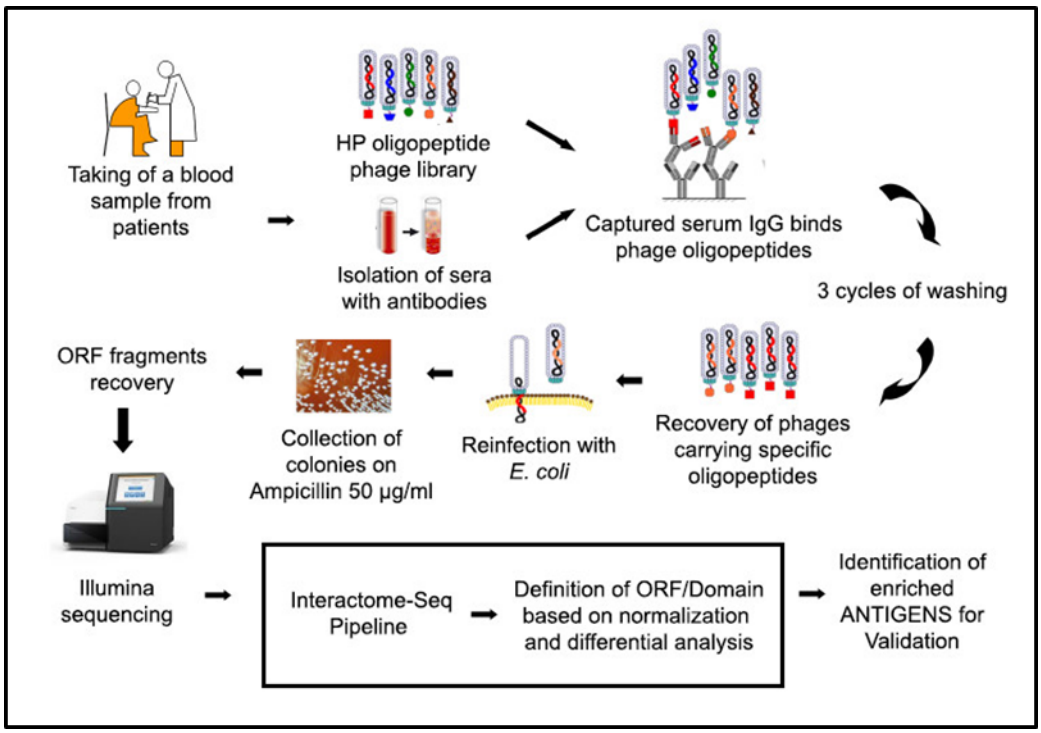


Figure 10 - Schematic overview of sera selection. Phage library were used for selection on patient serum antibodies immobilized on magnetic beads, three cycles of washes were followed by phage amplification on bacteria. *E.coli* cells plates supplemented with Amplicillin were deep-sequenced allowing the identification of enriched gene sequences. After data analysis, a list of putative antigens were produced and validated with an independent set of sera and assassed for their specificity.

6.6. Bioinformatics analysis and web-tools design

6.6.1. *De-Novo* Assembly of B128 strain

The assembly tool used in this study was Velvet 1.2.10 [134]. Large contigs (length >500nt) obtained were ordered and orientated with CONTIGuator version 2.7.4 [135]. Scaffolds were annotated with RAST [136].

6.6.2. Genomic comparison with other *H. pylori* strains

For comparative analysis, complete genome sequences of *H.pylori* were downloaded from NCBI database and compared using Gegenees tool (see Appendix 10.10) [137]. The tool utilizes a fragmented alignment algorithm to calculate average similarity among the compared genomes using BLASTn. The fragment size can be optimized according to the user. The tool was run with the fragment size set to 200 and a step size of 100 using BLASTn. The average similarity was calculated with a BLAST score threshold of 40% generating a heat plot matrix that was further used to deduce phylogenetic relationships exported in the form of a .nexus file. This nexus tree file was supplied as an input to SplitsTree [138] program for building an un-rooted phylogenetic tree employing Neighbor-Joining algorithm.

6.6.3. Implementation of Interactome-Seq pipeline and web-tool.

I developed a data analysis workflow called Interactome-Seq that is a combination of custom shell, AWK and python scripts. As part of the analysis several tools have been included in the pipeline, namely 1) FastQC (v. 0.11.4) for quality check and processing of fastq sequences, 2) Cutadapt (v. 1.10) [139] for trimming primers and discarding reads with low Phred quality score, 3) Bowtie2 (v. 2.2.9)[140] for alignment step, 4) Samtools (v. 0.1.19-44428cd) [141] for converting alignment in different formats, 5) Bedtools (v. v2.24.0) [142] for working with genomic intervals, 6) EdgeR (v. 3.14.0) [143] for differential signal analysis. The Interactome-Seq pipeline can be accessed

through an interactive web-based graphical user interface (http://cerbero.ba.itb.cnr.it/cricket/ITBHelicobacter).

Analysis steps are implemented and distributed inside server environment through a dispatching application and the whole execution can be launched and monitored using a common up-to-date web browser. Interactome-Seq is powered on a Ubuntu 14.04LTS Linux operating system. The Web Framework application used for development was Google AngularJS, combined with HTML5 and CSS3 standards, to enable a better user interaction The server was an Intel® Xeon® Processor, 64 GB of RAM, 9 TB (SATA) of hard disk.

6.7. ELISA validation

Helicobacter pylori ORFs obtained from the second round of selection, and resulting specific for progression towards pathological outcomes, were excised from the phagemid vector and cloned into a compatible pGEX-FLAG (D'angelo et al Clin Immunol, 2013) vector, creating for each output of selection a library for the expression of the selected ORFs as GST-fusion products. Plasmid DNA was isolated from an aliquot of these libraries and used as a template for an inverse PCR-based cloning strategy performed with two back-to-back outward specific primers designed centered on the epitope region identified by the overlapping reads. Colonies grown on Ampicillin plates were sequenced. Positive colonies were grown at 37°C until OD600nm=0,5 and induced by IPTG with a final concentration of 0,2mM for 16h at 28°C. Recombinant protein was purified with GST affinity resin (Sigma Aldrich) using standard procedures. An in-house ELISA assay was used to validate the antigenicity of the selected ORF. Briefly recombinant GST-protein was diluted in PBS to 2µg/ml and 100µl were coated in ELISA wells (Greiner), overnight at 4°C. Wells were washed with PBS and blocked with 200µl of blocking solution (PBS- 0.05% Tween20-2% milk) for 1h at 37°C. Wells were again washed for three times with PBS-0,05% Tween20 and incubated with sera samples diluted 1:500 in blocking solution, for 2h at 37°C. Extensive washes were performed with PBS-0,05% Tween20 and PBS. Secondary antibody was a goat anti-human-IgG HRP

conjugated (Sigma Aldrich) diluted 1:5000 in blocking solution, for 1h at 37°C. After extensive washing, immune-complexes were revealed with TMB and the plate read at 450nm. Samples with absorbance ≥ of the mean OD450 value obtained with control sera plus 2 standard deviations (SD) were considered positive.

7. RESULTS AND DISCUSSION

My thesis work is divided in four main parts:
1) construction and comparative analysis of two *H. pylori* genome ORF-filtering libraries; 2) bioinformatics analysis of sequencing data: development of a new data analysis pipeline and implementation of a web-tool, 3) phage-library production and analysis of the enriched domains after sera selection; 4) validation of the results.
I participated to all the parts of the work, but my personal involvement was mainly focused in the bioinformatics analysis of the sequencing data generated, on the pipeline development and on the web-tool implementation.

7.1 HP-26695 and HP-B128 ORF-filtering libraries comparison

H. pylori is known for its remarkably high level of genetic diversity creating a dynamic pool of genetic variants, this pool of genetic variants delivers a sufficient genetic diversity to allow *H. pylori* to occupy all the potential niches in the stomach (for example, antrum and corpus mucosa) [144].
Since one of the aims of my thesis work was to: identify novel biomarkers, to provide a "disease signature", by screening genomic DNA libraries created from *H. pylori* genome, directly with sera from infected patients with different pathological outcomes, first of all I wanted to be sure that the strategy here adopted and the instrument here proposed (i.e phage-library coupled with interactome-sequencing) could be considered universally valuable.
For this reason two ORF-filtering libraries were produced: one from the genome of the HP-26695 strain and another from the genome of the HP-B128 strain. Strain HP-26695 originates from patients suffering from a chronic gastritis and is considered the reference HP strain, having a complete and very well annotated genome sequence, strain HP-B128 originates from patients with gastric ulcer, its genome sequence is not complete but consists of 73 supercontigs. The strain HP-B128 is one of the strain most frequently found in infected patients and it is linked to infection progression towards gastric ulcer

and/or more serious pathologies such as gastric cancer, autoimmune gastritis and MALT lymphoma.

A comparative analysis of the genomic sequences of these two strains, and of the ORF portions that we found represented in their ORF filtering libraries, herein produced, was performed in order to understand if the *H. pylori* genetic diversity between these two strains could in some way introduce a bias into their ORF filtering libraries content and prevent their use as universal instrument for new specific diagnostic and prognostic markers discovery.

7.2. Assembly and phylogenetic analysis of HP-B128

Since the genome sequence of HP-B128 strain was not complete [145], before starting with the production of its ORF-filtering library, HP-B128 genome was re-sequenced in our laboratories. The reads obtained were then assembled and a comparative sequence analysis between HP-B128 strain new genome sequence and all the complete genome sequences of *H. pylori* strains available in public databases was performed.

In this way we estimated the level of genetic diversity that separates the HP-B128 strain from the HP-26695 reference strain to understand if the genome annotation of the reference strain could be used for the successive steps of functional analysis of the two ORF filtering libraries generated, and extended to the functional definition of the domains/antigens found enriched after sera selection.

7.2.1. HP-B128 genome assembly

The whole genome sequencing of HP-B128 isolate revealed that the chromosome size is around 1.66 Mb similar to others *H. pylori* strains. The genome also revealed a low G+C content of 38.76% which is another important characteristic of *H. pylori*. Detailed assembly metrics of the genome sequencing are reported in Table 4: we obtained a total of 61 contigs (>500nt) by assembling 180.967 reads produced with 454 GS-FLX+ platform (Roche). Afterwards 44 contigs were merged in a single FASTA file by the scaffolding

procedure, taking the HP-26695 genome as template, and a single scaffold of 1.664.509 bp length was obtained (see Appendix 10.1). The remaining 17 contigs were duplicates of parts of the 44 contigs assembled into the big scaffold, only repeated genomic regions were excluded from this scaffolding step. Thus we strongly improved the genome sequence of the HP-B128 strain previously available, that was composed of 73 contigs, and obtained a draft genome of this strain.

Sequencing Assembly of HP-B128	
Raw reads	180.967 SE
Total contigs	61
# contigs (>= 1000 bp)	34
# contigs (>= 5000 bp)	23
# contigs (>= 10000 bp)	21
# contigs (>= 25000 bp)	17
# contigs (>= 50000 bp)	9
Total contigs assembled	44
Largest contig	376433
Total length	1664509
GC (%)	38.76
N50	128427

Table 4 - Assembly metrics of HP-B8 strain. In this table are reported the number of raw reads (Raw reads), number of total contig (Total contigs), number of contig bigger that 1000,5000,10000,25000,50000 nt, total number of contigs assembled(Total contigs assembled), the most large contig assembled(Largest contig), total genome assembled (Total length), median of contig lengths (N50).

7.2.2. Analysis of HP-B128 genome sequence similarity

The HP-B128 draft genome obtained was automatically annotated by using RAST, 1711 CDSs were annotated, but their functional annotation was quite poor, about 25% of the CDSs were described as “hypothetical protein”. Thus to improve HP-B128 annotation I decided to compare its draft genome sequence to the genomic sequences of 52 strains available in the NCBI database. A similarity phylogenetic tree was constructed using Gegenees and Splitree software. The results of this phylogenetic comparison are shown in Figure 11 and the formation of distinct cluster (Green square), based on sequence similarity, between HP-B128 and HP-B8 strain can be clearly appreciated. HP-B128 draft genome showed a level of similarity of 96% with HP-B8 (NC_014256.1). The strain HP-B8 evolved from the human strain HP-B128 and was adapted to infecting gerbils, used as animal models for HP infection and gastric cancer development. This close phylogenetic association among HP-B128 and HP-B8 strain is in accordance with the comparative genome analysis of Farnbacher et al [145], in this work the authors found that about 98% of the whole genome sequence of strain B8, that they obtained, was covered by B128-supercontigs with a percentage of sequence similarity greater than 90%. They also performed a comparative analysis between the two strains revealing that more than 86% HP-26695 CDSs completely matched with the HP-B8 CDSs. At the same time it is possible to observe, from the phylogenetic tree, that even between the genome sequence of the strain HP-26695 and of the HP-B8 and HP-B128 respectively there is a very high level of similarity (80%).So after having verified the high level of similarity between our draft genome sequence of strain HP-B128 and the already published complete genome sequence of strain HP-B8 we associated the well curated genome annotation of this second strain to our draft genome. In this way we greatly improved HP-B128 CDSs functional annotation and made it comparable with the annotation of the whole HP-26695 CDSs set.

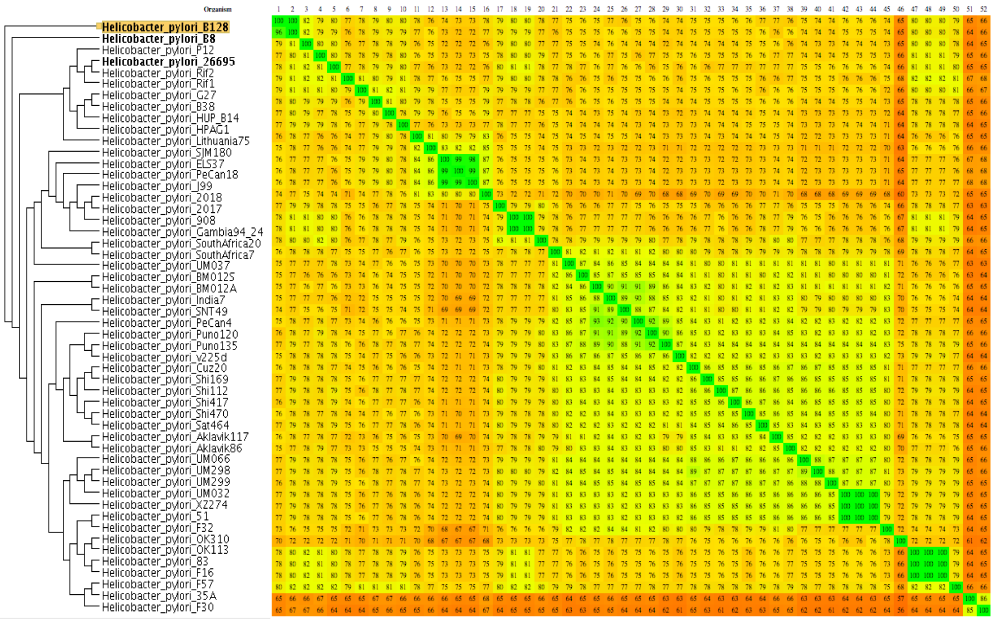


Figure 11- Gegenees heat-plot over a set of all complete *H. pylori* genomes**.** A fragmented alignment in BLASTn mode was performed with settings 200/200. The cutoff threshold for non-conserved material was 20%. A dendrogram was produced in SplitsTree 4 (using neighbor joining method) made from a distance matrix Nexus file exported from Gegenees.

7.2.3. Crossmapping comparison of HP-26695 and HP-B128 ORF-filtering libraries

As one of the main advantages of producing ORF-filtering libraries from intron-less genomes (i.e bacterial genomes) is the possibility to characterize their whole ORF-eome and/or Domain-ome, it is important first to evaluate if the ORFs/Domains filtered out by this technique can show relevant differences when ORF-filtering libraries are produced from genomes of different strains having high level of similarity, such in the case of HP-B128 and HP-26695 strains. This verification step is fundamental when the genomic ORF-filtering library, produced starting from the bacterial genome, should be used to perform selections/enrichments with sera from patients. Indeed in this case it is important to obtain a genomic ORF-filtering library that could be considered a universal instrument able to work equally with different sera regardless of the HP strain infecting patients.

To this purpose I performed a crossmapping comparison by aligning the reads obtained from the sequencing of both the genomic ORF-filtering libraries of HP-26695 and of HP-B128 against their respective genomic sequences of HP-29965 (NC_000915) and HP-B8 (NC_014256) and viceversa. Table 5 reports the metrics of all these crossmapping steps, both raw reads datasets were first trimmed then mapped with Bowtie2. In panel A, the 91,61% of the trimmed reads of HP-26695 ORF-filtering library aligned with its own reference genome while the 85,75% aligned to HP-B128 (HP-B8) sequence. In panel B the 84,7 %.of trimmed reads of HP-B128 ORF-filtering library aligned with the HP-B8 genome while again the 85,75% align with the HP-26695 genome. Considering that the cross-mapping metrics observed in particular for the HP-B128 ORF filtering library do not differ much when the reads are aligned against the HP-B8 and the HP-26695 genome sequences respectively, we can conclude that the genetic diversity that separates these two strains does not significantly affect the nature of ORFs/Domains sequences filtered out. So in the light of these results we decided to use as reference the HP-26695 genome and to use its CDSs annotation to functionally read out the results obtained after selection with sera from patients.

A) Reads mapping vs HP-26695 genome sequence		
sequencing metrics	**HP-26695 reads**	**HP-B128 reads**
Raw Reads	1.425.554	1.425.554
Reads after trimming	1.216.124	1.216.124
Mapping Reads	1.114.184	1.042.943
% of Mapping reads	91,61%	85,75%
Unmapping Reads	11.158	108.581
Multiple mapping reads	90.782	64.600
B) Reads mapping vs HP-B128(B8) genome sequence		
sequencing metrics	**HP-26695 reads**	**HP-B128 reads**
Raw Reads	1.031.956	1.031.956
Reads after trimming	954.895	954.895
Mapping Reads	746.325	802.021
% of Mapping reads	78,10%	84,70%
Unmapping Reads	158.301	92.896
Multiple mapping reads	50.269	59.978

Table 5 - Geonomic ORF filtering library crossmapping results for HP-26695 and HP-B128 strains. In the panel A are reported the mapping metrics of HP-26695 ORF-filtering library vs HP-26695 and HP-B8 strains. In panel B are reported the mapping metrics of HP-B128 ORF-filtering library vs HP-26695 and HP-B8 strains. Metrics are composed be number of Raw Reads (Raw Reads), number of Reads after trimming (Reads after trimming), number of Mapping reads (Mapping reads) and percentage (% Mapping reads), number of unmapped reads (Unmapping Reads) and number of reads that map in more than one point of genome (Multiple mapping reads).

7.3. Development of a New Bioinformatic Data Analysis Pipeline

7.3.1 The Idea behind

As previously explained in the paragraph 4.3, the output generated by ORF-filtering libraries sequencing can't be analysed with the existing bioinformatics resources, indeed this kind of data do not completely resemble other sequencing data like ChIP-Seq or RNA-Seq. Up to date, there aren't dedicated pipelines able to reveal enriched domains from dataset generated by Interactome-Seq technology. So, in order to fill this gap and to correctly analyze the Interactome sequencing data herein produced, I designed a new dedicated data analysis pipeline.

Even if Interactome-seq data stand apart from ChIP-seq and RNA-seq data, they have some common features, with these two kind of data, that should be taken in consideration.

1. One feature in common with ChIPseq analysis is essentially the step of detection/calling of enriched domains. For ChIP-Seq experiments the input DNA is mandatory to model the background noise of the experiment before the Peaks Finding step, even in Interactome-Seq experiments, background is necessary both for Genomic ORF-eome definition and for building the statistical model required to identify the enriched putative domains.
2. The features in common with RNAseq are two. The first one is the necessity of normalizing data between samples. The normalization can be done with an approach similar to that used for RNA-Seq experiments, indeed my pipeline counts the reads per kilobase per million mapped (RPKM), this normalization step, usually adopted to normalize RNA-Seq data, consists of multiplying the raw counts for each gene in each sample by a factor incorporating both sequencing depth and putative domains lenght and facilitates a trasparent comparison for putative domains within and between the selections. The second feature in common with RNA-Seq experiments is the statistical model used to identify the enriched/differential domains. Like in RNA-Seq the data count is fitted as a binomial negative (NB) distribution and the false discovery rate (FDR) of p-values obtained are controlled with the Benjamini-Hochberg procedure [147].

The pipeline is divided into 4 main steps of analysis (Figure 12): in the first step, all alignment files are scanned for the detection of putative domains (Figure 12 Step 1). Starting from the genome origin of replication (OriC base N.1) the coverage depth for each genome position is calculated, then start and end coordinates of each domains are defined. The starting coordinate of a putative domain is fixed on the first genome position having a coverage depth equal or greater than 1, while the ending coordinate of the same putative

domain is the next postion having a coverage depth equal to 0. The second step of the analysis is: the annotation inside CDSs of the putative domains previously determined (more than one domain can be associated to one CDS) and the determination of domains enrichment in the selections respect to their representation in the genomic phage library, which is used as reference/background for the step of “differential analysis” (Figure 12 Step 2). In the third and fourth steps of the analysis the lists of domains, significantly differentially enriched in the selections respect to the reference genomic phage library, are compared, and a subtractive step allows the identification of those domains/antigens specific of selections with the sera of sick patients (Figure 12 Step 3). Thus, all the antigens equally recognized by antibodies of healthy patients and of patients affected by Gastric Cancer, Autoimmune gastritis, and MALT lymphoma are removed, and lists of antigens specific infection progression are obtained (Figure 12 Step 4). These lists of specific antigens are given in output as ranking lists associated with statistical values (FDR) thus allowing a guided selection of the best targets for validation (i.e. top lists antigens).

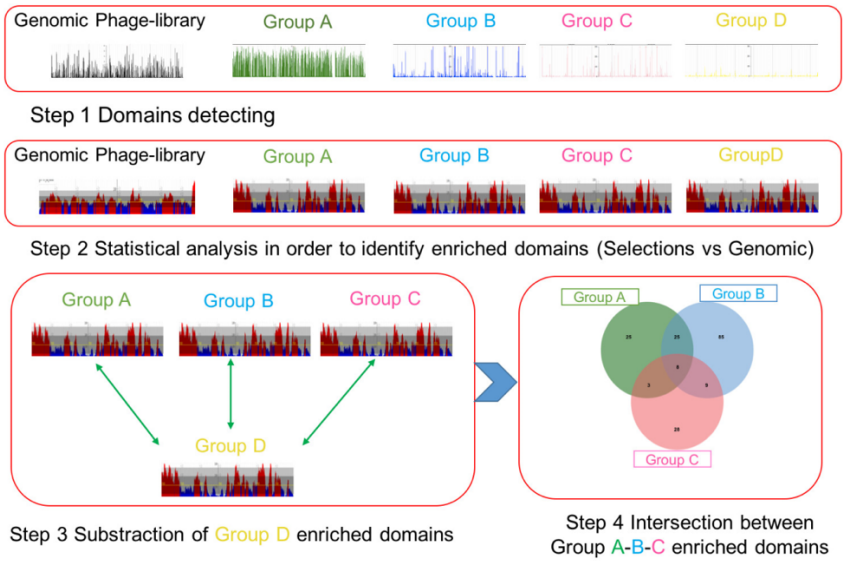


Figure 12 - Workflow of Interactome-Sequencing analysis. Analysis workflow is composed by four steps, Step 1 and 2 allow the detection of putative domains and the identification of only those that are enriched compared to the Genomic Phage library used as background. Steps 3 and 4 allow the identification of the enriched putative domains specific for each Selection group by consecutive subtraction of all the domains common to the Group D selection (healty patients) and intersection of the specific antigens lists obtai

7.4. Overview of Interactome-Seq webtool

7.4.1 Input description

Concerning input data type, Interactome-Seq pipeline can handle Roche-454 reads in FASTA format, Illumina Fastq or alignment files (BAM), single end and paired-end reads are supported. The pipeline is flexible, more than one selection can be analysed, but Genomic dataset input is mandatory, and it can be also analysed alone, if the focus of research is the identification of all potential soluble domains on a genomic scale (Figure 13).

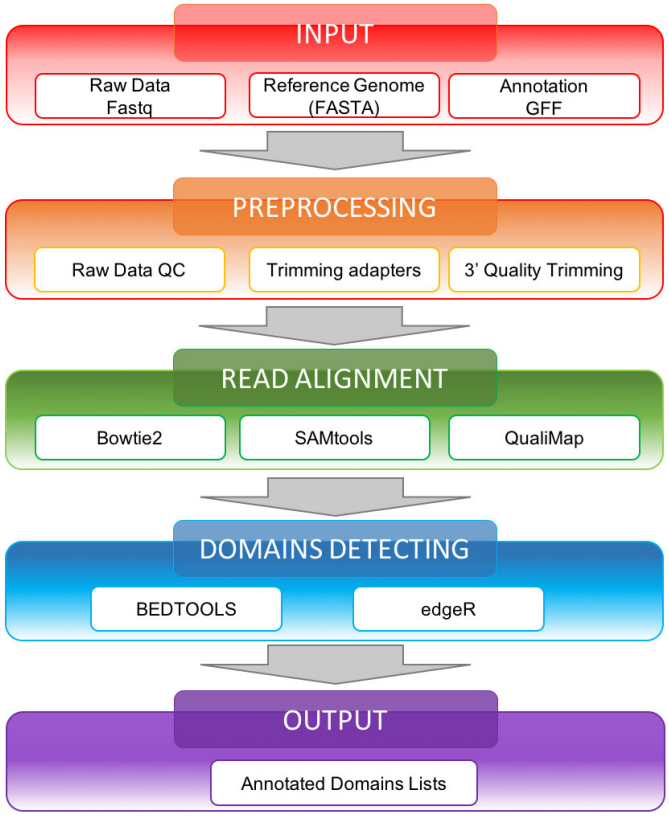


Figure 13 - Schematic steps of Interactome-Sequencing pipeline.

7.4.2. Pre-processing step

As shown in Figure 13 during the pre-processing step raw reads are analysed and quality checked using FastQC and Cutadapt tools. Reads with no identifiable adapters are discarded. During the pre-processing the dataset undergoes also quality trimming, a sliding window trimming is performed, this means that when the average quality within the window falls below a predetermined threshold (Q<30), the remaining part of the read is discarded.

7.4.3. Read alignment step

After the trimming reads are aligned with Bowtie2 to the genome sequence.
I used this mapping tool because it is optimized for working with the read length and error modes yielded by recent sequencers (Illumina, Roche 454, and Ion Torrent platforms). Users can decide the maximum number of mismatches allowed; the default mismatch option is set at 2% of reads length, only uniquely mapping reads are used for downstream analysis. A SAM file is generated and only reads with quality score greater than 30 (Q>30) are processed using SAMtools, and converted into a BAM file.

7.4.4. Putative domains Detection

The Interactome-Seq pipeline after the alignment step invokes bedtools to filter reads that overlap at least 80% inside CDSs, for each portion of CDSs covered by mapping reads, the coverage is calculated, and also max depth and focus values are calculated in according with NGS-Trex tool definition [131].
The coverage is the total number of sequences assigned to a gene; the depth is the maximum number of reads covering a specific genic position; the focus is an index obtained from the ratio between max depth and coverage, and its range is between 0 and 1.
Then, if the focus is higher than 0.8 and the coverage is higher than the average coverage observed for all mapping regions in the BAM file, the CDS portion is considered a putative domain/epitope.

When selections against sera or other interactors are available, domains lists are then statistically tested one at a time against the genomic phage library alignment file. Differential enrichment is calculated by using the R-package edgeR, in order to detect only domains that are enriched after the selections. The differential analysis is always performed testing the BAM files of the selections phage libraries against the BAM file of the genomic phage library that can be considered as the analysis background. .

7.4.5. Output parsing

The last step of the Interactome-Seq pipeline is the output parsing: the list of putative domains, resulted enriched in selection phage libraries respect to the genomic one, are annotated and a simple tabular file (.tab; .csv; .xls) containing the following fields is generated: Chromosome, Domain Start, Domain End, Gene Name, Focus and Annotation.

7.5. Overview of Interactome-Seq webtool

I decided to included Interactome-Seq pipeline in a web-tool to enable users without any bioinformatics or programming skills to perform Interactome-Seq analysis directly loading their reads through the graphical interface and to obtain their results in an easy and user friendly format.

7.5.2. Webtool design

Interactome-Seq webtool is available online at http://cerbero.ba.itb.cnr.it/cricket/ITBHelicobacter.

This webtool requires the following inputs:

1. **Genome sequence**: The Bowtie2 index of the genome of interest can be selected from a database of pre-loaded genomes. The default genome is set to Hp-26695 for the example.
2. **Raw reads file(s):** Users can choose to upload raw reads formatted files in two different ways: (i) a single file in '.fastq'

(Illumina platform) or '.fasta' (Roche 454 platform) format, (ii) alignment files ('.bam' or '.sam'). Alternatively, the user can also select an example file to run the analysis.

3. **Annotation.** User can upload custom annotation or use standard annotation by interrogating the internal database.
4. **Setting Parameters.** Number of Mismatches and sequences of Trimming adaptors should be set before launching the analysis.

There are two main components of the web-tool: the first one is the execution of Genomic analysis and the second is the performance of Case Sample Files. Genomic analysis is mandatory, a user can decide to perform only this step of the analysis if his/her objective is the definition of the ORF-eome of a intron-less genome. The analysis on Case Sample files can be performed only if Genomic analysis has been executed, and users can include one or more Case Samples in the analysis. The web-tool provides a progress bar that is located at the bottom of the web page and shows the processing step of a job in real time (Figure 14).

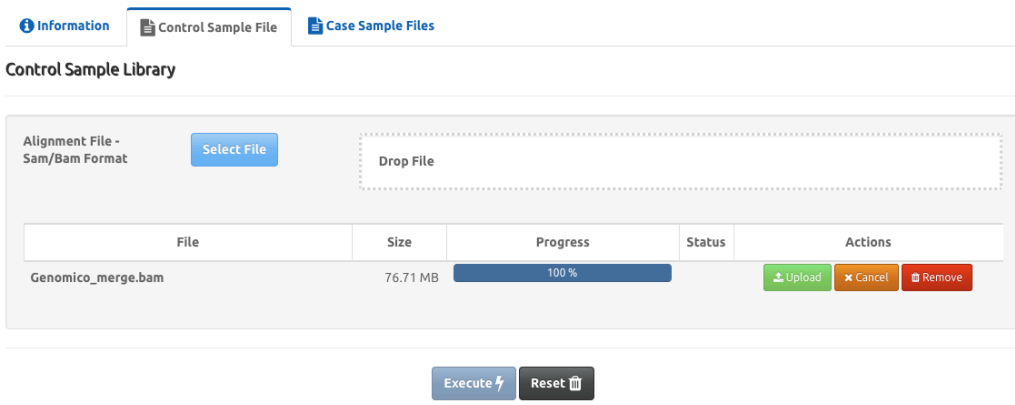


Figure 14 - Screenshot of Interactome-Sequencing webtool input. Input files allowed are FastQ single or paired end, Fasta or alignment file in format BAM. Others input files mandatory for the execution are the genome sequence and annotation file.

7.5.3. Output visualization and description

As shown in Figure 15 the output result of an analysis, performed with the Interactome-Seq web-tool, is a tabular file in which all the putative domains detected and/or found statistically enriched, respect to the Genomic Sample, are reported. The output file is organized in eight columns containing the following information and values (Figure 15):

- Chromosome
- Start of putative domain
- End of putative domain
- Strand
- Gene (ID or name of the gene associated with domain)
- Description (Gene Annotation)
- Q-Value
- Fold Change

These two last values can be used to order domains and/or filter out only statistically significantly enriched domains (for example: Q-Value<0.05 and Fold Change>2 or <0.5) The web-tool allows to sort lists for each of the fields previously described and to download files in tab separated format.

The Interactome-Seq web-tool that I implemented will be useful for scientists without any programming skills, and will provide a fast method to quickly analyse Interactome-seq datasets.

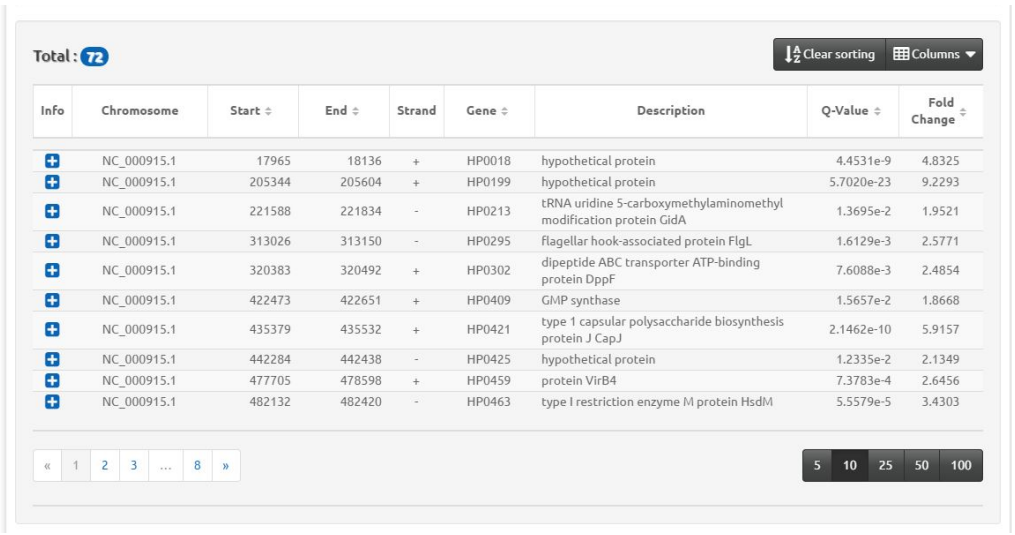
Previously, as I mentioned in the Introduction part of this Thesis, the tool used up to date, for interactome sequencing data analysis, was NGS-Trex [131], but this web-tool was originally created for analysing RNAseq data thus resulting quite inconsistent for the analysis of phage libraries sequencing data, being also no longer suitable to work with Illumina reads..

Recently a paper, in which an approach similar to our Interactome sequencing analysis was applied, has been published [145]. In this paper phage diplay was coupled with NGS and data analysis was performed. A statistical approach, similar to the one that we here describe, based on a two sided

parametric t-test, was applied. Sequencing data deriving from an experiment in which a phage library that displays random 7-mer peptides was challenged against suspensions of *M. tuberculosis,* were analysed to identify peptides that bind mycobacteria.

I used the statistical approach described in this recent paper as a scaffold to implement my pipeline and to integrate all the consecutive data analysis steps previously described.

The result is a new pipeline statistically consistent and highly flexible that, thanks to its integration into the web-tool will be available to a wide plethora of users.


Total : 72

Clear sorting | Columns

Info	Chromosome	Start	End	Strand	Gene	Description	Q-Value	Fold Change
+	NC_000915.1	17965	18136	+	HP0018	hypothetical protein	4.4531e-9	4.8325
+	NC_000915.1	205344	205604	+	HP0199	hypothetical protein	5.7020e-23	9.2293
+	NC_000915.1	221588	221834	-	HP0213	tRNA uridine 5-carboxymethylaminomethyl modification protein GidA	1.3695e-2	1.9521
+	NC_000915.1	313026	313150	-	HP0295	flagellar hook-associated protein FlgL	1.6129e-3	2.5771
+	NC_000915.1	320383	320492	+	HP0302	dipeptide ABC transporter ATP-binding protein DppF	7.6088e-3	2.4854
+	NC_000915.1	422473	422651	+	HP0409	GMP synthase	1.5657e-2	1.8668
+	NC_000915.1	435379	435532	+	HP0421	type 1 capsular polysaccharide biosynthesis protein J CapJ	2.1462e-10	5.9157
+	NC_000915.1	442284	442438	-	HP0425	hypothetical protein	1.2335e-2	2.1349
+	NC_000915.1	477705	478598	+	HP0459	protein VirB4	7.3783e-4	2.6456
+	NC_000915.1	482132	482420	-	HP0463	type I restriction enzyme M protein HsdM	5.5579e-5	3.4303

« 1 2 3 ... 8 »

5 10 25 50 100

Figure 15 - Screenshot of Interactome-Sequencing webtool output. The figure shows an example of output generated, for each putative domains the information reported is Chromosome, that is the name of chromosome used for alignment of the reads,Start,End,Strand,Gene Name that are fields with information about coordinates of putative domains and CDS associated, Q-Value and Fold Change are fields derived from statistical testing. Output can be downloaded or sorted using the arrows near the colums name.

7.6. Novel biomarker identification

The web-tool previously described was used to analyse data deriving both from genomic phage libraries and from selections phage libraries sequencing.

7.6.1. Analysis of the two genomic phage libraries

The two genomic phage libraries constructed, starting from the two strains HP-26695 and HP-B128, were analysed with my new Interactome–seq pipeline in order to define the Domains filtered out by our filtering strategy and to compare them between the two strains.

In Table 6 the sequencing metrics for all the phage libraries obtained are reported. More than 1.4 million and more than 1 million reads were produced for HP-26695 and for HP-B128 phage libraries respectively, thus reaching a total genome coverage of more than 76X and 77X for HP-26695 and HP-B128(B8). When analysing the coverage inside CDSs we found that more than 93% and 89% of the total CDSs of HP-26695 and of HP-B128 where represented in their respective genomic phage libraries, and the percentages of nucleotide covered by reads inside CDSs were respectively 73.5% and 76.8%. The first step of the Interactome-seq pipeline (i.e. Putative Domain Detection) identified a total of 1761 Domains represented into the HP-26695 phage library and 1237 Domains in the HP-B128 phage library (Table 7) (see APPENDIX 10.3); it should be noted that the number of domains detected can be higher than the number of total CDSs of the genome because more than one domain can be found inside a single CDS.

To understand if the domains, found in common CDSs, represented inside the two phage libraries have the same features, in terms the portions of the CDS filtered out, I compared their aminoacidic sequences by blasting all the HP-26695 domains against all the HP-B128 domains using BLASTx (search protein databases using a translated nucleotide query).The blasting step was performed by imposing the following parameters: aminoacidic sequence identity>30% and length overlapping>50%. Among the 1268 common domains 98,82% have an aminoacid sequence identity more than 30% and 60% of

them have an overlap of the aminoacid sequence lenght higher than 50%. Thus after this preliminary comparative evaluation we can conclude that the two instruments constructed (i.e. HP-26695 and HP-B128 phage libraries) can be considered equally representative of *H. pylori* domainome and can be used both separated or mixed together for the selection steps with sera from patients.

7.6.2. Analysis of the three selections phage libraries

The selections against sera from patients who developed: Gastric Cancer (selection A), Autoimmune Gastritis (selection B) and MALT lymphoma (Seleciton C) (Table3), were performed by using both genomic phage libraries previously obtained and described. The two genomic phage libraries of HP-26695 and of HP-B128 were mixed up in equimolar concentration and then the pool of the two libraries was challenged against pools of sera from patients with the three different pathological outcomes. In the three selection steps we pooled together respectively 2 gastric cancer (sel A), 3 autoimmune gastritis (sel B), 2 MALT lymphoma (sel C) sera from patients HP positive, and 6 sera from healthy patients both HP positive or negative.

The pooling of the two phage libraries was aimed at exploiting at maximum level their potentiality in terms of HP domains repertoire and it was validly justified by the previously described comparative analysis that demonstrated their complete functional similarity. The pooling of sera was necessary to reduce the inter-individual variability of antibody titer. In Table 6 the sequencing metrics for all the three selection phage libraries are reported.

A total of more than 9 million paired end reads of 250 bp in length, were produced by sequencing the four selection phage libraries. For each selection library we generated 2.169.178 reads for Control, 815.891 reads for Selection A, 3.737.010 reads for Selection B and 1.633.626 reads for Selection C. After the alignment step, we mapped respectively 1.294.576 (71,97%), 503.401 (73,94%), 2.563.391 (81,37%), 961.686 (75,22%) reads for the Control and for the three Selections A, B and C, and we observed that the total number of

CDSs covered at least by 10 reads were 1225, 1130, 1331, 1263 respectively. However, even if the number of CDSs represented inside the four selection phage libraries was high, and comparable with that obtained for the genomic phage library (Table 6), the percentage of nucleotides covered inside those CDSs decreased to 48,44% (control), 61,86% (sel A), 52,01%(sel B) and 44,97% (sel C). This result shows that the diversity of the libraries obtained after selections with patients' sera was significantly reduced demonstrating that specific enrichment of precise portions of CDSs/Domains occurred, and that probably specific epitopes have been recognized by the antibodies, presents inside the sera from patients, as expected.

	HP-26695	HP-B128	Control	SelA	SelB	SelC
Raw Reads	1.425.554	1.031.956	2.169.178	815.891	3.737.010	1.633.626
Reads after trimming	1.216.124	954.895	1.798.722	680.764	3.149.939	1.278.335
Mapping Reads	1.114.184	746.325	1.294.576	503.401	2.563.391	961.686
% of Mapping reads	91,61%	89,10%	71,97%	73,94%	81,37%	75,22%
Unmapping Reads	11158	158301	166,831	113,397	257,481	128,346
Mean Coverage	76,81X	76,81X	96,13X	31,19X	155,82X	56,97X
CDS covered	1372	1597	1225	1130	1331	1263
% CDS covered	93,46%	93,80%	83,44%	76,97%	90,66%	86,03%
Nucleotides covered inside CDS	1068020	1164096	703739	898694	755447	653331
% of Nucleotide covered inside CDS	73,51%	76,80%	48,44%	61,86%	52,01%	44,97%

Table 6 - Summary of mapping metrics of Genomic phage library and Selection libraries. The metrics reported are the total number of reads sequenced (Raw reads), number of reads after trimming step (Reads after trimming), number of mapping reads (Mapping reads) and percentage of mapping compare to the total number of reads (% of Mapping reads), number of unmapped reads (Unmapped reads), mean coverage is a parameter derived from number of reads * read length / genome size (Mean Coverage), the number of CDS covered is the number of CDS with almost 10 reads mapping, the percentage of CDS covered is number of CDS covered * 100/ total number of CDS annotated,number of nucleotides is the number of bases with depth different of 0 and the percentage of nucleotide covered inside CDS is the Nucleotides covered inside CDS* 100/total CDS nucleotide length.

7.6.2. Identification of common and/or specific biomarkers for HP infection progression

The Interactome–seq data analysis pipeline identified 535, 627, 780 and 652 putative domains respectively present in the Control and A, B, C selection phage libraries (Table 7). Afterwards through the second step of the analysis the domains differentially enriched (Q-value<0.05 and Focus>0.8) in the selection phage libraries respect to the genomic phage libraries were determined. As shown in Table 7, the number of domains found significantly enriched are 115, 125, 183 and 117 respectively for the Control, A, B and C selections. Then the last step of the analysis was applied and output lists were parsed to generate sub-lists of domains/epitopes specific for the pathological outcomes and absent in the healthy controls. This step was carried out in order to eliminate poly-reactive domains and consists of a simple subtractive comparison between the domains lists of Selections A, B and C and the list of the Control selection

	HP-26695	Control	SelA	SelB	SelC
Putative Domains	1761	535	627	780	652
DEG Domains	###	115	125	183	117
Domains subtracted Control	###	###	61	127	49

Table 7 - Overview of putative domains detected, enriched and specific for selections. The first line of table shows the total of Putative domains that are detected for HP-26695 ORF-filtering library, Control and all Selections, the second line of table shows the total number of domains of Contol and Selections that are enriched compare to Genomic HP-26695 ORF-filtering library, in the third line are reported the domains of Selection A, B and C that are resulting after the exclusion of that domains in common with Control Selection. DEG: Differentially Expressed Gene

The Venn diagram in Figure 16 shows the result of this last part of the analysis. It is evident that the most of the enriched domains, found after selections against sera (in total 138), are specific for the three pathological outcomes: Selection A (25), Selection B (85) and Selection C (28), while only 45 enriched domains are common to two or to all the selections.

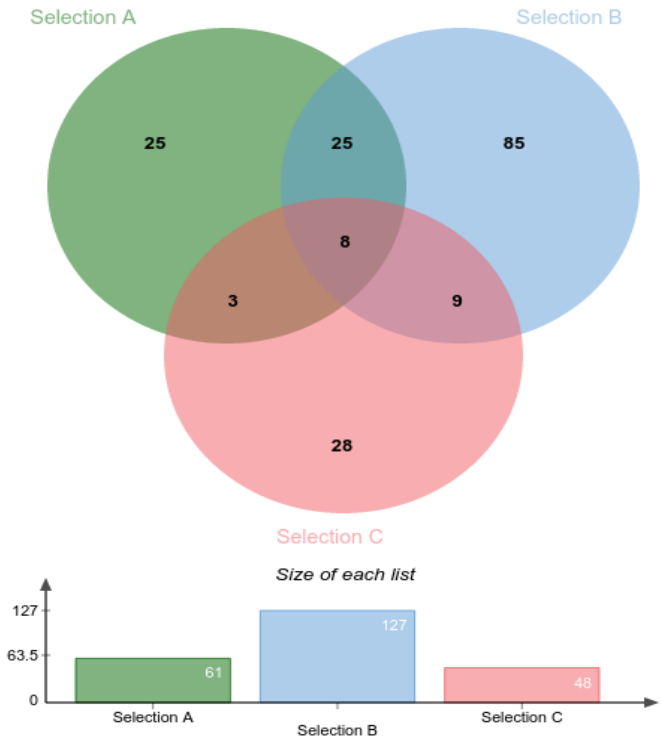


Figure 16 - Venn diagram of specific and common putative antigens. The Venn diagram shows the number of different CDS shared between the libraries. Group A represents sera from individuals affected by gastric adenocarcinoma, Group B represents sera from individuals affected by autoimmune gastritis and group C is sera from patients affected by MALT lymphoma.

This result is really promising, indeed our Interactome-Seq approach coupled with Interactome web-tool data analysis, allowed the identification of many putative domains specifc for the three different pathological outcomes under study.

7.6.3. Validation of one target for HP infection progression

As previously elucidated, it is very important to find novel biomarkers that can provide a “disease signature” for different pathological outcomes, or at least biomarkers associated to HP infection progression towards serious diseases.
Among the sub-lists of domains/epitopes specific for Gastric Cancer, Autoimmune Gastritis, MALT lymphoma and absent in the Healthy controls we selected the top list ones, having the highest Fold Change values (see APPENDIX 10.3-10.4-10.5-10.6-10.7-10.8-10.9) as putative targets for validation with ELISA assay.
Our attention was, from the beginning, captured by the HP0527 gene domains filtered out by our approach and resulting significantly enriched after the selections. This gene also named cagY/cag7 encodes for a large protein of 1927 aminoacids that has been described by Liu et al [149] as one of the main components of *H. pylori cag* T4SS-associated pilum. Interestingly, CagY has an unusual domain structure, in which a number of direct repeats is predicted to cause rearrangements that invariably yield in-frame insertions or deletions [150]. Recent infection studies in murine and non-human primate models have shown that the rearrangements in CagY are sufficient to cause gain or loss of function in the *H. pylori* T4SS and are driven by the host immune system [151]. It was therefore proposed that CagY may function as a sort of molecular switch that alters the function of the T4SS and tunes CagA injection and host pro-inflammatory responses. A majority of *H. pylori*-positive MALT lymphoma patients were serum positive for CagA (9, 10), leading to the hypothesis that CagA might be one causative factor in MALT lymphoma. This observation could be supported by our results, as a matter of fact we find a portion of CagY highly enriched among the top list domains specific for the MALT selection (sel C). At the same time we found highly enriched also other portions of the CagY gene among the top list domains specific for the Gastric Cancer selection (sel A) but also among the top list domains common to all the three pathological conditions.

So we selected two different portions of this gene corresponding to:

1. the domain identified between aminoacidic positions 600 and 853
2. the domain identified between aminoacidic positions 344 and 396

The first one resulted commonly enriched and differential in all the three selections (Figure 17 panel A red box) while the second one is specific for the MALT selection (sel C) and has a very high Fold Change of 7,71 and a length of 761 bp (Figure 17 panel A blue box).

Both these domains were excised from the phagemid vector and cloned into a compatible pGEX-FLAG [130] vector, creating for each output of selection a library for the expression of the selected domain as GST-fusion product (see materials and methods section 6.6)(Figure 17 panel B).

Up to date it was possible to purify the correct recombinant protein only for the first of the two domains, the second one seems to be less soluble and further attempts for purification are ongoing.

To investigate possible cross-reactivity and specificity of the first domain an antibody-capture ELISA assay was implemented (Figure 17 panel B) and a large number of sera, from HP positive and negative Healthy patients (51 sera), from patients with Gastric cancer (29 sera), Autoimmune gastritis (10 sera) and MALT lymphoma (6 sera) were tested (Figure 17 panel C).

It should be noted that the number of patients affected by Autoimmune Gastritis and MALT lymphoma, that was possible to recruit, is much less than the number of patients affected by Gastric Cancer because these two outcomes are rare diseases.

ELISA results show that all the Control Sera (healthy patients) are non-reactive independently from being HP positive or negative, so from this first result we can infer that the domain tested is specific for infection progression towards serious outcomes and not simply related to HP infection. Then looking at the results obtained with the sera from sick patients we found that almost all the sera of HP negative ill patients resulted non-reactive while for the sera of HP positive patients we obtained very high percentages of positive reactivity: 90%

for Gastric Cancer (9 over 10 sera); 75% for Autoimmune Gatsritis (3 over 4 sera) and finally 83% for MALT Lymphoma (5 over 6 sera) (Figure 17 panel D). Thus after ELISA validation we can conclude that this first CagY domain could become a good new biomarker for HP infection progression towards serious pathologies. We expect to obtain much more specific results for the second CagY domain, that we are trying to purify, especially for MALT sera, because, as it can be appreciated from Figure 17 panel A blue box, this portion of the gene seems to be specifically enriched after selection with MALT sera (Sel C).
We have already started to purify other specific domains indentified from our analysis for further ELISA validation assays.
In the next future we will increase the number of sera of patients, that will be recruited over time; furthermore, we will continue with ELISA test validation on other new markers specific for gastric adenocarcinoma, autoimmune gastritis and MALT lymphoma.

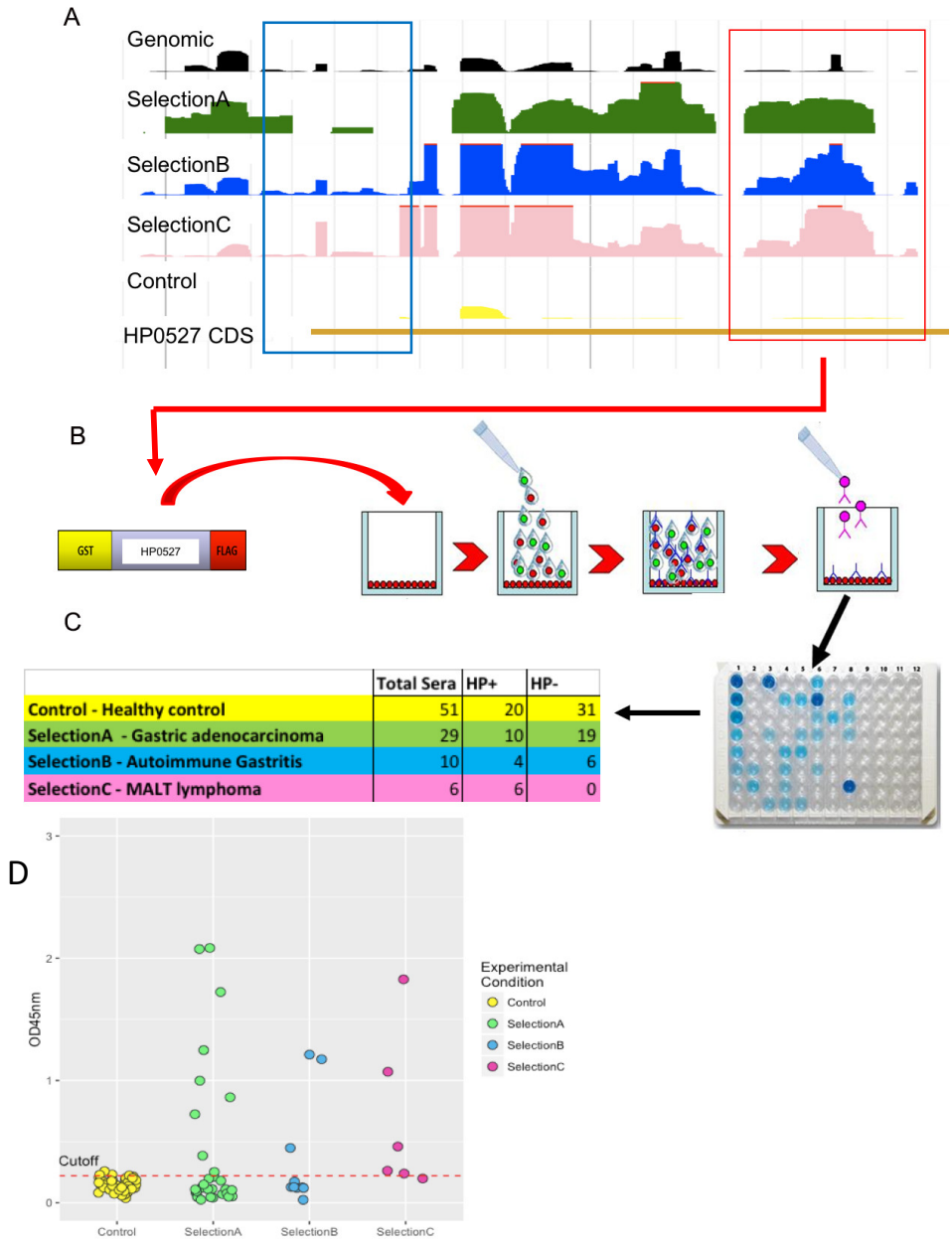


Figure 17 - Schematic overview of HP0527 ELISA validation. Panel A shows genomic browser tracks of HP0527 for genomic and all selections. Panel B shows domains that were excised from the phagemid vector and cloned into a compatible pGEX-FLAG vector. Panel C shows ELISA test sera number used for validation. Panel D shows results of ELISA assay.

8. CONCLUSION AND FUTURE DIRECTIONS

In this section I will summarize the results obtained during the Thesis work and the future perspectives.

The diagnosis of Gastric Cancer and MALT lymphoma is not easy because patients do not show notable symptoms of illness in early onset stages, at the same time they progress rapidly. The possibility of doing an early diagnosis increases the opportunity to save time for active therapy and to improve the survival rate of patients. Many of the previous studies, aiming at biomarkers discovery, were hypothesis-driven, limiting the identification of clinically significant antigens to those directed against proteins that had been already described in pathological outcomes associated with *H. pylori* infection. The Interactome Sequencing approach, here proposed and described, is an ideal unbiased and powerful high-throughput strategy to identify novel antigenic targets. We have exploited the huge data output generated by high throughput sequencing to investigate samples having great complexity, specifically sera samples from patients with different pathological outcomes.

In this thesis work I was able to identify lists of new putative common and specific antigens by: i) constructing ORF genomic filtered libraries from two strains of *H. pylori,* HP-26695 and HP-B128; ii) developing a dedicated web-tool for Interactome-seq data analysis iii) selecting the genomic phage libraries with sera from patients affected by gastric adenocarcinoma, autoimmune gastritis and MALT lymphoma and iv) validating one enriched protein domain specific for infection progression.

Overall, we produced a panel of putative specific domains for three different pathological outcomes, these targets could be validated in the next future comprehensibly associated to the onset of serious diseases related to HP infection.

The new data analysis pipeline here developed and the web-tool implemented has been designed taking into consideration several factors: first of all it is compatible with all sequencing platforms outputs, secondly the “blind” detection of putative domains is very important for reducing the noise

background typical of approaches based on phage display coupled with NGS. Compared to softwares previously used for performing this kind of analysis (NGS-TRex), the Interactome-seq pipeline allows the identification of more than one putative domain/epitope from each protein, thus enabling to discriminate, for example, the portion of CagY specifically recognized by MALT sera from the portion commonly reactive by selections performed with all the three sets of sera tested.

In order to make public and accessible this new pipeline a a web-tool was implemented; this is a user-friendly web interface that doesn't require any programming skills, thus allowing every user to perform Interactome-Seq analysis starting from raw-sequencing data, to sort and visualize the outputs and if needed to download the domain lists in tabular format for interpolation with other results.

In conclusion this thesis work demonstrated that the Interactome-Seq technology is an easily available and powerful strategy, which can help deciphering the molecular mechanisms involved in modulating host immune response.

Providing a new web-tool for data analysis the way to detection of new biomarkers by applying innovative approaches based on interactome sequencing strategy was open.

The release of this tool is timely and relevant for the scientific and clinical community as the importance of using new unbiased experimental and data analysis approaches to discover new biomarkers, for improving early diagnosis of *H. pylori* infection progression towards serious pathological outcome, is among the WHO priorities.

9. REFERENCES

1 B.J. Marshall, J.R. Warren, "UNIDENTIFIED CURVED BACILLI IN THE STOMACH OF PATIENTS WITH GASTRITIS AND PEPTIC ULCERATION",*The Lancet*, Vol. 323, no. 8390, 1984.

2 C. De Martel, J. Ferlay, S. Franceschi, et al., "Global burden of cancers attributable to infections in 2008: A review and synthetic analysis",*The Lancet Oncology*, Vol. 13, no. 6, 2012.

3 D. Makola, D. a Peura, S.E. Crowe, "Helicobacter pylori infection and related gastrointestinal diseases.",*Journal of clinical gastroenterology*, Vol. 41, no. 6, 2007.

4 C.S. Goodwin, W. McConnell, R.K. McCulloch, et al., "Cellular fatty acid composition of Campylobacter pylori from primates and ferrets compared with those of other campylobacters.",*Journal of clinical microbiology*, Vol. 27, no. 5, 1989.

5 M. Camorlinga-Ponce, G. Perez-Perez, G. Gonzalez-Valencia, et al., "Helicobacter pylori genotyping from american indigenous groups shows novel amerindian vacA and cagA alleles and Asian, African and European admixture",*PLoS ONE*, Vol. 6, no. 11, 2011.

6 R.E. Pounder, D. Ng, "The prevalence of Helicobacter pylori infection in different countries.",*Alimentary pharmacology & therapeutics*, Vol. 9 Suppl 2, no. 1995.

7 H.M. Malaty, D.Y. Graham, "Importance of childhood socioeconomic status on the current prevalence of Helicobacter pylori infection.",*Gut*, Vol. 35, no. 6, 1994.

8 J. Khatoon, "Role of Helicobacter pylori in gastric cancer: Updates",*World Journal of Gastrointestinal Oncology*, Vol. 8, no. 2, 2016.

9 C. Josenhans, R.L. Ferrero, A. Labigne, et al., "Cloning and allelic exchange mutagenesis of two flagellin genes of Helicobacter felis",*Molecular Microbiology*, Vol. 33, no. 2, 1999.

10 S.K. Sinha, B. Martin, M. Sargent, et al., "Age at acquisition of

Helicobacter pylori in a pediatric Canadian First Nations population",*Helicobacter*, Vol. 7, no. 2, 2002.

11 R.P. Allaker, K.A. Young, J.M. Hardie, et al., "Prevalence of Helicobacter pylori at oral and gastrointestinal sites in children: Evidence for possible oral-to-oral transmission",*Journal of Medical Microbiology*, Vol. 51, no. 4, 2002.

12 J.F. Tomb, O. White, a R. Kerlavage, et al., "The complete genome sequence of the gastric pathogen Helicobacter pylori.",*Nature*, Vol. 388, no. 6642, 1997.

13 K.P. Haley, J.A. Gaddy, K.P. Haley, et al., "*Helicobacter pylori*: Genomic Insight into the Host-Pathogen Interaction",*International Journal of Genomics*, Vol. 2015, no. 2015.

14 M. Asahi, Y. Tanaka, T. Izumi, et al., "Helicobacter pylori CagA containing ITAM-like sequences localized to lipid rafts negatively regulates VacA-induced signaling in vivo",*Helicobacter*, Vol. 8, no. 1, 2003.

15 W. Fischer, U. Breithaupt, B. Kern, et al., "A comprehensive analysis of Helicobacter pylori plasticity zones reveals that they are integrating conjugative elements with intermediate integration specificity.",*BMC genomics*, Vol. 15, no. 1, 2014.

16 B. Linz, H.M. Windsor, J.J. McGraw, et al., "A mutation burst during the acute phase of Helicobacter pylori infection in humans and rhesus macaques",*Nature Communications*, Vol. 5, no. May, 2014.

17 I.E. Brodsky, R. Medzhitov, "Targeting of immune signalling networks by bacterial pathogens",*Nature Cell Biology*, Vol. 11, no. 5, 2009.

18 B. Linz, H.M. Windsor, J.P. Gajewski, et al., "Helicobacter pylori genomic microevolution during naturally occurring transmission between adults",*PLoS ONE*, Vol. 8, no. 12, 2013.

19 L.E. Wroblewski, R.M. Peek, K.T. Wilson, "Helicobacter pylori and gastric cancer: Factors that modulate disease risk",*Clinical Microbiology Reviews*, Vol. 23, no. 4, 2010.

20 T.L. Cover, M.J. Blaser, "Helicobacter pylori in health and disease.",*Gastroenterology*, Vol. 136, no. 6, 2009.

21 A. Talebi Bezmin Abadi, "Helicobacter pylori: A Beneficial Gastric Pathogen?",*Frontiers in medicine*, Vol. 1, no. 2014.

22 L.E. Wroblewski, L. Shen, S. Ogden, et al., "Helicobacter pylori dysregulation of gastric epithelial tight junctions by urease-mediated myosin II activation.",*Gastroenterology*, Vol. 136, no. 1, 2009.

23 E. Hildebrandt, D.J. McGee, M. Amieva, et al., "Helicobacter pylori lipopolysaccharide modification, Lewis antigen expression, and gastric colonization are cholesterol-dependent",*BMC Microbiology*, Vol. 9, no. 1, 2009.

24 T.L. Testerman, J. Morris, "Beyond the stomach: an updated view of Helicobacter pylori pathogenesis, diagnosis, and treatment.",*World journal of gastroenterology*, Vol. 20, no. 36, 2014.

25 P. Correa, "Human gastric carcinogenesis: a multistep and multifactorial process--First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention.",*Cancer research*, Vol. 52, no. 24, 1992.

26 J. Ferlay, I. Soerjomataram, R. Dikshit, et al., "Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012",*International Journal of Cancer*, Vol. 136, no. 5, 2015.

27 D.B. Polk, R.M. Peek, "Helicobacter pylori: gastric cancer and beyond.",*Nature reviews. Cancer*, Vol. 10, no. 6, 2010.

28 D.E. Guggenheim, M.A. Shah, "Gastric cancer epidemiology and risk factors",*Journal of Surgical Oncology*, Vol. 107, no. 3, 2013.

29 M. Mackenzie, K. Spithoff, D. Jonker, "Systemic therapy for advanced gastric cancer: a clinical practice guideline.",*Current oncology (Toronto, Ont.)*, Vol. 18, no. 2011.

30 S. Tsugane, "Salt, salted food intake, and risk of gastric cancer: Epidemiologic evidence",*Cancer Science*, Vol. 96, no. 1, 2005.

31 X.Q. Wang, P.D. Terry, H. Yan, "Review of salt consumption and

stomach cancer risk: Epidemiological and biological evidence",*World Journal of Gastroenterology*, Vol. 15, no. 18, 2009.

32 K. Shikata, Y. Doi, K. Yonemoto, et al., "Population-based prospective study of the combined influence of cigarette smoking and Helicobacter pylori infection on gastric cancer incidence: the Hisayama Study",*Am J Epidemiol*, Vol. 168, no. 12, 2008.

33 J.M. Yeh, S.J. Goldie, K.M. Kuntz, et al., "Effects of Helicobacter pylori infection and smoking on gastric cancer incidence in China: A population-level analysis of trends and projections",*Cancer Causes and Control*, Vol. 20, no. 10, 2009.

34 J. Watari, N. Chen, P.S. Amenta, et al., "Helicobacter pylori associated chronic gastritis, clinical syndromes, precancerous lesions, and pathogenesis of gastric cancer development",*World Journal of Gastroenterology*, Vol. 20, no. 18, 2014.

35 Y. Matsumoto, H. Marusawa, K. Kinoshita, et al., "Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium.",*Nature medicine*, Vol. 13, no. 4, 2007.

36 R.-G. Zhang, "Role of *Helicobacter pylori* infection in pathogenesis of gastric carcinoma",*World Journal of Gastrointestinal Pathophysiology*, Vol. 7, no. 1, 2016.

37 S. Alzahrani, T.T. Lina, J. Gonzalez, et al., "Effect of Helicobacter pylori on gastric epithelial cells",*World Journal of Gastroenterology*, Vol. 20, no. 36, 2014.

38 W. Fischer, S. Prassl, R. Haas, "Virulence mechanisms and persistence strategies of the human gastric pathogen Helicobacter pylori",*Current Topics in Microbiology and Immunology*, Vol. 337, no. 1, 2009.

39 X. Yong, B. Tang, B.-S. Li, et al., "Helicobacter pylori virulence factor CagA promotes tumorigenesis of gastric cancer via multiple signaling pathways.",*Cell communication and signaling : CCS*, Vol. 13, no. 2015.

40 V. Ricci, "Relationship between VacA Toxin and Host Cell Autophagy in Helicobacter pylori Infection of the Human Stomach: A Few Answers,

Many Questions.",*Toxins*, Vol. 8, no. 7, 2016.

41 L.K. Greenfield, N.L. Jones, "Modulation of autophagy by Helicobacter pylori and its role in gastric carcinogenesis",*Trends in Microbiology*, Vol. 21, no. 11, 2013.

42 W.-C. Lin, H.-F. Tsai, H.-J. Liao, et al., "Helicobacter pylori sensitizes TNF-related apoptosis-inducing ligand (TRAIL)-mediated apoptosis in human gastric epithelial cells through regulation of FLIP.",*Cell death & disease*, Vol. 5, no. 2014.

43 M. Miftahussurur, A.F. Syam, D. Makmun, et al., "Helicobacter pylori virulence genes in the five largest islands of Indonesia",*Gut Pathogens*, Vol. 7, no. 1, 2015.

44 H. Ashktorab, R.H. Dashwood, M.M. Dashwood, et al., "H. pylori-induced apoptosis in human gastric cancer cells mediated via the release of apoptosis-inducing factor from mitochondria",*Helicobacter*, Vol. 13, no. 6, 2008.

45 M.R. Amieva, E.M. El-Omar, "Host-Bacterial Interactions in Helicobacter pylori Infection",*Gastroenterology*, Vol. 134, no. 1, 2008.

46 A. Sokic-Milutinovic, T. Alempijevic, T. Milosavljevic, "Role of Helicobacter pylori infection in gastric carcinogenesis: Current knowledge and future directions",*World Journal of Gastroenterology*, Vol. 21, no. 41, 2015.

47 K. Iijima, Y. Abe, R. Kikuchi, et al., "Serum biomarker tests are useful in delineating between patients with gastric atrophy and normal, healthy stomach",*World Journal of Gastroenterology*, Vol. 15, no. 7, 2009.

48 M. Vieth, H. Neumann, C. Falkeis, "The diagnosis of gastritis",*Diagnostic Histopathology*, Vol. 20, no. 6, 2014.

49 B.H. Toh, "Diagnosis and classification of autoimmune gastritis",*Autoimmunity Reviews*, Vol. 13, nos. 4–5, 2014.

50 A. Antico, M. Tampoia, D. Villalta, et al., "Clinical usefulness of the serological gastric biopsy for the diagnosis of chronic autoimmune gastritis",*Clinical and Developmental Immunology*, Vol. 2012, no. 2012.

51 I. Coati, M. Fassan, F. Farinati, et al., "Autoimmune gastritis: Pathologist's viewpoint.",*World journal of gastroenterology*, Vol. 21, no. 42, 2015.

52 S.M. Cohen, M. Petryk, M. Varma, et al., "Non-Hodgkin's lymphoma of mucosa-associated lymphoid tissue.",*The oncologist*, Vol. 11, no. 10, 2007.

53 N. Chaudhary, H. Ozer, D. Huard, et al., "Successful treatment of Helicobacter Pylori-negative gastric MALT lymphoma with rituximab",*Digestive Diseases and Sciences*, Vol. 51, no. 4, 2006.

54 A. Zullo, C. Hassan, L. Ridola, et al., "Gastric MALT lymphoma: old and new insights.",*Annals of gastroenterology : quarterly publication of the Hellenic Society of Gastroenterology*, Vol. 27, no. 1, 2014.

55 A. Zullo, C. Hassan, A. Andriani, et al., "Primary low-grade and high-grade gastric MALT-lymphoma presentation.",*Journal of clinical gastroenterology*, Vol. 44, no. 5, 2010.

56 E. Zucca, F. Bertoni, "The spectrum of MALT lymphoma at different sites: biological and therapeutic relevance", Vol. no. n.d.

57 N. Vakil, F. Megraud, "Eradication Therapy for Helicobacter pylori",*Gastroenterology*, Vol. 133, no. 3, 2007.

58 J.P. Gisbert, X. Calvet, "Review article: Common misconceptions in the management of Helicobacter pylori-associated gastric MALT-lymphoma",*Alimentary Pharmacology and Therapeutics*, Vol. 34, no. 9, 2011.

59 T. Hussell, P.G. Isaacson, J. Spencer, et al., "The response of cells from low-grade B-cell gastric lymphomas of mucosa-associated lymphoid tissue to Helicobacter pylori",*The Lancet*, Vol. 342, no. 8871, 1993.

60 G. Driessens, J. Kline, T.F. Gajewski, "Costimulatory and coinhibitory receptors in anti-tumor immunity.",*Immunological reviews*, Vol. 229, no. 1, 2009.

61 K. Troppan, K. Wenzl, P. Neumeister, et al., "Molecular Pathogenesis of MALT Lymphoma",*Gastroenterol Res Pract*, Vol. 2015, no. 2015.

62 Q. Hu, Y. Zhang, X. Zhang, et al., "Gastric mucosa-associated lymphoid tissue lymphoma and Helicobacter pylori infection: a review of current diagnosis and management.",*Biomarker research*, Vol. 4, no. 2016.

63 M.G. Mancino, M. Bianchi, V. Festa, et al., "Recent advances in endoscopic management of gastrointestinal cancers",*Translational Gastrointestinal Cancer*, Vol. 3, no. 2, 2014.

64 C.M. Bacon, M.-Q. Du, A. Dogan, "Mucosa-associated lymphoid tissue (MALT) lymphoma: a practical guide for pathologists.",*Journal of clinical pathology*, Vol. 60, no. 4, 2007.

65 A.C. Wotherspoon, T.C. Diss, L. Pan, et al., "Regression of primary low-grade B-cell gastric lymphoma of mucosa-associated lymphoid tissue type after eradication of Helicobacter pylori",*The Lancet*, Vol. 342, no. 8871, 1993.

66 S. Kazemi, H. Tavakkoli, M.R. Habizadeh, et al., "Diagnostic values of Helicobacter pylori diagnostic tests: stool antigen test, urea breath test, rapid urease test, serology and histology.",*Journal of research in medical sciences : the official journal of Isfahan University of Medical Sciences*, Vol. 16, no. 9, 2011.

67 N. Ahmed, M.F. Loke, N. Kumar, et al., "Helicobacter pylori in 2013: Multiplying Genomes, Emerging Insights", Vol. no. n.d.

68 G. Spohn, V. Scarlato, "Motility, Chemotaxis, and Flagella",*Helicobacter pylori: Physiology and Genetics*, Vol. no. 2001.

69 R.M. Peek, C. Fiske, K.T. Wilson, "Role of innate immunity in Helicobacter pylori-induced gastric malignancy.",*Physiological reviews*, Vol. 90, no. 3, 2010.

70 J. Haiko, B. Westerlund-Wikström, "The role of the bacterial flagellum in adhesion and virulence.",*Biology*, Vol. 2, no. 4, 2013.

71 C.-Y. Kao, B.-S. Sheu, J.-J. Wu, "Helicobacter pylori infection: An overview of bacterial virulence factors and pathogenesis",*Biomedical Journal*, Vol. 39, no. 1, 2016.

72 J.G. Kusters, A.H.M. Van Vliet, E.J. Kuipers, "Pathogenesis of

Helicobacter pylori infection",*Clinical Microbiology Reviews*, Vol. 19, no. 3, 2006.

73 M. Rektorschek, A. Buhmann, D. Weeks, et al., "Acid resistance of Helicobacter pylori depends on the UreI membrane protein and an inner membrane proton barrier.",*Molecular microbiology*, Vol. 36, no. 1, 2000.

74 R. McNulty, J.P. Ulmschneider, H. Luecke, et al., "Mechanisms of molecular transport through the urea channel of Helicobacter pylori",*Nature Communications*, Vol. 4, no. 2013.

75 G. Sachs, Y. Wen, D.R. Scott, "Gastric infection by Helicobacter pylori.",*Current gastroenterology reports*, Vol. 11, no. 6, 2009.

76 P. Lertsethtakarn, K.M. Ottemann, D.R. Hendrixson, "Motility and chemotaxis in Campylobacter and Helicobacter",*Annual Review of Microbiology*, Vol. 65, no. 1, 2011.

77 H. Asakura, Y. Churin, B. Bauer, et al., "Helicobacter pylori HP0518 affects flagellin glycosylation to alter bacterial motility",*Molecular Microbiology*, Vol. 78, no. 5, 2010.

78 C. Berne, A. Ducret, G.G. Hardy, et al., "Adhesins Involved in Attachment to Abiotic Surfaces by Gram-Negative Bacteria.",*Microbiology spectrum*, Vol. 3, no. 4, 2015.

79 B. Kalali, R. Mejías-Luque, A. Javaheri, et al., "H. pylori virulence factors: influence on immune system and pathology.",*Mediators of inflammation*, Vol. 2014, no. 2014.

80 M. Oleastro, A. Ménard, "The Role of Helicobacter pylori Outer Membrane Proteins in Adherence and Pathogenesis.",*Biology*, Vol. 2, no. 3, 2013.

81 H.-W. Fu, "Helicobacter pylori neutrophil-activating protein: From molecular pathogenesis to clinical applications.",*World journal of gastroenterology : WJG*, Vol. 20, no. 18, 2014.

82 Y. Yamaoka, "Roles of Helicobacter pylori BabA in gastroduodenal pathogenesis.",*World journal of gastroenterology*, Vol. 14, no. 27, 2008.

83 J.R. White, J.A. Winter, K. Robinson, "Differential inflammatory

response to Helicobacter pylori infection: etiology and clinical outcomes.",*Journal of inflammation research*, Vol. 8, no. 2015.

84 G. Kaul, H. Thippeswamy, "Role of heat shock proteins in diseases and their therapeutic potential.",*Indian journal of microbiology*, Vol. 51, no. 2, 2011.

85 A. Tanaka, T. Kamada, K. Yokota, et al., "Helicobacter pylori heat shock protein 60 antibodies are associated with gastric cancer",*Pathology, research and practice*, Vol. 205, no. 10, 2009.

86 Y. Zhao, K. Yokota, K. Ayada, et al., "Helicobacter pylori heat-shock protein 60 induces interleukin-8 via a Toll-like receptor (TLR)2 and mitogen-activated protein (MAP) kinase pathway in human monocytes",*Journal of Medical Microbiology*, Vol. 56, no. PART 2, 2007.

87 G. Posselt, S. Backert, S. Wessler, "The functional interplay of Helicobacter pylori factors with gastric epithelial cells induces a multi-step process in pathogenesis.",*Cell communication and signaling : CCS*, Vol. 11, no. 1, 2013.

88 H. Parker, J.I. Keenan, "Composition and function of Helicobacter pylori outer membrane vesicles.",*Microbes and infection / Institut Pasteur*, Vol. 14, no. 1, 2012.

89 B. Busch, R. Weimer, C. Woischke, et al., "Helicobacter pylori interferes with leukocyte migration via the outer membrane protein HopQ and via CagA translocation",*International Journal of Medical Microbiology*, Vol. 305, no. 3, 2015.

90 P. Olbermann, C. Josenhans, Y. Moodley, et al., "A global overview of the genetic and functional diversity in the helicobacter pylori cag pathogenicity island",*PLoS Genetics*, Vol. 6, no. 8, 2010.

91 P. Correa, M.B. Piazuelo, "Evolutionary history of the Helicobacter pylori genome: Implications for gastric carcinogenesis",*Gut and Liver*, Vol. 6, no. 1, 2012.

92 R.M. Ferreira, J.C. Machado, C. Figueiredo, "Clinical relevance of Helicobacter pylori vacA and cagA genotypes in gastric

carcinoma.",*Best practice & research. Clinical gastroenterology*, Vol. 28, no. 6, 2014.

93 J.L. Rhead, D.P. Letley, M. Mohammadi, et al., "A New Helicobacter pylori Vacuolating Cytotoxin Determinant, the Intermediate Region, Is Associated With Gastric Cancer",*Gastroenterology*, Vol. 133, no. 3, 2007.

94 T.L. Cover, "*Helicobacter pylori* Diversity and Gastric Cancer Risk",*mBio*, Vol. 7, no. 1, 2016.

95 J.C. Atherton, "The pathogenesis of Helicobacter pylori-induced gastroduodenal diseases.",*Annual review of pathology*, Vol. 1, no. 2006.

96 J. Rassow, M. Meinecke, "Helicobacter pylori VacA: A new perspective on an invasive chloride channel",*Microbes and Infection*, Vol. 14, no. 12, 2012.

97 V. Ricci, M. Giannouli, M. Romano, et al., "*Helicobacter pylori* gamma-glutamyl transpeptidase and its pathogenic role.",*World journal of gastroenterology : WJG*, Vol. 20, no. 3, 2014.

98 H. Lu, P.I. Hsu, D.Y. Graham, et al., "Duodenal ulcer promoting gene of Helicobacter pylori",*Gastroenterology*, Vol. 128, no. 4, 2005.

99 S.W. Jung, M. Sugimoto, S. Shiota, et al., "The intact dupA cluster is a more reliable Helicobacter pylori virulence marker than dupA alone.",*Infection and immunity*, Vol. 80, no. 1, 2012.

100 Y.-K. Wang, F.-C. Kuo, C.-J. Liu, et al., "Diagnosis of Helicobacter pylori infection: Current options and developments.",*World journal of gastroenterology*, Vol. 21, no. 40, 2015.

101 J.Y. Lee, N. Kim, "Diagnosis of Helicobacter pylori by invasive test: histology.",*Annals of translational medicine*, Vol. 3, no. 1, 2015.

102 T. Uotani, D.Y. Graham, "Diagnosis of Helicobacter pylori using the rapid urease test.",*Annals of translational medicine*, Vol. 3, no. 1, 2015.

103 E. Garza-González, G.I. Perez-Perez, H.J. Maldonado-Garza, et al., "A review of Helicobacter pylori diagnosis, treatment, and methods to detect eradication.",*World journal of gastroenterology*, Vol. 20, no. 6,

2014.

104 W.A. Gramley, A. Asghar, H.F. Frierson, et al., "Detection of Helicobacter pylori DNA in fecal samples from infected individuals.",*Journal of clinical microbiology*, Vol. 37, no. 7, 1999.

105 S.K. Patel, C.B. Pratap, A.K. Jain, et al., "Diagnosis of Helicobacter pylori: what should be the gold standard?",*World journal of gastroenterology*, Vol. 20, no. 36, 2014.

106 M. Miftahussurur, Y. Yamaoka, "Diagnostic Methods of Helicobacter pylori Infection for Epidemiological Studies: Critical Importance of Indirect Test Validation.",*BioMed research international*, Vol. 2016, no. 2016.

107 M. Foroutan, B. Loloei, S. Irvani, et al., "Accuracy of rapid urease test in diagnosing Helicobacter pylori infection in patients using NSAIDs.",*Saudi journal of gastroenterology : official journal of the Saudi Gastroenterology Association*, Vol. 16, no. 2, 2010.

108 A. Iranikhah, M.-R. Ghadir, S. Sarkeshikian, et al., "Stool antigen tests for the detection of Helicobacter pylori in children.",*Iranian journal of pediatrics*, Vol. 23, no. 2, 2013.

109 P.A. Marchildon, T. Sugiyama, Y. Fukuda, et al., "Evaluation of the effects of strain-specific antigen variation on the accuracy of serologic diagnosis of Helicobacter pylori infection.",*Journal of clinical microbiology*, Vol. 41, no. 4, 2003.

110 A.I. Lopes, F.F. Vale, M. Oleastro, "Helicobacter pylori infection - recent developments in diagnosis.",*World journal of gastroenterology*, Vol. 20, no. 28, 2014.

111 L. Formichella, L. Romberg, C. Bolz, et al., "A novel line immunoassay based on recombinant virulence factors enables highly specific and sensitive serologic diagnosis of Helicobacter pylori infection.",*Clinical and vaccine immunology : CVI*, Vol. 20, no. 11, 2013.

112 K. Chandramouli, P.-Y. Qian, "Proteomics: challenges, techniques and possibilities to overcome biological sample complexity.",*Human*

genomics and proteomics : HGP, Vol. 2009, no. 2009.

113 P. Molek, B. Strukelj, T. Bratkovic, "Peptide phage display as a tool for drug discovery: Targeting membrane receptors",*Molecules*, Vol. 16, no. 1, 2011.

114 S. Naseem, J. Meens, J. Jores, et al., "Phage display-based identification and potential diagnostic application of novel antigens from Mycoplasma mycoides subsp. mycoides small colony type",*Veterinary Microbiology*, Vol. 142, nos. 3–4, 2010.

115 T. Meyer, T. Schirrmann, A. Frenzel, et al., "Identification of immunogenic proteins and generation of antibodies against Salmonella Typhimurium using phage display.",*BMC biotechnology*, Vol. 12, no. 2012.

116 D.O. Connor, J. Zantow, M. Hust, et al., "Identification of Novel Immunogenic Proteins of Neisseria gonorrhoeae by Phage Display",*PLoS ONE*, Vol. 11, no. 2, 2016.

117 L. Aghebati-Maleki, B. Bakhshinejad, B. Baradaran, et al., "Phage display as a promising approach for vaccine development.",*Journal of biomedical science*, Vol. 23, no. 1, 2016.

118 J. Bazan, I. Całkosiński, A. Gamian, "Phage display--a powerful technique for immunotherapy: 1. Introduction and potential of therapeutic applications.",*Human vaccines & immunotherapeutics*, Vol. 8, no. 12, 2012.

119 W. Li, N.B. Caberoy, "New perspective for phage display as an efficient and versatile technology of functional proteomics",*Applied Microbiology and Biotechnology*, Vol. 85, no. 4, 2010.

120 H. Imaizumi-Anraku, J. Webb, N. Takeda, et al., "Interaction network containing conserved and essential protein complexes in Escherichia coli",*Nature*, Vol. 433, no. 7025, 2005.

121 R.M. Ewing, P. Chu, F. Elisma, et al., "Large-scale mapping of human protein–protein interactions by mass spectrometry",*Molecular Systems Biology*, Vol. 3, no. 1, 2007.

122 N.B. Caberoy, Y. Zhou, X. Jiang, et al., "Efficient identification of tubby-binding proteins by an improved system of T7 phage display",*Journal of Molecular Recognition*, Vol. 23, no. 1, 2010.

123 A.R.M. Bradbury, S. Sidhu, S. Dübel, et al., "Beyond natural antibodies: the power of in vitro display technologies.",*Nature biotechnology*, Vol. 29, no. 3, 2011.

124 N.B. Caberoy, Y. Zhou, G. Alvarado, et al., "Efficient identification of phosphatidylserine-binding proteins by ORF phage display",*Biochemical and Biophysical Research Communications*, Vol. 386, no. 1, 2009.

125 N.A.C. Ngubane, L. Gresh, T.R. loerger, et al., "High-throughput sequencing enhanced phage display identifies peptides that bind mycobacteria",*PLoS ONE*, Vol. 8, no. 11, 2013.

126 D. Hou, C. Chen, E.J. Seely, et al., "High-Throughput Sequencing-Based Immune Repertoire Study during Infectious Disease.",*Frontiers in immunology*, Vol. 7, no. 2016.

127 E. Dias-Neto, D.N. Nunes, R.J. Giordano, et al., "Next-generation phage display: integrating and comparing available molecular tools to enable cost-effective high-throughput analysis.",*PloS one*, Vol. 4, no. 12, 2009.

128 L.J. Gourlay, C. Peano, C. Deantonio, et al., "Selecting soluble/foldable protein domains through single-gene or genomic ORF filtering: Structure of the head domain of Burkholderia pseudomallei antigen BPSL2063",*Acta Crystallographica Section D: Biological Crystallography*, Vol. 71, no. Pt 11, 2015.

129 R. Di Niro, A.M. Sulic, F. Mignone, et al., "Rapid interactome profiling by massive sequencing",*Nucleic Acids Research*, Vol. 38, no. 9, 2010.

130 S. D'Angelo, F. Mignone, C. Deantonio, et al., "Profiling celiac disease antibody repertoire",*Clinical Immunology*, Vol. 148, no. 1, 2013.

131 I. Boria, L. Boatti, G. Pesole, et al., "NGS-Trex: Next Generation Sequencing Transcriptome profile explorer.",*BMC bioinformatics*, Vol. 14 Suppl 7, no. SUPPL7, 2013.

132 S. D'Angelo, N. Velappan, F. Mignone, et al., "Filtering "genic" open

reading frames from genomic DNA samples for advanced annotation.",*BMC genomics*, Vol. 12 Suppl 1, no. SUPPL. 1, 2011.

133 D. Sblattero, A. Bradbury, "Exploiting recombination in single bacteria to make large phage antibody libraries.",*Nature biotechnology*, Vol. 18, no. 2000.

134 D.R. Zerbino, E. Birney, "Velvet: Algorithms for de novo short read assembly using de Bruijn graphs",*Genome Research*, Vol. 18, no. 5, 2008.

135 M. Galardini, E.G. Biondi, M. Bazzicalupo, et al., "CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes",*Source Code for Biology and Medicine*, Vol. 6, no. 1, 2011.

136 R.K. Aziz, D. Bartels, A. a Best, et al., "The RAST Server: rapid annotations using subsystems technology.",*BMC genomics*, Vol. 9, no. 2008.

137 J. Ågren, A. Sundström, T. Håfström, et al., "Gegenees: Fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups",*PLoS ONE*, Vol. 7, no. 6, 2012.

138 D.H. Huson, D. Bryant, "Application of phylogenetic networks in evolutionary studies",*Molecular Biology and Evolution*, Vol. 23, no. 2, 2006.

139 M. Martin, "Cutadapt removes adapter sequences from high-throughput sequencing reads",*EMBnet.journal*, Vol. 17, no. 1, 2011.

140 B. Langmead, S.L. Salzberg, "Fast gapped-read alignment with Bowtie 2",*Nature Methods*, Vol. 9, no. 4, 2012.

141 H. Li, B. Handsaker, A. Wysoker, et al., "The Sequence Alignment/Map format and SAMtools",*Bioinformatics*, Vol. 25, no. 16, 2009.

142 A.R. Quinlan, "BEDTools: The Swiss-Army tool for genome feature analysis",*Current Protocols in Bioinformatics*, Vol. 2014, no. 2014.

143 M.D. Robinson, D.J. McCarthy, G.K. Smyth, "edgeR: A Bioconductor package for differential expression analysis of digital gene expression

data",*Bioinformatics*, Vol. 26, no. 1, 2009.

144 J. Kang, M.J. Blaser, "Bacterial populations as perfect gases: genomic integrity and diversification tensions in Helicobacter pylori.",*Nature reviews. Microbiology*, Vol. 4, no. 11, 2006.

145 M. Farnbacher, T. Jahns, D. Willrodt, et al., "Sequencing, annotation, and comparative genome analysis of the gerbil-adapted Helicobacter pylori strain B8.",*BMC genomics*, Vol. 11, no. 2010.

146 R. Overbeek, R. Olson, G.D. Pusch, et al., "The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST)",*Nucleic Acids Research*, Vol. 42, no. D1, 2014.

147 Y. Benjamini, Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing",*Source Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 57, no. 1, 1995.

148 B.M. Scott, W.L. Matochko, R.F. Gierczak, et al., "Phage Display of the Serpin Alpha-1 Proteinase Inhibitor Randomized at Consecutive Residues in the Reactive Centre Loop and Biopanned with or without Thrombin",*PLoS ONE*, Vol. 9, no. 1, 2014.

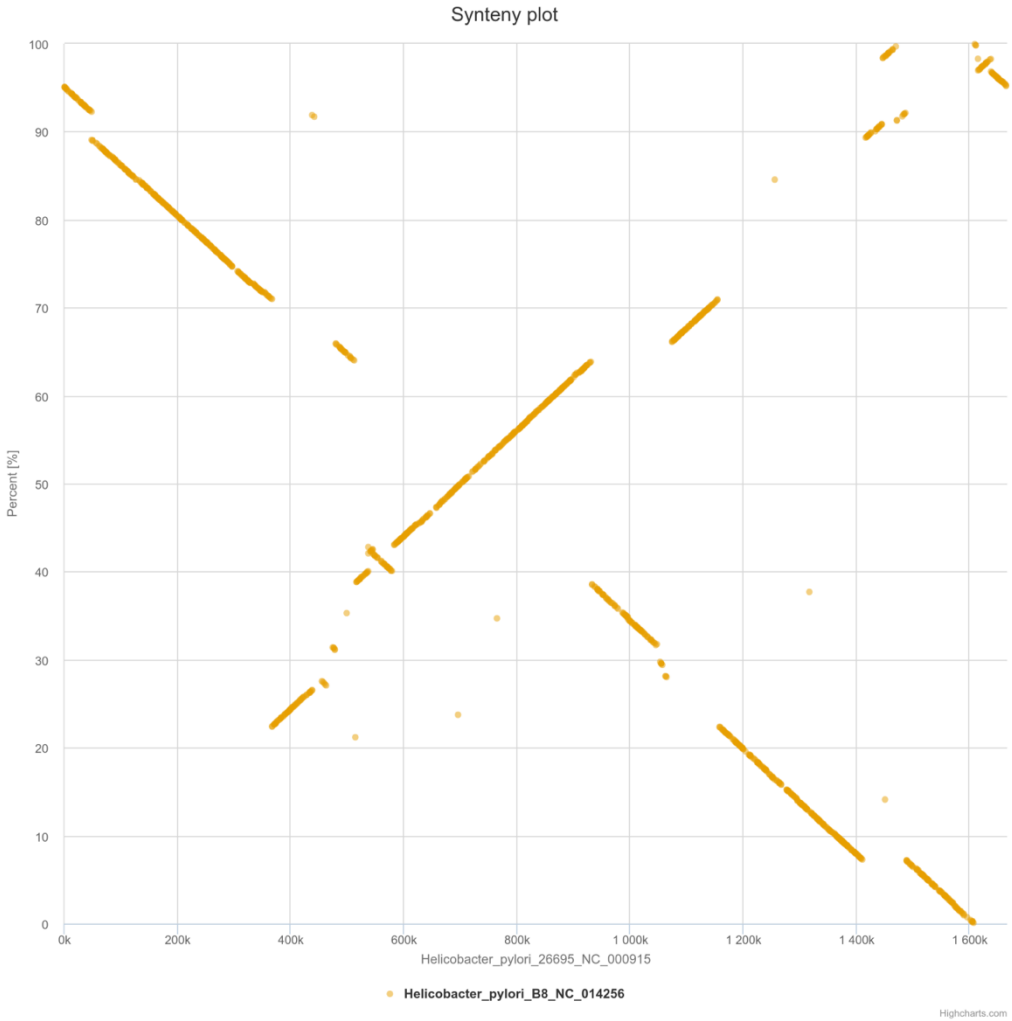
149 G. Liu, T.K. McDaniel, S. Falkow, et al., "Sequence anomalies in the Cag7 gene of the Helicobacter pylori pathogenicity island.",*Proceedings of the National Academy of Sciences of the United States of America*, Vol. 96, no. 12, 1999.

150 R.A. Aras, W. Fischer, G.I. Perez-Perez, et al., "Plasticity of repetitive DNA sequences within a bacterial (Type IV) secretion system component.",*The Journal of experimental medicine*, Vol. 198, no. 9, 2003.

151 R.M. Barrozo, C.L. Cooke, L.M. Hansen, et al., "Functional Plasticity in the Type IV Secretion System of Helicobacter pylori", Vol. no. 2013.

10. APPENDIX

10.1. Synteny plot of HP-26695 and HP-B128



10.2. HP-26695 Genomic Domains

CHR	START	END	GENE_ID
NC_000915	2750	3018	HP0005
NC_000915	9339	9574	HP0011
NC_000915	16862	17165	HP0018
NC_000915	20596	20890	HP0021
NC_000915	27157	27358	HP0026
NC_000915	27642	28180	HP0027

NC_000915	28900	29293	HP0028
NC_000915	31960	32344	HP0031
NC_000915	34508	34722	HP0033
NC_000915	41162	41382	HP0043
NC_000915	43242	43482	HP0045
NC_000915	47004	47542	HP0048
NC_000915	50497	50927	HP0051
NC_000915	52551	53065	HP0053
NC_000915	61618	61828	HP0057
NC_000915	67353	67792	HP0063
NC_000915	75066	75320	HP0071
NC_000915	84358	84667	HP0080
NC_000915	87828	88141	HP0082
NC_000915	98567	98901	HP0092
NC_000915	100684	101444	HP0096
NC_000915	101645	102356	HP0097
NC_000915	112851	113200	HP0105
NC_000915	115793	115996	HP0108
NC_000915	124752	125086	HP0116
NC_000915	128034	128489	HP0118
NC_000915	128534	128964	HP0118
NC_000915	135989	136323	HP0123
NC_000915	143701	144650	HP0133
NC_000915	144688	144835	HP0133
NC_000915	149234	149334	HP0138
NC_000915	152748	153301	HP0141
NC_000915	153513	153628	HP0141
NC_000915	156608	156999	HP0144
NC_000915	160724	161118	HP0150
NC_000915	167975	168396	HP0159
NC_000915	178792	178988	HP0172
NC_000915	181234	181434	HP0174
NC_000915	181498	181717	HP0174a
NC_000915	184797	185017	HP0178
NC_000915	185865	186178	HP0179
NC_000915	187996	188197	HP0181

NC_000915	188808	189236	HP0182
NC_000915	194280	194374	HP0186
NC_000915	201347	201606	HP0194
NC_000915	202115	202470	HP0195
NC_000915	210515	211056	HP0205
NC_000915	212346	212878	HP0207
NC_000915	213992	214434	HP0208
NC_000915	216274	217226	HP0210
NC_000915	219112	219302	HP0212
NC_000915	222219	222375	HP0214
NC_000915	222572	223080	HP0214
NC_000915	234129	234412	HP0226
NC_000915	237950	238457	HP0228
NC_000915	238752	239071	HP0229
NC_000915	243848	244379	HP0234
NC_000915	246048	246165	HP0235
NC_000915	246311	246588	HP0236
NC_000915	246637	247082	HP0237
NC_000915	252286	252662	HP0243
NC_000915	255066	255396	HP0246
NC_000915	258812	259536	HP0250
NC_000915	263313	264080	HP0254
NC_000915	264209	264599	HP0254
NC_000915	271082	271647	HP0262
NC_000915	271940	272436	HP0263
NC_000915	274800	275196	HP0264
NC_000915	275249	275546	HP0265
NC_000915	281385	281576	HP0271
NC_000915	285031	285280	HP0277
NC_000915	292078	292479	HP0283
NC_000915	297483	297726	HP0288
NC_000915	298023	298299	HP0289
NC_000915	305966	306705	HP0289
NC_000915	313299	313663	HP0295
NC_000915	315584	316040	HP0298
NC_000915	316398	317096	HP0298

NC_000915	319842	319981	HP0301
NC_000915	330903	331245	HP0316
NC_000915	339414	339957	HP0323
NC_000915	341636	342159	HP0326
NC_000915	342227	342588	HP0326
NC_000915	344522	344772	HP0329
NC_000915	344953	345222	HP0329
NC_000915	347606	347771	HP0333
NC_000915	347848	347942	HP0333
NC_000915	350699	350961	HP0339
NC_000915	351632	351926	HP0342
NC_000915	357559	358061	HP0349
NC_000915	358325	358586	HP0350
NC_000915	359148	359592	HP0351
NC_000915	364020	364398	HP0354
NC_000915	367318	367751	HP0357
NC_000915	371481	371824	HP0361
NC_000915	372084	372523	HP0362
NC_000915	372964	373223	HP0363
NC_000915	373261	373592	HP0363
NC_000915	375534	375745	HP0366
NC_000915	388335	388540	HP0378
NC_000915	394214	394278	HP0383
NC_000915	397675	398081	HP0388
NC_000915	399370	399797	HP0390
NC_000915	409033	409294	HP0398
NC_000915	409706	409782	HP0399
NC_000915	410825	411101	HP0399
NC_000915	411443	411768	HP0400
NC_000915	413340	413665	HP0402
NC_000915	415062	415369	HP0402
NC_000915	416701	416947	HP0404
NC_000915	417345	417787	HP0405
NC_000915	418393	418496	HP0406
NC_000915	426979	427076	HP0414
NC_000915	429602	429966	HP0416

NC_000915	430007	430078	HP0416
NC_000915	430898	431265	HP0417
NC_000915	435116	435532	HP0421
NC_000915	437633	437784	HP0422
NC_000915	441193	441440	HP0425
NC_000915	457407	458128	HP0440
NC_000915	465205	465587	HP0447
NC_000915	466275	466510	HP0448
NC_000915	468416	468662	HP0451
NC_000915	471452	471869	HP0453
NC_000915	471925	472218	HP0453
NC_000915	482819	483147	HP0464
NC_000915	485132	485328	HP0464
NC_000915	485678	485904	HP0464
NC_000915	487048	487349	HP0465
NC_000915	487385	487874	HP0465
NC_000915	487933	488302	HP0466
NC_000915	488473	488677	HP0466
NC_000915	489251	489441	HP0468
NC_000915	491430	491662	HP0470
NC_000915	491706	491872	HP0470
NC_000915	492420	492697	HP0470
NC_000915	494390	494939	HP0472
NC_000915	495941	496238	HP0474
NC_000915	497521	497891	HP0476
NC_000915	501188	501462	HP0478
NC_000915	505373	505686	HP0482
NC_000915	523544	523700	HP0497
NC_000915	525005	525134	HP0498
NC_000915	531423	531718	HP0505
NC_000915	534062	534201	HP0508
NC_000915	539972	540398	HP0513
NC_000915	540439	540659	HP0513
NC_000915	541679	541949	HP0513
NC_000915	542009	542405	HP0514
NC_000915	545917	546002	HP0518

NC_000915	546489	547094	HP0519
NC_000915	551488	551887	HP0524
NC_000915	553687	553922	HP0526
NC_000915	556188	556615	HP0527
NC_000915	561239	561572	HP0528
NC_000915	561680	562240	HP0529
NC_000915	563419	563993	HP0530
NC_000915	564421	565037	HP0531
NC_000915	566332	566693	HP0534
NC_000915	567168	567493	HP0535
NC_000915	571310	571583	HP0539
NC_000915	575313	576186	HP0544
NC_000915	576363	576712	HP0544
NC_000915	587251	587527	HP0552
NC_000915	590683	591061	HP0556
NC_000915	591724	592439	HP0557
NC_000915	593321	593616	HP0558
NC_000915	594171	594273	HP0559
NC_000915	595016	595247	HP0561
NC_000915	597616	597940	HP0565
NC_000915	602303	602676	HP0570
NC_000915	604551	604780	HP0572
NC_000915	605318	605737	HP0574
NC_000915	613103	613578	HP0582
NC_000915	614858	615183	HP0584
NC_000915	628704	629038	HP0595
NC_000915	634050	634279	HP0599
NC_000915	634313	634419	HP0599
NC_000915	635442	635932	HP0600
NC_000915	637289	637383	HP0601
NC_000915	639068	639378	HP0602
NC_000915	646838	647002	HP0608
NC_000915	658734	658965	HP0614
NC_000915	659821	660296	HP0615
NC_000915	664090	664396	HP0618
NC_000915	668383	668698	HP0621

NC_000915	669845	669981	HP0623
NC_000915	675641	676047	HP0629
NC_000915	681576	682018	HP0634
NC_000915	683275	683592	HP0635
NC_000915	691274	691581	HP0645
NC_000915	695504	696012	HP0649
NC_000915	696062	696326	HP0650
NC_000915	710343	710500	HP0662
NC_000915	711840	712255	HP0664
NC_000915	712583	713073	HP0665
NC_000915	719861	720393	HP0670
NC_000915	722804	723057	HP0672
NC_000915	724866	725065	HP0674
NC_000915	732159	732513	HP0682
NC_000915	736161	736382	HP0686
NC_000915	743319	743580	HP0693
NC_000915	748765	749324	HP0696
NC_000915	752549	752745	HP0701
NC_000915	762808	763267	HP0709
NC_000915	766157	766390	HP0711
NC_000915	773728	773939	HP0721
NC_000915	794040	794302	HP0739
NC_000915	798735	798905	HP0743
NC_000915	801910	802058	HP0746
NC_000915	809666	809937	HP0755
NC_000915	809993	810267	HP0755
NC_000915	810370	810748	HP0757
NC_000915	811715	812383	HP0758
NC_000915	813548	813910	HP0759
NC_000915	815553	816147	HP0761
NC_000915	817931	818133	HP0764
NC_000915	819099	819378	HP0765
NC_000915	819409	819836	HP0766
NC_000915	819925	820213	HP0766
NC_000915	822798	823195	HP0770
NC_000915	823318	823477	HP0771

NC_000915	826297	826409	HP0773
NC_000915	831369	831751	HP0778
NC_000915	835101	835587	HP0781
NC_000915	837900	838228	HP0783
NC_000915	839632	839954	HP0786
NC_000915	842312	842790	HP0787
NC_000915	844821	844938	HP0788
NC_000915	847584	847882	HP0791
NC_000915	848115	848599	HP0791
NC_000915	850760	850974	HP0793
NC_000915	851070	851296	HP0794
NC_000915	857557	858127	HP0803
NC_000915	858233	858360	HP0804
NC_000915	859765	860171	HP0805
NC_000915	867029	867507	HP0813
NC_000915	869926	870322	HP0817
NC_000915	872480	872733	HP0819
NC_000915	872930	873353	HP0820
NC_000915	876506	876789	HP0823
NC_000915	876903	877108	HP0824
NC_000915	879410	879638	HP0827
NC_000915	880025	880542	HP0828
NC_000915	888470	888690	HP0836
NC_000915	888960	889051	HP0837
NC_000915	891218	891388	HP0839
NC_000915	893756	894062	HP0842
NC_000915	894148	894494	HP0842
NC_000915	895836	895969	HP0844
NC_000915	898918	899443	HP0846
NC_000915	907123	907489	HP0854
NC_000915	909944	910335	HP0856
NC_000915	910541	910836	HP0857
NC_000915	912335	912447	HP0859
NC_000915	915419	915893	HP0863
NC_000915	916013	916527	HP0863
NC_000915	919264	919490	HP0867

NC_000915	922840	923104	HP0871
NC_000915	923868	924139	HP0872
NC_000915	930747	931075	HP0879
NC_000915	932856	933115	HP0883
NC_000915	933177	933328	HP0883
NC_000915	933394	933857	HP0884
NC_000915	934470	934742	HP0884
NC_000915	937838	938190	HP0886
NC_000915	942346	942720	HP0888
NC_000915	946427	946611	HP0894
NC_000915	950227	950360	HP0897
NC_000915	963697	963773	HP0911
NC_000915	964084	964571	HP0911
NC_000915	969287	969691	HP0914
NC_000915	969741	969859	HP0914
NC_000915	975003	975682	HP0919
NC_000915	994172	994373	HP0933
NC_000915	995566	995987	HP0935
NC_000915	997101	997701	HP0936
NC_000915	999225	999413	HP0938
NC_000915	1000759	1001003	HP0940
NC_000915	1001336	1001949	HP0941
NC_000915	1003728	1004032	HP0943
NC_000915	1004307	1004458	HP0943
NC_000915	1004970	1005306	HP0944
NC_000915	1005801	1006038	HP0944a
NC_000915	1011804	1012177	HP0951
NC_000915	1015890	1016317	HP0957
NC_000915	1020161	1020276	HP0961
NC_000915	1020454	1020650	HP0962
NC_000915	1021742	1021911	HP0963
NC_000915	1024407	1024567	HP0965
NC_000915	1026447	1026657	HP0967
NC_000915	1032097	1032478	HP0971
NC_000915	1033052	1033407	HP0972
NC_000915	1033472	1033884	HP0972

NC_000915	1034744	1034969	HP0973
NC_000915	1035031	1035197	HP0973
NC_000915	1040627	1041110	HP0978
NC_000915	1051287	1051472	HP0988
NC_000915	1055621	1055829	HP0994
NC_000915	1058567	1058764	HP0996
NC_000915	1061684	1062105	HP0998
NC_000915	1062696	1063273	HP1000
NC_000915	1063435	1063691	HP1001
NC_000915	1064579	1064865	HP1002
NC_000915	1068092	1068362	HP1006
NC_000915	1070070	1070167	HP1008
NC_000915	1073747	1073937	HP1010
NC_000915	1077507	1077607	HP1013
NC_000915	1077878	1078203	HP1014
NC_000915	1079526	1079615	HP1016
NC_000915	1082513	1082727	HP1019
NC_000915	1085696	1086096	HP1022
NC_000915	1088600	1088766	HP1025
NC_000915	1089778	1090034	HP1026
NC_000915	1090378	1090496	HP1027
NC_000915	1091219	1091682	HP1029
NC_000915	1093317	1093639	HP1031
NC_000915	1101723	1101913	HP1041
NC_000915	1103157	1103563	HP1042
NC_000915	1107082	1107262	HP1045
NC_000915	1109196	1109494	HP1046
NC_000915	1112679	1112781	HP1048
NC_000915	1113953	1114376	HP1051
NC_000915	1116345	1116514	HP1054
NC_000915	1124840	1125113	HP1065
NC_000915	1126656	1126891	HP1068
NC_000915	1127734	1127994	HP1069
NC_000915	1129922	1130004	HP1071
NC_000915	1132791	1132946	HP1073
NC_000915	1135600	1135871	HP1076

NC_000915	1138149	1138495	HP1079
NC_000915	1138897	1139274	HP1080
NC_000915	1148973	1149187	HP1089
NC_000915	1152212	1152543	HP1090
NC_000915	1153779	1153960	HP1091
NC_000915	1154694	1154869	HP1091
NC_000915	1158084	1158505	HP1096
NC_000915	1166498	1166655	HP1105
NC_000915	1170806	1170873	HP1109
NC_000915	1170927	1171077	HP1109
NC_000915	1173477	1173861	HP1112
NC_000915	1174043	1174284	HP1112
NC_000915	1174394	1174721	HP1112
NC_000915	1177137	1177472	HP1114
NC_000915	1181738	1181976	HP1117
NC_000915	1187736	1187938	HP1121
NC_000915	1188624	1188715	HP1123
NC_000915	1191488	1191945	HP1126
NC_000915	1193310	1193695	HP1130
NC_000915	1199218	1199653	HP1137
NC_000915	1201115	1201308	HP1139
NC_000915	1201480	1201820	HP1140
NC_000915	1217000	1217500	HP1153
NC_000915	1227036	1227441	HP1161
NC_000915	1231304	1231489	HP1166
NC_000915	1236909	1237113	HP1169
NC_000915	1237337	1237411	HP1170
NC_000915	1237931	1238373	HP1171
NC_000915	1246053	1246369	HP1178
NC_000915	1250863	1251116	HP1182
NC_000915	1252456	1252759	HP1183
NC_000915	1262097	1262393	HP1191
NC_000915	1262487	1262920	HP1191
NC_000915	1268024	1268205	HP1197
NC_000915	1278505	1278953	HP1201
NC_000915	1283324	1283823	HP1206

NC_000915	1287311	1287484	HP1210
NC_000915	1290705	1291093	HP1214
NC_000915	1293588	1293872	HP1217
NC_000915	1295052	1295159	HP1218
NC_000915	1295199	1295391	HP1218
NC_000915	1301546	1301688	HP1226
NC_000915	1304802	1305336	HP1230
NC_000915	1305466	1305932	HP1231
NC_000915	1308085	1308592	HP1234
NC_000915	1308723	1308981	HP1234
NC_000915	1316882	1317051	HP1242
NC_000915	1320697	1321038	HP1245
NC_000915	1323240	1323634	HP1248
NC_000915	1324883	1325418	HP1249
NC_000915	1327169	1327568	HP1252
NC_000915	1342119	1342789	HP1269
NC_000915	1343454	1343645	HP1271
NC_000915	1345308	1345576	HP1272
NC_000915	1351730	1352015	HP1276
NC_000915	1356738	1356953	HP1281
NC_000915	1359894	1360354	HP1283
NC_000915	1362133	1362349	HP1285
NC_000915	1363199	1363711	HP1287
NC_000915	1369403	1369725	HP1295
NC_000915	1370618	1370862	HP1299
NC_000915	1373120	1373205	HP1302
NC_000915	1373542	1373762	HP1303
NC_000915	1377988	1378100	HP1314
NC_000915	1380913	1381004	HP1320
NC_000915	1381105	1381187	HP1320
NC_000915	1383358	1383689	HP1323
NC_000915	1385782	1386064	HP1326
NC_000915	1391144	1391455	HP1329
NC_000915	1392295	1392495	HP1331
NC_000915	1393798	1394101	HP1333
NC_000915	1396544	1396851	HP1335

NC_000915	1398076	1398202	HP1337
NC_000915	1398406	1398773	HP1338
NC_000915	1404034	1404242	HP1344
NC_000915	1404307	1404701	HP1344
NC_000915	1408652	1409376	HP1349
NC_000915	1410347	1410786	HP1350
NC_000915	1410860	1411136	HP1350
NC_000915	1411540	1412039	HP1351
NC_000915	1417966	1418060	HP1356
NC_000915	1420222	1420538	HP1359
NC_000915	1439500	1439757	HP1375
NC_000915	1439835	1440194	HP1375
NC_000915	1440901	1441274	HP1377
NC_000915	1448465	1448550	HP1386
NC_000915	1451036	1451490	HP1390
NC_000915	1451807	1452053	HP1391
NC_000915	1455330	1455763	HP1394
NC_000915	1456023	1456117	HP1395
NC_000915	1457699	1458465	HP1397
NC_000915	1461441	1461832	HP1400
NC_000915	1463488	1463949	HP1400
NC_000915	1467843	1468219	HP1403
NC_000915	1469991	1470103	HP1403
NC_000915	1471413	1471901	HP1406
NC_000915	1472305	1472411	HP1407
NC_000915	1491052	1491686	HP1421
NC_000915	1495383	1495675	HP1424
NC_000915	1500414	1501468	HP1430
NC_000915	1504984	1505478	HP1433
NC_000915	1505690	1505945	HP1433
NC_000915	1506158	1506399	HP1434
NC_000915	1506509	1506963	HP1434
NC_000915	1513703	1513922	HP1440
NC_000915	1514042	1514508	HP1441
NC_000915	1518702	1518907	HP1450
NC_000915	1551682	1552025	HP1479

NC_000915	1553479	1554095	HP1479
NC_000915	1554239	1554866	HP1480
NC_000915	1556767	1557257	HP1483
NC_000915	1557297	1557730	HP1484
NC_000915	1559814	1560455	HP1487
NC_000915	1566597	1566727	HP1493
NC_000915	1566765	1567093	HP1493
NC_000915	1567365	1567460	HP1494
NC_000915	1572432	1572562	HP1499
NC_000915	1574912	1575645	HP1503
NC_000915	1582409	1582503	HP1508
NC_000915	1583275	1583534	HP1509
NC_000915	1586946	1587022	HP1512
NC_000915	1598455	1599164	HP1521
NC_000915	1603617	1603912	HP1523
NC_000915	1604072	1604388	HP1524
NC_000915	1606198	1606588	HP1527
NC_000915	1609151	1609335	HP1530
NC_000915	1609367	1609491	HP1530
NC_000915	1614361	1614764	HP1535
NC_000915	1619538	1620295	HP1541
NC_000915	1621953	1622038	HP1542
NC_000915	1621953	1622038	HP1543
NC_000915	1628475	1628749	HP1549
NC_000915	1628829	1629246	HP1549
NC_000915	1631259	1631650	HP1552
NC_000915	1636517	1636755	HP1555
NC_000915	1638022	1638497	HP1556
NC_000915	1640037	1640328	HP1557
NC_000915	1645303	1645820	HP1563
NC_000915	1653484	1653716	HP1572
NC_000915	1654700	1655013	HP1574
NC_000915	1655316	1655571	HP1575
NC_000915	1661399	1661849	HP1583
NC_000915	1662677	1663071	HP1584
NC_000915	1663172	1663361	HP1584

NC_000915	1667177	1667672	HP1589

9.3

CHR	START	END	GENE_ID
NC_014256	154	1527	HPB8_1_154_1527
NC_014256	1780	3186	HPB8_2_1780_3186
NC_014256	5193	7038	HPB8_6_5193_7038
NC_014256	7126	8537	HPB8_7_7126_8537
NC_014256	8572	12060	HPB8_9_8572_12060
NC_014256	12178	13128	HPB8_10_12178_13128
NC_014256	13346	15741	HPB8_10_13346_15741
NC_014256	15784	16064	HPB8_11_15784_16064
NC_014256	16575	16955	HPB8_12_16575_16955
NC_014256	17400	17611	HPB8_13_17400_17611
NC_014256	17997	18992	HPB8_14_17997_18992
NC_014256	19439	20448	HPB8_15_19439_20448
NC_014256	20675	23264	HPB8_17_20675_23264
NC_014256	23485	23755	HPB8_18_23485_23755
NC_014256	23787	24133	HPB8_19_23787_24133
NC_014256	24232	24779	HPB8_20_24232_24779
NC_014256	30591	32878	HPB8_26_30591_32878
NC_014256	32928	33357	HPB8_27_32928_33357
NC_014256	33551	34675	HPB8_28_33551_34675
NC_014256	34837	34970	HPB8_29_34837_34970
NC_014256	35098	36365	HPB8_29_35098_36365
NC_014256	36847	37146	HPB8_30_36847_37146
NC_014256	37458	37711	HPB8_31_37458_37711
NC_014256	38109	38230	HPB8_31_38109_38230
NC_014256	38428	39392	HPB8_32_38428_39392
NC_014256	40631	41553	HPB8_35_40631_41553
NC_014256	41600	42831	HPB8_36_41600_42831
NC_014256	42926	43482	HPB8_37_42926_43482
NC_014256	43551	43793	HPB8_38_43551_43793
NC_014256	43942	45534	HPB8_39_43942_45534
NC_014256	45646	46101	HPB8_40_45646_46101
NC_014256	66762	69290	HPB8_61_66762_69290
NC_014256	69444	70186	HPB8_62_69444_70186
NC_014256	70241	70649	HPB8_63_70241_70649
NC_014256	76396	77448	HPB8_71_76396_77448
NC_014256	77596	81944	HPB8_72_77596_81944
NC_014256	82057	82357	HPB8_74_82057_82357
NC_014256	82741	83355	HPB8_77_82741_83355
NC_014256	85587	85728	HPB8_81_85587_85728
NC_014256	85806	88235	HPB8_83_85806_88235

NC_014256	93324	93637	HPB8_90_93324_93637
NC_014256	95188	95317	HPB8_93_95188_95317
NC_014256	95361	95582	HPB8_94_95361_95582
NC_014256	95615	96098	HPB8_95_95615_96098
NC_014256	96222	96968	HPB8_96_96222_96968
NC_014256	99457	99798	HPB8_97_99457_99798
NC_014256	100186	100619	HPB8_98_100186_100619
NC_014256	100667	100977	HPB8_99_100667_100977
NC_014256	101290	101644	HPB8_101_101290_101644
NC_014256	101774	102014	HPB8_102_101774_102014
NC_014256	102087	102378	HPB8_103_102087_102378
NC_014256	102584	102841	HPB8_105_102584_102841
NC_014256	103035	103760	HPB8_106_103035_103760
NC_014256	103812	104662	HPB8_107_103812_104662
NC_014256	104716	105039	HPB8_108_104716_105039
NC_014256	105307	105609	HPB8_108_105307_105609
NC_014256	106241	106536	HPB8_112_106241_106536
NC_014256	106796	107605	HPB8_113_106796_107605
NC_014256	107649	109499	HPB8_114_107649_109499
NC_014256	111921	112142	HPB8_117_111921_112142
NC_014256	112442	112751	HPB8_118_112442_112751
NC_014256	112982	113251	HPB8_119_112982_113251
NC_014256	113401	114021	HPB8_120_113401_114021
NC_014256	118103	119649	HPB8_124_118103_119649
NC_014256	119688	120467	HPB8_126_119688_120467
NC_014256	121531	121665	HPB8_128_121531_121665
NC_014256	121757	121802	HPB8_128_121757_121802
NC_014256	122042	123079	HPB8_129_122042_123079
NC_014256	126051	126968	HPB8_133_126051_126968
NC_014256	127107	128267	HPB8_134_127107_128267
NC_014256	130155	130295	HPB8_137_130155_130295
NC_014256	130332	130453	HPB8_137_130332_130453
NC_014256	130570	130951	HPB8_137_130570_130951
NC_014256	130998	131089	HPB8_137_130998_131089
NC_014256	131244	132184	HPB8_137_131244_132184
NC_014256	135792	136084	HPB8_143_135792_136084
NC_014256	136169	137245	HPB8_144_136169_137245
NC_014256	137297	138428	HPB8_145_137297_138428
NC_014256	138553	139647	HPB8_147_138553_139647
NC_014256	139890	140282	HPB8_148_139890_140282
NC_014256	144820	145994	HPB8_152_144820_145994
NC_014256	146063	146440	HPB8_153_146063_146440
NC_014256	146615	148279	HPB8_154_146615_148279
NC_014256	148310	148898	HPB8_156_148310_148898

NC_014256	148932	149476	HPB8_157_148932_149476
NC_014256	149722	150806	HPB8_158_149722_150806
NC_014256	150973	151287	HPB8_159_150973_151287
NC_014256	151343	151899	HPB8_160_151343_151899
NC_014256	152064	152581	HPB8_161_152064_152581
NC_014256	152709	152845	HPB8_162_152709_152845
NC_014256	152882	153681	HPB8_163_152882_153681
NC_014256	161666	161790	HPB8_181_161666_161790
NC_014256	168526	169128	HPB8_194_168526_169128
NC_014256	169254	169789	HPB8_195_169254_169789
NC_014256	169915	170644	HPB8_196_169915_170644
NC_014256	170720	171770	HPB8_197_170720_171770
NC_014256	178956	179016	HPB8_204_178956_179016
NC_014256	183601	184183	HPB8_207_183601_184183
NC_014256	185558	185931	HPB8_208_185558_185931
NC_014256	186120	186440	HPB8_209_186120_186440
NC_014256	200243	200974	HPB8_226_200243_200974
NC_014256	204874	205354	HPB8_230_204874_205354
NC_014256	205611	206225	HPB8_231_205611_206225
NC_014256	206263	207162	HPB8_232_206263_207162
NC_014256	207212	208179	HPB8_232_207212_208179
NC_014256	208215	208511	HPB8_233_208215_208511
NC_014256	208627	208829	HPB8_233_208627_208829
NC_014256	208894	209207	HPB8_233_208894_209207
NC_014256	210419	210620	HPB8_236_210419_210620
NC_014256	213815	214044	HPB8_241_213815_214044
NC_014256	214163	215332	HPB8_242_214163_215332
NC_014256	217740	218740	HPB8_246_217740_218740
NC_014256	220666	221910	HPB8_249_220666_221910
NC_014256	221984	222166	HPB8_250_221984_222166
NC_014256	222213	222457	HPB8_250_222213_222457
NC_014256	222555	222810	HPB8_250_222555_222810
NC_014256	223174	223605	HPB8_252_223174_223605
NC_014256	227906	228196	HPB8_258_227906_228196
NC_014256	228291	229328	HPB8_259_228291_229328
NC_014256	229405	229771	HPB8_260_229405_229771
NC_014256	230915	233735	HPB8_263_230915_233735
NC_014256	235249	235315	HPB8_266_235249_235315
NC_014256	235745	236058	HPB8_268_235745_236058
NC_014256	236798	238060	HPB8_270_236798_238060
NC_014256	239402	240816	HPB8_273_239402_240816
NC_014256	243786	244092	HPB8_276_243786_244092
NC_014256	244713	244836	HPB8_278_244713_244836
NC_014256	244875	245212	HPB8_278_244875_245212

NC_014256	249018	249577	HPB8_284_249018_249577
NC_014256	250278	251474	HPB8_285_250278_251474
NC_014256	251517	251657	HPB8_286_251517_251657
NC_014256	251924	252103	HPB8_287_251924_252103
NC_014256	253148	253840	HPB8_290_253148_253840
NC_014256	253961	254449	HPB8_291_253961_254449
NC_014256	254497	254871	HPB8_292_254497_254871
NC_014256	257755	257930	HPB8_293_257755_257930
NC_014256	263858	264260	HPB8_294_263858_264260
NC_014256	264341	266835	HPB8_296_264341_266835
NC_014256	267396	268325	HPB8_297_267396_268325
NC_014256	268604	268756	HPB8_299_268604_268756
NC_014256	268904	269928	HPB8_300_268904_269928
NC_014256	270013	270714	HPB8_301_270013_270714
NC_014256	274074	274795	HPB8_304_274074_274795
NC_014256	275071	276246	HPB8_306_275071_276246
NC_014256	278984	279734	HPB8_310_278984_279734
NC_014256	279834	281084	HPB8_311_279834_281084
NC_014256	281401	282630	HPB8_312_281401_282630
NC_014256	285409	285736	HPB8_316_285409_285736
NC_014256	286032	286082	HPB8_316_286032_286082
NC_014256	289384	289515	HPB8_320_289384_289515
NC_014256	289581	290885	HPB8_321_289581_290885
NC_014256	291032	292186	HPB8_322_291032_292186
NC_014256	292469	293007	HPB8_324_292469_293007
NC_014256	293129	293962	HPB8_325_293129_293962
NC_014256	294135	294756	HPB8_326_294135_294756
NC_014256	296296	298350	HPB8_329_296296_298350
NC_014256	298674	300073	HPB8_330_298674_300073
NC_014256	300391	302028	HPB8_331_300391_302028
NC_014256	302092	302965	HPB8_332_302092_302965
NC_014256	303037	303252	HPB8_332_303037_303252
NC_014256	303315	304230	HPB8_333_303315_304230
NC_014256	304266	304453	HPB8_334_304266_304453
NC_014256	304578	305054	HPB8_336_304578_305054
NC_014256	305250	305743	HPB8_337_305250_305743
NC_014256	305830	306208	HPB8_338_305830_306208
NC_014256	306318	306501	HPB8_339_306318_306501
NC_014256	315164	317767	HPB8_348_315164_317767
NC_014256	317798	319090	HPB8_349_317798_319090
NC_014256	320374	320986	HPB8_353_320374_320986
NC_014256	321107	321430	HPB8_354_321107_321430
NC_014256	322220	322378	HPB8_355_322220_322378
NC_014256	322416	322838	HPB8_356_322416_322838

NC_014256	325440	325853	HPB8_358_325440_325853
NC_014256	326074	327362	HPB8_359_326074_327362
NC_014256	327411	327680	HPB8_360_327411_327680
NC_014256	327873	328651	HPB8_360_327873_328651
NC_014256	328845	329200	HPB8_360_328845_329200
NC_014256	329253	329425	HPB8_360_329253_329425
NC_014256	329595	329770	HPB8_360_329595_329770
NC_014256	333335	333769	HPB8_364_333335_333769
NC_014256	333818	334255	HPB8_365_333818_334255
NC_014256	334311	334813	HPB8_366_334311_334813
NC_014256	337318	338671	HPB8_369_337318_338671
NC_014256	338729	339100	HPB8_370_338729_339100
NC_014256	339134	339680	HPB8_371_339134_339680
NC_014256	339732	340097	HPB8_372_339732_340097
NC_014256	340152	340901	HPB8_373_340152_340901
NC_014256	340942	342074	HPB8_374_340942_342074
NC_014256	343789	344346	HPB8_377_343789_344346
NC_014256	344458	344584	HPB8_378_344458_344584
NC_014256	344653	344853	HPB8_379_344653_344853
NC_014256	345019	345833	HPB8_380_345019_345833
NC_014256	346074	346470	HPB8_381_346074_346470
NC_014256	346510	348320	HPB8_382_346510_348320
NC_014256	348394	350204	HPB8_384_348394_350204
NC_014256	350472	351242	HPB8_385_350472_351242
NC_014256	351520	351625	HPB8_386_351520_351625
NC_014256	351782	352448	HPB8_386_351782_352448
NC_014256	355762	357426	HPB8_390_355762_357426
NC_014256	357482	357725	HPB8_390_357482_357725
NC_014256	357803	358532	HPB8_391_357803_358532
NC_014256	358638	359960	HPB8_392_358638_359960
NC_014256	362679	363228	HPB8_396_362679_363228
NC_014256	363507	364142	HPB8_397_363507_364142
NC_014256	364229	365457	HPB8_398_364229_365457
NC_014256	365657	366663	HPB8_399_365657_366663
NC_014256	366851	367870	HPB8_400_366851_367870
NC_014256	368042	368996	HPB8_401_368042_368996
NC_014256	373616	374480	HPB8_406_373616_374480
NC_014256	374730	375468	HPB8_407_374730_375468
NC_014256	375507	376951	HPB8_408_375507_376951
NC_014256	384302	384706	HPB8_416_384302_384706
NC_014256	385105	385567	HPB8_417_385105_385567
NC_014256	387763	388319	HPB8_420_387763_388319
NC_014256	388544	390532	HPB8_421_388544_390532
NC_014256	390667	391347	HPB8_422_390667_391347

NC_014256	391439	391779	HPB8_423_391439_391779
NC_014256	391851	392839	HPB8_424_391851_392839
NC_014256	392915	396390	HPB8_426_392915_396390
NC_014256	397688	398984	HPB8_429_397688_398984
NC_014256	399117	399899	HPB8_430_399117_399899
NC_014256	399930	401114	HPB8_431_399930_401114
NC_014256	401332	401835	HPB8_432_401332_401835
NC_014256	401907	402645	HPB8_433_401907_402645
NC_014256	402848	402994	HPB8_435_402848_402994
NC_014256	403088	403288	HPB8_436_403088_403288
NC_014256	403328	405142	HPB8_437_403328_405142
NC_014256	405253	405926	HPB8_438_405253_405926
NC_014256	405999	406630	HPB8_439_405999_406630
NC_014256	406854	407354	HPB8_440_406854_407354
NC_014256	407614	410556	HPB8_442_407614_410556
NC_014256	413056	414836	HPB8_447_413056_414836
NC_014256	414896	416420	HPB8_448_414896_416420
NC_014256	416486	416966	HPB8_449_416486_416966
NC_014256	417100	418665	HPB8_450_417100_418665
NC_014256	424335	424499	HPB8_456_424335_424499
NC_014256	424599	425910	HPB8_457_424599_425910
NC_014256	425947	426537	HPB8_458_425947_426537
NC_014256	426641	429031	HPB8_459_426641_429031
NC_014256	431272	432054	HPB8_463_431272_432054
NC_014256	436496	437611	HPB8_467_436496_437611
NC_014256	439046	439447	HPB8_468_439046_439447
NC_014256	440717	441842	HPB8_471_440717_441842
NC_014256	441920	443748	HPB8_472_441920_443748
NC_014256	444522	445179	HPB8_475_444522_445179
NC_014256	445446	447228	HPB8_476_445446_447228
NC_014256	447408	447722	HPB8_477_447408_447722
NC_014256	452485	452648	HPB8_482_452485_452648
NC_014256	452709	452755	HPB8_482_452709_452755
NC_014256	453450	453550	HPB8_484_453450_453550
NC_014256	453652	454107	HPB8_484_453652_454107
NC_014256	454400	454929	HPB8_484_454400_454929
NC_014256	455488	455736	HPB8_485_455488_455736
NC_014256	455994	456330	HPB8_486_455994_456330
NC_014256	456900	457157	HPB8_486_456900_457157
NC_014256	458570	459764	HPB8_487_458570_459764
NC_014256	460190	460437	HPB8_488_460190_460437
NC_014256	460601	461167	HPB8_489_460601_461167
NC_014256	461444	462117	HPB8_490_461444_462117
NC_014256	462243	462717	HPB8_491_462243_462717

NC_014256	462906	463501	HPB8_492_462906_463501
NC_014256	463629	463884	HPB8_493_463629_463884
NC_014256	464776	464924	HPB8_496_464776_464924
NC_014256	465243	465772	HPB8_497_465243_465772
NC_014256	465884	466269	HPB8_498_465884_466269
NC_014256	467038	467685	HPB8_501_467038_467685
NC_014256	467939	468274	HPB8_501_467939_468274
NC_014256	468572	468910	HPB8_501_468572_468910
NC_014256	469222	470406	HPB8_503_469222_470406
NC_014256	470644	470727	HPB8_504_470644_470727
NC_014256	470816	471472	HPB8_505_470816_471472
NC_014256	471747	472107	HPB8_506_471747_472107
NC_014256	472436	479082	HPB8_506_472436_479082
NC_014256	479458	479590	HPB8_507_479458_479590
NC_014256	479641	479826	HPB8_508_479641_479826
NC_014256	479902	480500	HPB8_509_479902_480500
NC_014256	480631	481352	HPB8_510_480631_481352
NC_014256	484238	485514	HPB8_512_484238_485514
NC_014256	486679	487173	HPB8_515_486679_487173
NC_014256	488451	488922	HPB8_517_488451_488922
NC_014256	489238	489657	HPB8_518_489238_489657
NC_014256	489714	490200	HPB8_518_489714_490200
NC_014256	490731	491969	HPB8_519_490731_491969
NC_014256	493724	493920	HPB8_521_493724_493920
NC_014256	494189	494383	HPB8_521_494189_494383
NC_014256	494711	494937	HPB8_522_494711_494937
NC_014256	495334	495487	HPB8_522_495334_495487
NC_014256	496776	497081	HPB8_525_496776_497081
NC_014256	497113	497422	HPB8_525_497113_497422
NC_014256	500601	501608	HPB8_528_500601_501608
NC_014256	505645	506313	HPB8_532_505645_506313
NC_014256	506360	506737	HPB8_533_506360_506737
NC_014256	506866	507152	HPB8_534_506866_507152
NC_014256	507660	508190	HPB8_535_507660_508190
NC_014256	508251	508697	HPB8_536_508251_508697
NC_014256	508766	510812	HPB8_537_508766_510812
NC_014256	510846	512962	HPB8_538_510846_512962
NC_014256	513025	513066	HPB8_539_513025_513066
NC_014256	513353	513391	HPB8_539_513353_513391
NC_014256	513577	514418	HPB8_540_513577_514418
NC_014256	515320	515645	HPB8_542_515320_515645
NC_014256	515733	516257	HPB8_543_515733_516257
NC_014256	516295	517682	HPB8_543_516295_517682
NC_014256	517921	519122	HPB8_544_517921_519122

NC_014256	519196	520774	HPB8_545_519196_520774
NC_014256	520910	521407	HPB8_546_520910_521407
NC_014256	522090	522931	HPB8_548_522090_522931
NC_014256	522989	523441	HPB8_548_522989_523441
NC_014256	524489	524613	HPB8_548_524489_524613
NC_014256	525153	525383	HPB8_551_525153_525383
NC_014256	525768	525998	HPB8_553_525768_525998
NC_014256	526398	526448	HPB8_554_526398_526448
NC_014256	527245	527467	HPB8_555_527245_527467
NC_014256	527696	528268	HPB8_556_527696_528268
NC_014256	529230	529881	HPB8_558_529230_529881
NC_014256	530456	530520	HPB8_560_530456_530520
NC_014256	530931	531598	HPB8_561_530931_531598
NC_014256	531695	531785	HPB8_562_531695_531785
NC_014256	532010	532567	HPB8_563_532010_532567
NC_014256	532657	533220	HPB8_564_532657_533220
NC_014256	533554	534700	HPB8_566_533554_534700
NC_014256	537921	539210	HPB8_569_537921_539210
NC_014256	542331	544429	HPB8_574_542331_544429
NC_014256	544479	545667	HPB8_575_544479_545667
NC_014256	545741	546751	HPB8_576_545741_546751
NC_014256	546796	550468	HPB8_577_546796_550468
NC_014256	550682	552287	HPB8_580_550682_552287
NC_014256	552330	552605	HPB8_581_552330_552605
NC_014256	556135	556373	HPB8_583_556135_556373
NC_014256	556514	556888	HPB8_584_556514_556888
NC_014256	559160	559869	HPB8_587_559160_559869
NC_014256	559952	560871	HPB8_588_559952_560871
NC_014256	561070	561792	HPB8_590_561070_561792
NC_014256	561952	562623	HPB8_591_561952_562623
NC_014256	562665	563291	HPB8_592_562665_563291
NC_014256	565287	566125	HPB8_596_565287_566125
NC_014256	566170	566577	HPB8_597_566170_566577
NC_014256	566632	567601	HPB8_598_566632_567601
NC_014256	568100	568412	HPB8_602_568100_568412
NC_014256	568784	570243	HPB8_603_568784_570243
NC_014256	571191	571535	HPB8_604_571191_571535
NC_014256	571582	572808	HPB8_605_571582_572808
NC_014256	572855	574154	HPB8_606_572855_574154
NC_014256	574221	575347	HPB8_607_574221_575347
NC_014256	576952	577299	HPB8_610_576952_577299
NC_014256	577384	577555	HPB8_611_577384_577555
NC_014256	577659	578469	HPB8_611_577659_578469
NC_014256	579109	579780	HPB8_613_579109_579780

NC_014256	579830	580278	HPB8_613_579830_580278
NC_014256	584771	585149	HPB8_618_584771_585149
NC_014256	591922	599022	HPB8_627_591922_599022
NC_014256	599197	600195	HPB8_628_599197_600195
NC_014256	600338	600949	HPB8_629_600338_600949
NC_014256	601162	604818	HPB8_630_601162_604818
NC_014256	605005	605823	HPB8_632_605005_605823
NC_014256	606158	607444	HPB8_634_606158_607444
NC_014256	607918	609317	HPB8_636_607918_609317
NC_014256	613235	613400	HPB8_640_613235_613400
NC_014256	616917	617361	HPB8_643_616917_617361
NC_014256	617420	619230	HPB8_644_617420_619230
NC_014256	619454	620252	HPB8_645_619454_620252
NC_014256	620303	621854	HPB8_646_620303_621854
NC_014256	622111	623604	HPB8_647_622111_623604
NC_014256	625480	626117	HPB8_652_625480_626117
NC_014256	626169	627433	HPB8_654_626169_627433
NC_014256	632724	633223	HPB8_661_632724_633223
NC_014256	636035	639973	HPB8_666_636035_639973
NC_014256	642747	643029	HPB8_668_642747_643029
NC_014256	643215	643439	HPB8_669_643215_643439
NC_014256	643530	644260	HPB8_669_643530_644260
NC_014256	644334	644992	HPB8_669_644334_644992
NC_014256	645047	645434	HPB8_670_645047_645434
NC_014256	646125	646399	HPB8_671_646125_646399
NC_014256	649551	649840	HPB8_674_649551_649840
NC_014256	650220	650961	HPB8_675_650220_650961
NC_014256	651102	651354	HPB8_675_651102_651354
NC_014256	651518	651648	HPB8_676_651518_651648
NC_014256	651792	651860	HPB8_677_651792_651860
NC_014256	651915	652628	HPB8_677_651915_652628
NC_014256	652674	653417	HPB8_678_652674_653417
NC_014256	655088	655263	HPB8_680_655088_655263
NC_014256	655333	655720	HPB8_681_655333_655720
NC_014256	655971	656601	HPB8_682_655971_656601
NC_014256	658164	658627	HPB8_683_658164_658627
NC_014256	658741	659698	HPB8_684_658741_659698
NC_014256	659771	659911	HPB8_685_659771_659911
NC_014256	660041	660875	HPB8_685_660041_660875
NC_014256	660926	663228	HPB8_686_660926_663228
NC_014256	663350	664189	HPB8_687_663350_664189
NC_014256	664253	664531	HPB8_688_664253_664531
NC_014256	666967	667365	HPB8_692_666967_667365
NC_014256	667400	667803	HPB8_692_667400_667803

NC_014256	668024	669376	HPB8_693_668024_669376
NC_014256	669425	670157	HPB8_694_669425_670157
NC_014256	670637	670855	HPB8_696_670637_670855
NC_014256	672885	674751	HPB8_699_672885_674751
NC_014256	676010	677314	HPB8_702_676010_677314
NC_014256	679307	680173	HPB8_705_679307_680173
NC_014256	680232	681362	HPB8_706_680232_681362
NC_014256	682552	682702	HPB8_708_682552_682702
NC_014256	682552	682702	HPB8_709_682552_682702
NC_014256	683744	683872	HPB8_710_683744_683872
NC_014256	684245	685022	HPB8_711_684245_685022
NC_014256	685232	685685	HPB8_712_685232_685685
NC_014256	686072	686482	HPB8_713_686072_686482
NC_014256	686539	686647	HPB8_713_686539_686647
NC_014256	686835	688425	HPB8_714_686835_688425
NC_014256	688486	689921	HPB8_715_688486_689921
NC_014256	691528	691703	HPB8_716_691528_691703
NC_014256	691783	691998	HPB8_716_691783_691998
NC_014256	693129	693171	HPB8_716_693129_693171
NC_014256	693367	693455	HPB8_716_693367_693455
NC_014256	693499	693650	HPB8_716_693499_693650
NC_014256	693859	695531	HPB8_716_693859_695531
NC_014256	695834	696214	HPB8_717_695834_696214
NC_014256	697765	699528	HPB8_719_697765_699528
NC_014256	699618	699833	HPB8_721_699618_699833
NC_014256	702753	703086	HPB8_725_702753_703086
NC_014256	703259	703754	HPB8_726_703259_703754
NC_014256	703875	704105	HPB8_726_703875_704105
NC_014256	704351	705782	HPB8_728_704351_705782
NC_014256	705923	706376	HPB8_729_705923_706376
NC_014256	706608	706948	HPB8_729_706608_706948
NC_014256	707238	707608	HPB8_729_707238_707608
NC_014256	712538	712595	HPB8_735_712538_712595
NC_014256	712681	713040	HPB8_735_712681_713040
NC_014256	713121	713395	HPB8_736_713121_713395
NC_014256	713569	713636	HPB8_736_713569_713636
NC_014256	714134	714646	HPB8_738_714134_714646
NC_014256	715093	715170	HPB8_739_715093_715170
NC_014256	715221	715396	HPB8_739_715221_715396
NC_014256	716234	716422	HPB8_740_716234_716422
NC_014256	716856	718850	HPB8_741_716856_718850
NC_014256	718928	720370	HPB8_741_718928_720370
NC_014256	721387	722149	HPB8_743_721387_722149
NC_014256	722190	723504	HPB8_744_722190_723504

NC_014256	723770	723944	HPB8_745_723770_723944
NC_014256	724020	724858	HPB8_746_724020_724858
NC_014256	724932	725557	HPB8_747_724932_725557
NC_014256	727450	727863	HPB8_750_727450_727863
NC_014256	730640	730871	HPB8_755_730640_730871
NC_014256	731102	731830	HPB8_756_731102_731830
NC_014256	731889	732096	HPB8_757_731889_732096
NC_014256	732204	733407	HPB8_758_732204_733407
NC_014256	733619	733853	HPB8_760_733619_733853
NC_014256	733892	734531	HPB8_761_733892_734531
NC_014256	734663	735459	HPB8_762_734663_735459
NC_014256	735629	736620	HPB8_763_735629_736620
NC_014256	736771	737621	HPB8_764_736771_737621
NC_014256	740324	740857	HPB8_767_740324_740857
NC_014256	740903	741442	HPB8_768_740903_741442
NC_014256	741556	741832	HPB8_769_741556_741832
NC_014256	741883	742338	HPB8_770_741883_742338
NC_014256	742379	742988	HPB8_771_742379_742988
NC_014256	744989	745702	HPB8_774_744989_745702
NC_014256	745758	746913	HPB8_774_745758_746913
NC_014256	747127	747322	HPB8_775_747127_747322
NC_014256	747533	748491	HPB8_776_747533_748491
NC_014256	751913	752552	HPB8_782_751913_752552
NC_014256	752629	752785	HPB8_783_752629_752785
NC_014256	752831	753597	HPB8_784_752831_753597
NC_014256	753629	753704	HPB8_784_753629_753704
NC_014256	753792	755343	HPB8_784_753792_755343
NC_014256	755477	756621	HPB8_785_755477_756621
NC_014256	758322	759085	HPB8_788_758322_759085
NC_014256	759128	759654	HPB8_789_759128_759654
NC_014256	760666	761300	HPB8_791_760666_761300
NC_014256	761691	763130	HPB8_793_761691_763130
NC_014256	767117	767477	HPB8_796_767117_767477
NC_014256	767569	768632	HPB8_797_767569_768632
NC_014256	768906	769426	HPB8_798_768906_769426
NC_014256	770010	770576	HPB8_799_770010_770576
NC_014256	770830	772350	HPB8_800_770830_772350
NC_014256	772511	773136	HPB8_801_772511_773136
NC_014256	773214	773718	HPB8_802_773214_773718
NC_014256	780828	790388	HPB8_808_780828_790388
NC_014256	790707	790820	HPB8_809_790707_790820
NC_014256	791263	791839	HPB8_810_791263_791839
NC_014256	792229	792542	HPB8_812_792229_792542
NC_014256	792660	792904	HPB8_813_792660_792904

NC_014256	792993	794612	HPB8_813_792993_794612
NC_014256	794697	795608	HPB8_814_794697_795608
NC_014256	798210	798494	HPB8_817_798210_798494
NC_014256	798675	798763	HPB8_817_798675_798763
NC_014256	799434	800139	HPB8_818_799434_800139
NC_014256	800501	800913	HPB8_819_800501_800913
NC_014256	808664	809648	HPB8_826_808664_809648
NC_014256	810216	811952	HPB8_828_810216_811952
NC_014256	812026	812597	HPB8_829_812026_812597
NC_014256	812769	813923	HPB8_831_812769_813923
NC_014256	816379	816880	HPB8_834_816379_816880
NC_014256	816963	817938	HPB8_835_816963_817938
NC_014256	817970	818420	HPB8_835_817970_818420
NC_014256	818453	818847	HPB8_836_818453_818847
NC_014256	818966	819418	HPB8_837_818966_819418
NC_014256	819599	820494	HPB8_838_819599_820494
NC_014256	820595	821281	HPB8_839_820595_821281
NC_014256	821374	822117	HPB8_841_821374_822117
NC_014256	822215	822541	HPB8_841_822215_822541
NC_014256	824761	826080	HPB8_845_824761_826080
NC_014256	826122	826359	HPB8_846_826122_826359
NC_014256	826400	827989	HPB8_847_826400_827989
NC_014256	830631	832033	HPB8_851_830631_832033
NC_014256	832078	832657	HPB8_852_832078_832657
NC_014256	834096	834719	HPB8_855_834096_834719
NC_014256	835026	835237	HPB8_856_835026_835237
NC_014256	835547	836209	HPB8_857_835547_836209
NC_014256	836333	836598	HPB8_857_836333_836598
NC_014256	856120	856338	HPB8_872_856120_856338
NC_014256	857071	857098	HPB8_873_857071_857098
NC_014256	858519	859331	HPB8_874_858519_859331
NC_014256	859490	860656	HPB8_876_859490_860656
NC_014256	862831	863871	HPB8_878_862831_863871
NC_014256	863970	864379	HPB8_879_863970_864379
NC_014256	864455	864787	HPB8_880_864455_864787
NC_014256	864896	865211	HPB8_880_864896_865211
NC_014256	865327	866266	HPB8_881_865327_866266
NC_014256	866336	868700	HPB8_882_866336_868700
NC_014256	868888	869131	HPB8_883_868888_869131
NC_014256	869236	869354	HPB8_883_869236_869354
NC_014256	869415	869713	HPB8_884_869415_869713
NC_014256	869935	871231	HPB8_885_869935_871231
NC_014256	871276	872014	HPB8_886_871276_872014
NC_014256	872109	874362	HPB8_887_872109_874362

NC_014256	874654	876582	HPB8_888_874654_876582
NC_014256	876725	877531	HPB8_889_876725_877531
NC_014256	880343	880629	HPB8_893_880343_880629
NC_014256	880660	881707	HPB8_893_880660_881707
NC_014256	881870	881945	HPB8_894_881870_881945
NC_014256	881979	882809	HPB8_894_881979_882809
NC_014256	888389	889045	HPB8_900_888389_889045
NC_014256	889829	892281	HPB8_902_889829_892281
NC_014256	894445	894865	HPB8_907_894445_894865
NC_014256	895548	898355	HPB8_909_895548_898355
NC_014256	898538	899359	HPB8_910_898538_899359
NC_014256	899535	900428	HPB8_912_899535_900428
NC_014256	900482	900826	HPB8_913_900482_900826
NC_014256	901030	901929	HPB8_914_901030_901929
NC_014256	904353	905516	HPB8_916_904353_905516
NC_014256	905716	906045	HPB8_918_905716_906045
NC_014256	909469	910340	HPB8_923_909469_910340
NC_014256	910449	911032	HPB8_924_910449_911032
NC_014256	914622	915636	HPB8_928_914622_915636
NC_014256	915714	917042	HPB8_929_915714_917042
NC_014256	917145	919107	HPB8_930_917145_919107
NC_014256	921308	922265	HPB8_933_921308_922265
NC_014256	923794	924013	HPB8_936_923794_924013
NC_014256	925570	926046	HPB8_939_925570_926046
NC_014256	927426	928711	HPB8_941_927426_928711
NC_014256	928795	929206	HPB8_942_928795_929206
NC_014256	929265	930324	HPB8_943_929265_930324
NC_014256	930474	930950	HPB8_944_930474_930950
NC_014256	931045	931122	HPB8_945_931045_931122
NC_014256	931175	932039	HPB8_945_931175_932039
NC_014256	934823	935752	HPB8_949_934823_935752
NC_014256	944184	946614	HPB8_959_944184_946614
NC_014256	947319	947978	HPB8_962_947319_947978
NC_014256	948157	949017	HPB8_964_948157_949017
NC_014256	949162	950335	HPB8_965_949162_950335
NC_014256	950601	950755	HPB8_966_950601_950755
NC_014256	950809	951586	HPB8_966_950809_951586
NC_014256	954047	954481	HPB8_969_954047_954481
NC_014256	957290	957497	HPB8_974_957290_957497
NC_014256	957563	958803	HPB8_974_957563_958803
NC_014256	959089	960030	HPB8_977_959089_960030
NC_014256	960219	960694	HPB8_978_960219_960694
NC_014256	960729	961778	HPB8_979_960729_961778
NC_014256	963947	965033	HPB8_982_963947_965033

NC_014256	965083	966240	HPB8_983_965083_966240
NC_014256	966349	968819	HPB8_984_966349_968819
NC_014256	968955	969583	HPB8_986_968955_969583
NC_014256	969694	970335	HPB8_987_969694_970335
NC_014256	970513	973064	HPB8_988_970513_973064
NC_014256	973147	973355	HPB8_989_973147_973355
NC_014256	973681	974959	HPB8_990_973681_974959
NC_014256	977460	977853	HPB8_993_977460_977853
NC_014256	982263	983745	HPB8_996_982263_983745
NC_014256	984109	984246	HPB8_998_984109_984246
NC_014256	984307	984604	HPB8_998_984307_984604
NC_014256	985339	987385	HPB8_999_985339_987385
NC_014256	987436	988930	HPB8_1000_987436_988930
NC_014256	988976	989460	HPB8_1001_988976_989460
NC_014256	989495	990052	HPB8_1002_989495_990052
NC_014256	990093	991428	HPB8_1003_990093_991428
NC_014256	991642	992324	HPB8_1004_991642_992324
NC_014256	992452	993187	HPB8_1005_992452_993187
NC_014256	995098	995646	HPB8_1010_995098_995646
NC_014256	995746	996574	HPB8_1011_995746_996574
NC_014256	996663	996826	HPB8_1012_996663_996826
NC_014256	996885	997654	HPB8_1012_996885_997654
NC_014256	997802	998544	HPB8_1013_997802_998544
NC_014256	998720	999274	HPB8_1014_998720_999274
NC_014256	999343	1000834	HPB8_1015_999343_1000834
NC_014256	1000875	1001721	HPB8_1015_1000875_1001721
NC_014256	1001941	1002229	HPB8_1016_1001941_1002229
NC_014256	1004152	1004355	HPB8_1020_1004152_1004355
NC_014256	1004415	1005166	HPB8_1020_1004415_1005166
NC_014256	1008738	1010299	HPB8_1026_1008738_1010299
NC_014256	1010424	1010579	HPB8_1027_1010424_1010579
NC_014256	1011328	1011646	HPB8_1029_1011328_1011646
NC_014256	1011722	1013446	HPB8_1030_1011722_1013446
NC_014256	1013483	1014739	HPB8_1031_1013483_1014739
NC_014256	1014783	1015055	HPB8_1032_1014783_1015055
NC_014256	1015200	1015433	HPB8_1033_1015200_1015433
NC_014256	1015508	1016431	HPB8_1034_1015508_1016431
NC_014256	1016726	1017532	HPB8_1035_1016726_1017532
NC_014256	1017772	1017998	HPB8_1036_1017772_1017998
NC_014256	1018417	1018968	HPB8_1037_1018417_1018968
NC_014256	1023486	1024370	HPB8_1043_1023486_1024370
NC_014256	1024446	1024944	HPB8_1044_1024446_1024944
NC_014256	1024981	1025751	HPB8_1044_1024981_1025751
NC_014256	1025985	1026269	HPB8_1045_1025985_1026269

NC_014256	1026978	1027571	HPB8_1047_1026978_1027571
NC_014256	1035351	1038467	HPB8_1056_1035351_1038467
NC_014256	1038623	1038892	HPB8_1057_1038623_1038892
NC_014256	1039691	1041262	HPB8_1058_1039691_1041262
NC_014256	1041653	1042405	HPB8_1059_1041653_1042405
NC_014256	1042465	1042970	HPB8_1060_1042465_1042970
NC_014256	1043162	1043378	HPB8_1060_1043162_1043378
NC_014256	1043547	1043873	HPB8_1061_1043547_1043873
NC_014256	1043914	1044232	HPB8_1061_1043914_1044232
NC_014256	1048677	1050031	HPB8_1067_1048677_1050031
NC_014256	1053206	1053330	HPB8_1072_1053206_1053330
NC_014256	1053462	1054563	HPB8_1072_1053462_1054563
NC_014256	1055698	1056116	HPB8_1075_1055698_1056116
NC_014256	1056183	1057265	HPB8_1076_1056183_1057265
NC_014256	1057357	1057664	HPB8_1077_1057357_1057664
NC_014256	1057733	1058086	HPB8_1078_1057733_1058086
NC_014256	1058147	1060296	HPB8_1079_1058147_1060296
NC_014256	1060526	1060763	HPB8_1081_1060526_1060763
NC_014256	1060810	1061254	HPB8_1082_1060810_1061254
NC_014256	1061625	1061894	HPB8_1084_1061625_1061894
NC_014256	1062012	1062214	HPB8_1085_1062012_1062214
NC_014256	1062327	1063174	HPB8_1086_1062327_1063174
NC_014256	1063350	1064836	HPB8_1087_1063350_1064836
NC_014256	1065162	1068011	HPB8_1088_1065162_1068011
NC_014256	1068141	1068852	HPB8_1090_1068141_1068852
NC_014256	1069928	1070042	HPB8_1091_1069928_1070042
NC_014256	1070232	1070969	HPB8_1092_1070232_1070969
NC_014256	1071149	1071226	HPB8_1093_1071149_1071226
NC_014256	1071483	1072834	HPB8_1094_1071483_1072834
NC_014256	1072975	1074539	HPB8_1095_1072975_1074539
NC_014256	1074825	1075697	HPB8_1096_1074825_1075697
NC_014256	1076052	1076347	HPB8_1097_1076052_1076347
NC_014256	1076847	1077800	HPB8_1098_1076847_1077800
NC_014256	1078015	1078457	HPB8_1099_1078015_1078457
NC_014256	1078751	1078982	HPB8_1100_1078751_1078982
NC_014256	1079087	1081274	HPB8_1102_1079087_1081274
NC_014256	1081328	1081458	HPB8_1103_1081328_1081458
NC_014256	1081521	1082193	HPB8_1103_1081521_1082193
NC_014256	1082351	1082726	HPB8_1103_1082351_1082726
NC_014256	1082861	1083569	HPB8_1103_1082861_1083569
NC_014256	1083984	1084067	HPB8_1104_1083984_1084067
NC_014256	1084484	1084644	HPB8_1104_1084484_1084644
NC_014256	1084993	1086374	HPB8_1106_1084993_1086374
NC_014256	1086503	1087278	HPB8_1107_1086503_1087278

NC_014256	1087339	1087965	HPB8_1108_1087339_1087965
NC_014256	1088063	1088732	HPB8_1109_1088063_1088732
NC_014256	1088971	1089513	HPB8_1111_1088971_1089513
NC_014256	1089912	1091100	HPB8_1112_1089912_1091100
NC_014256	1096072	1097909	HPB8_1118_1096072_1097909
NC_014256	1097969	1098449	HPB8_1119_1097969_1098449
NC_014256	1098489	1098659	HPB8_1120_1098489_1098659
NC_014256	1103059	1103577	HPB8_1123_1103059_1103577
NC_014256	1104490	1105689	HPB8_1126_1104490_1105689
NC_014256	1105747	1106494	HPB8_1126_1105747_1106494
NC_014256	1106588	1107520	HPB8_1127_1106588_1107520
NC_014256	1107606	1108922	HPB8_1128_1107606_1108922
NC_014256	1108955	1109639	HPB8_1129_1108955_1109639
NC_014256	1112050	1113401	HPB8_1133_1112050_1113401
NC_014256	1113527	1114916	HPB8_1134_1113527_1114916
NC_014256	1114974	1116182	HPB8_1135_1114974_1116182
NC_014256	1116307	1117097	HPB8_1136_1116307_1117097
NC_014256	1118300	1119550	HPB8_1141_1118300_1119550
NC_014256	1120966	1121652	HPB8_1144_1120966_1121652
NC_014256	1121710	1122132	HPB8_1144_1121710_1122132
NC_014256	1126873	1127659	HPB8_1152_1126873_1127659
NC_014256	1130673	1131151	HPB8_1156_1130673_1131151
NC_014256	1131626	1132254	HPB8_1157_1131626_1132254
NC_014256	1132295	1132540	HPB8_1157_1132295_1132540
NC_014256	1132617	1132865	HPB8_1158_1132617_1132865
NC_014256	1133008	1133691	HPB8_1159_1133008_1133691
NC_014256	1133732	1135173	HPB8_1159_1133732_1135173
NC_014256	1135240	1136274	HPB8_1160_1135240_1136274
NC_014256	1136642	1137313	HPB8_1161_1136642_1137313
NC_014256	1137667	1138779	HPB8_1162_1137667_1138779
NC_014256	1138975	1140957	HPB8_1163_1138975_1140957
NC_014256	1144928	1145807	HPB8_1169_1144928_1145807
NC_014256	1148401	1149049	HPB8_1173_1148401_1149049
NC_014256	1149196	1149340	HPB8_1174_1149196_1149340
NC_014256	1149374	1150077	HPB8_1174_1149374_1150077
NC_014256	1151830	1152631	HPB8_1177_1151830_1152631
NC_014256	1152671	1153631	HPB8_1178_1152671_1153631
NC_014256	1153721	1154138	HPB8_1179_1153721_1154138
NC_014256	1154196	1154432	HPB8_1180_1154196_1154432
NC_014256	1157282	1157862	HPB8_1185_1157282_1157862
NC_014256	1161710	1162341	HPB8_1191_1161710_1162341
NC_014256	1162420	1164846	HPB8_1192_1162420_1164846
NC_014256	1165608	1165757	HPB8_1194_1165608_1165757
NC_014256	1165822	1166834	HPB8_1194_1165822_1166834

NC_014256	1166999	1167487	HPB8_1195_1166999_1167487
NC_014256	1167540	1168513	HPB8_1196_1167540_1168513
NC_014256	1169594	1170213	HPB8_1198_1169594_1170213
NC_014256	1170566	1171094	HPB8_1199_1170566_1171094
NC_014256	1173502	1174892	HPB8_1202_1173502_1174892
NC_014256	1175195	1176118	HPB8_1203_1175195_1176118
NC_014256	1176185	1176696	HPB8_1204_1176185_1176696
NC_014256	1176776	1177312	HPB8_1205_1176776_1177312
NC_014256	1177383	1178194	HPB8_1206_1177383_1178194
NC_014256	1178262	1180498	HPB8_1207_1178262_1180498
NC_014256	1180627	1182876	HPB8_1208_1180627_1182876
NC_014256	1182940	1185069	HPB8_1209_1182940_1185069
NC_014256	1185329	1186609	HPB8_1210_1185329_1186609
NC_014256	1186664	1187462	HPB8_1211_1186664_1187462
NC_014256	1188152	1188294	HPB8_1214_1188152_1188294
NC_014256	1194693	1196322	HPB8_1220_1194693_1196322
NC_014256	1196381	1196620	HPB8_1221_1196381_1196620
NC_014256	1196683	1197035	HPB8_1221_1196683_1197035
NC_014256	1197347	1198242	HPB8_1223_1197347_1198242
NC_014256	1198295	1198902	HPB8_1223_1198295_1198902
NC_014256	1202575	1202926	HPB8_1229_1202575_1202926
NC_014256	1203175	1203579	HPB8_1230_1203175_1203579
NC_014256	1203707	1204334	HPB8_1231_1203707_1204334
NC_014256	1206684	1207364	HPB8_1235_1206684_1207364
NC_014256	1207547	1208368	HPB8_1236_1207547_1208368
NC_014256	1208873	1209710	HPB8_1237_1208873_1209710
NC_014256	1212705	1213881	HPB8_1243_1212705_1213881
NC_014256	1213957	1214560	HPB8_1244_1213957_1214560
NC_014256	1214664	1216469	HPB8_1246_1214664_1216469
NC_014256	1216575	1217330	HPB8_1247_1216575_1217330
NC_014256	1217703	1218923	HPB8_1249_1217703_1218923
NC_014256	1225027	1226020	HPB8_1258_1225027_1226020
NC_014256	1230737	1232321	HPB8_1264_1230737_1232321
NC_014256	1233318	1235781	HPB8_1267_1233318_1235781
NC_014256	1235887	1236906	HPB8_1268_1235887_1236906
NC_014256	1237104	1238181	HPB8_1269_1237104_1238181
NC_014256	1238224	1238774	HPB8_1269_1238224_1238774
NC_014256	1238812	1239573	HPB8_1270_1238812_1239573
NC_014256	1239688	1241177	HPB8_1272_1239688_1241177
NC_014256	1241225	1249970	HPB8_1273_1241225_1249970
NC_014256	1250042	1250515	HPB8_1274_1250042_1250515
NC_014256	1258032	1259117	HPB8_1281_1258032_1259117
NC_014256	1259277	1260071	HPB8_1282_1259277_1260071
NC_014256	1260240	1261048	HPB8_1283_1260240_1261048

NC_014256	1261133	1262622	HPB8_1284_1261133_1262622
NC_014256	1262705	1262947	HPB8_1285_1262705_1262947
NC_014256	1264890	1265286	HPB8_1288_1264890_1265286
NC_014256	1265319	1265750	HPB8_1289_1265319_1265750
NC_014256	1267361	1267924	HPB8_1292_1267361_1267924
NC_014256	1268027	1269307	HPB8_1293_1268027_1269307
NC_014256	1269352	1269582	HPB8_1294_1269352_1269582
NC_014256	1269653	1270829	HPB8_1295_1269653_1270829
NC_014256	1272789	1272849	HPB8_1298_1272789_1272849
NC_014256	1272906	1272978	HPB8_1298_1272906_1272978
NC_014256	1273020	1275294	HPB8_1298_1273020_1275294
NC_014256	1275437	1276190	HPB8_1299_1275437_1276190
NC_014256	1277086	1277609	HPB8_1302_1277086_1277609
NC_014256	1277861	1278810	HPB8_1303_1277861_1278810
NC_014256	1278862	1280124	HPB8_1304_1278862_1280124
NC_014256	1286833	1287822	HPB8_1312_1286833_1287822
NC_014256	1287878	1289409	HPB8_1313_1287878_1289409
NC_014256	1301576	1301905	HPB8_1327_1301576_1301905
NC_014256	1302035	1303093	HPB8_1328_1302035_1303093
NC_014256	1303828	1304358	HPB8_1331_1303828_1304358
NC_014256	1307120	1307806	HPB8_1335_1307120_1307806
NC_014256	1308045	1309331	HPB8_1336_1308045_1309331
NC_014256	1309739	1310894	HPB8_1337_1309739_1310894
NC_014256	1311295	1312150	HPB8_1338_1311295_1312150
NC_014256	1312206	1312634	HPB8_1338_1312206_1312634
NC_014256	1313350	1314150	HPB8_1341_1313350_1314150
NC_014256	1314493	1315498	HPB8_1342_1314493_1315498
NC_014256	1315678	1317048	HPB8_1343_1315678_1317048
NC_014256	1317138	1317369	HPB8_1344_1317138_1317369
NC_014256	1317535	1318515	HPB8_1345_1317535_1318515
NC_014256	1318546	1319701	HPB8_1346_1318546_1319701
NC_014256	1319873	1320337	HPB8_1347_1319873_1320337
NC_014256	1320483	1321032	HPB8_1348_1320483_1321032
NC_014256	1328084	1328836	HPB8_1354_1328084_1328836
NC_014256	1329039	1330840	HPB8_1356_1329039_1330840
NC_014256	1331059	1332361	HPB8_1357_1331059_1332361
NC_014256	1332520	1333591	HPB8_1359_1332520_1333591
NC_014256	1333713	1334799	HPB8_1360_1333713_1334799
NC_014256	1334899	1335290	HPB8_1361_1334899_1335290
NC_014256	1336072	1336670	HPB8_1361_1336072_1336670
NC_014256	1336915	1338299	HPB8_1361_1336915_1338299
NC_014256	1341129	1341262	HPB8_1367_1341129_1341262
NC_014256	1341295	1341617	HPB8_1368_1341295_1341617
NC_014256	1341700	1342041	HPB8_1369_1341700_1342041

NC_014256	1342147	1343286	HPB8_1370_1342147_1343286
NC_014256	1352171	1353697	HPB8_1379_1352171_1353697
NC_014256	1353778	1353888	HPB8_1380_1353778_1353888
NC_014256	1359164	1359472	HPB8_1386_1359164_1359472
NC_014256	1359679	1360424	HPB8_1386_1359679_1360424
NC_014256	1365549	1366097	HPB8_1392_1365549_1366097
NC_014256	1366271	1367008	HPB8_1393_1366271_1367008
NC_014256	1367048	1368214	HPB8_1394_1367048_1368214
NC_014256	1368284	1369357	HPB8_1395_1368284_1369357
NC_014256	1372465	1373142	HPB8_1400_1372465_1373142
NC_014256	1373207	1374427	HPB8_1401_1373207_1374427
NC_014256	1376804	1377559	HPB8_1405_1376804_1377559
NC_014256	1377818	1378675	HPB8_1406_1377818_1378675
NC_014256	1381061	1381291	HPB8_1410_1381061_1381291
NC_014256	1383719	1384051	HPB8_1413_1383719_1384051
NC_014256	1384113	1384525	HPB8_1413_1384113_1384525
NC_014256	1384659	1385346	HPB8_1414_1384659_1385346
NC_014256	1386728	1387809	HPB8_1418_1386728_1387809
NC_014256	1392846	1393498	HPB8_1424_1392846_1393498
NC_014256	1393571	1394501	HPB8_1424_1393571_1394501
NC_014256	1394551	1395611	HPB8_1425_1394551_1395611
NC_014256	1395656	1396200	HPB8_1425_1395656_1396200
NC_014256	1396346	1397146	HPB8_1426_1396346_1397146
NC_014256	1399262	1399720	HPB8_1429_1399262_1399720
NC_014256	1400381	1401725	HPB8_1431_1400381_1401725
NC_014256	1401907	1403121	HPB8_1433_1401907_1403121
NC_014256	1403158	1404494	HPB8_1434_1403158_1404494
NC_014256	1404678	1404771	HPB8_1435_1404678_1404771
NC_014256	1405144	1406004	HPB8_1436_1405144_1406004
NC_014256	1406447	1406847	HPB8_1438_1406447_1406847
NC_014256	1407213	1408060	HPB8_1440_1407213_1408060
NC_014256	1408265	1408536	HPB8_1441_1408265_1408536
NC_014256	1409559	1411316	HPB8_1444_1409559_1411316
NC_014256	1411372	1411503	HPB8_1445_1411372_1411503
NC_014256	1411683	1414115	HPB8_1446_1411683_1414115
NC_014256	1415522	1416455	HPB8_1448_1415522_1416455
NC_014256	1416582	1418658	HPB8_1449_1416582_1418658
NC_014256	1418824	1420368	HPB8_1450_1418824_1420368
NC_014256	1420426	1420782	HPB8_1451_1420426_1420782
NC_014256	1420950	1422273	HPB8_1451_1420950_1422273
NC_014256	1422347	1422614	HPB8_1452_1422347_1422614
NC_014256	1422679	1423103	HPB8_1453_1422679_1423103
NC_014256	1423151	1423268	HPB8_1453_1423151_1423268
NC_014256	1423490	1423927	HPB8_1454_1423490_1423927

NC_014256	1424921	1426771	HPB8_1456_1424921_1426771
NC_014256	1427090	1427585	HPB8_1457_1427090_1427585
NC_014256	1429809	1430264	HPB8_1460_1429809_1430264
NC_014256	1430425	1432119	HPB8_1461_1430425_1432119
NC_014256	1432373	1434065	HPB8_1462_1432373_1434065
NC_014256	1434102	1434812	HPB8_1463_1434102_1434812
NC_014256	1434956	1435563	HPB8_1464_1434956_1435563
NC_014256	1435872	1436095	HPB8_1466_1435872_1436095
NC_014256	1436181	1436880	HPB8_1466_1436181_1436880
NC_014256	1437189	1438968	HPB8_1467_1437189_1438968
NC_014256	1439172	1440629	HPB8_1468_1439172_1440629
NC_014256	1440737	1441450	HPB8_1469_1440737_1441450
NC_014256	1441664	1442608	HPB8_1470_1441664_1442608
NC_014256	1442681	1443211	HPB8_1471_1442681_1443211
NC_014256	1443334	1444205	HPB8_1472_1443334_1444205
NC_014256	1451844	1453391	HPB8_1479_1451844_1453391
NC_014256	1454670	1456410	HPB8_1483_1454670_1456410
NC_014256	1456463	1456660	HPB8_1483_1456463_1456660
NC_014256	1456974	1458746	HPB8_1484_1456974_1458746
NC_014256	1459022	1459782	HPB8_1485_1459022_1459782
NC_014256	1460226	1461234	HPB8_1487_1460226_1461234
NC_014256	1461358	1461627	HPB8_1488_1461358_1461627
NC_014256	1469466	1469786	HPB8_1499_1469466_1469786
NC_014256	1469846	1472280	HPB8_1501_1469846_1472280
NC_014256	1479664	1479873	HPB8_1511_1479664_1479873
NC_014256	1480253	1483743	HPB8_1512_1480253_1483743
NC_014256	1483779	1484083	HPB8_1513_1483779_1484083
NC_014256	1484263	1485250	HPB8_1513_1484263_1485250
NC_014256	1485309	1485432	HPB8_1514_1485309_1485432
NC_014256	1485774	1486457	HPB8_1516_1485774_1486457
NC_014256	1486597	1486775	HPB8_1517_1486597_1486775
NC_014256	1491489	1491623	HPB8_1522_1491489_1491623
NC_014256	1491980	1492073	HPB8_1523_1491980_1492073
NC_014256	1492327	1492581	HPB8_1524_1492327_1492581
NC_014256	1492944	1494526	HPB8_1525_1492944_1494526
NC_014256	1499064	1499888	HPB8_1531_1499064_1499888
NC_014256	1500032	1500230	HPB8_1532_1500032_1500230
NC_014256	1500363	1501030	HPB8_1532_1500363_1501030
NC_014256	1504805	1505208	HPB8_1535_1504805_1505208
NC_014256	1505275	1505916	HPB8_1536_1505275_1505916
NC_014256	1509706	1510553	HPB8_1539_1509706_1510553
NC_014256	1510616	1511196	HPB8_1539_1510616_1511196
NC_014256	1511431	1512084	HPB8_1540_1511431_1512084
NC_014256	1512179	1513220	HPB8_1541_1512179_1513220

NC_014256	1516091	1516462	HPB8_1545_1516091_1516462
NC_014256	1516712	1516824	HPB8_1546_1516712_1516824
NC_014256	1516878	1517540	HPB8_1547_1516878_1517540
NC_014256	1517582	1520065	HPB8_1548_1517582_1520065
NC_014256	1520126	1520904	HPB8_1549_1520126_1520904
NC_014256	1521135	1521966	HPB8_1551_1521135_1521966
NC_014256	1522089	1522840	HPB8_1552_1522089_1522840
NC_014256	1523039	1523900	HPB8_1552_1523039_1523900
NC_014256	1524094	1524749	HPB8_1553_1524094_1524749
NC_014256	1524823	1526398	HPB8_1553_1524823_1526398
NC_014256	1526808	1527796	HPB8_1555_1526808_1527796
NC_014256	1527837	1528492	HPB8_1556_1527837_1528492
NC_014256	1533342	1533555	HPB8_1562_1533342_1533555
NC_014256	1533701	1535103	HPB8_1563_1533701_1535103
NC_014256	1535241	1535812	HPB8_1564_1535241_1535812
NC_014256	1536130	1536315	HPB8_1566_1536130_1536315
NC_014256	1536396	1538235	HPB8_1567_1536396_1538235
NC_014256	1539415	1539829	HPB8_1570_1539415_1539829
NC_014256	1539947	1540077	HPB8_1571_1539947_1540077
NC_014256	1542768	1542946	HPB8_1574_1542768_1542946
NC_014256	1547604	1547698	HPB8_1578_1547604_1547698
NC_014256	1550131	1551551	HPB8_1582_1550131_1551551
NC_014256	1554587	1555642	HPB8_1587_1554587_1555642
NC_014256	1557418	1559297	HPB8_1591_1557418_1559297
NC_014256	1559371	1559489	HPB8_1591_1559371_1559489
NC_014256	1563304	1564557	HPB8_1597_1563304_1564557
NC_014256	1564762	1566042	HPB8_1598_1564762_1566042
NC_014256	1566602	1568518	HPB8_1600_1566602_1568518
NC_014256	1568937	1569022	HPB8_1601_1568937_1569022
NC_014256	1569305	1570865	HPB8_1603_1569305_1570865
NC_014256	1573761	1575201	HPB8_1607_1573761_1575201
NC_014256	1575430	1577371	HPB8_1608_1575430_1577371
NC_014256	1577513	1577723	HPB8_1608_1577513_1577723
NC_014256	1578468	1579295	HPB8_1611_1578468_1579295
NC_014256	1582399	1584420	HPB8_1615_1582399_1584420
NC_014256	1584765	1586291	HPB8_1616_1584765_1586291
NC_014256	1586394	1586651	HPB8_1616_1586394_1586651
NC_014256	1586778	1586864	HPB8_1617_1586778_1586864
NC_014256	1591958	1592274	HPB8_1625_1591958_1592274
NC_014256	1592733	1593476	HPB8_1626_1592733_1593476
NC_014256	1593579	1593691	HPB8_1627_1593579_1593691
NC_014256	1594223	1594618	HPB8_1629_1594223_1594618
NC_014256	1594713	1594991	HPB8_1630_1594713_1594991
NC_014256	1595090	1595878	HPB8_1631_1595090_1595878

NC_014256	1596725	1597737	HPB8_1633_1596725_1597737
NC_014256	1599661	1600668	HPB8_1638_1599661_1600668
NC_014256	1601274	1601676	HPB8_1639_1601274_1601676
NC_014256	1603301	1603515	HPB8_1641_1603301_1603515
NC_014256	1604522	1605271	HPB8_1644_1604522_1605271
NC_014256	1605381	1606470	HPB8_1645_1605381_1606470
NC_014256	1609811	1611876	HPB8_1652_1609811_1611876
NC_014256	1612375	1613190	HPB8_1655_1612375_1613190
NC_014256	1613336	1613932	HPB8_1656_1613336_1613932
NC_014256	1614173	1614508	HPB8_1657_1614173_1614508
NC_014256	1614619	1614865	HPB8_1657_1614619_1614865
NC_014256	1614996	1615132	HPB8_1657_1614996_1615132
NC_014256	1615372	1616024	HPB8_1658_1615372_1616024
NC_014256	1616063	1616216	HPB8_1658_1616063_1616216
NC_014256	1616402	1617568	HPB8_1659_1616402_1617568
NC_014256	1618823	1619149	HPB8_1663_1618823_1619149
NC_014256	1619307	1621134	HPB8_1664_1619307_1621134
NC_014256	1621450	1621751	HPB8_1665_1621450_1621751
NC_014256	1621918	1622715	HPB8_1665_1621918_1622715
NC_014256	1637672	1638103	HPB8_1678_1637672_1638103
NC_014256	1638181	1638488	HPB8_1679_1638181_1638488
NC_014256	1643029	1643823	HPB8_1682_1643029_1643823
NC_014256	1643879	1644890	HPB8_1683_1643879_1644890
NC_014256	1645812	1646823	HPB8_1686_1645812_1646823
NC_014256	1646954	1647721	HPB8_1688_1646954_1647721
NC_014256	1648341	1649277	HPB8_1689_1648341_1649277
NC_014256	1649353	1649877	HPB8_1690_1649353_1649877
NC_014256	1650001	1650334	HPB8_1690_1650001_1650334
NC_014256	1651543	1652027	HPB8_1692_1651543_1652027
NC_014256	1652172	1653656	HPB8_1693_1652172_1653656
NC_014256	1653699	1654534	HPB8_1694_1653699_1654534
NC_014256	1654720	1655428	HPB8_1696_1654720_1655428
NC_014256	1656572	1657225	HPB8_1699_1656572_1657225
NC_014256	1657360	1658488	HPB8_1700_1657360_1658488
NC_014256	1658692	1659647	HPB8_1701_1658692_1659647
NC_014256	1660005	1662516	HPB8_1702_1660005_1662516
NC_014256	1662620	1663327	HPB8_1703_1662620_1663327
NC_014256	1663620	1666264	HPB8_1704_1663620_1666264
NC_014256	1666391	1668838	HPB8_1705_1666391_1668838
NC_014256	1669787	1669959	HPB8_1706_1669787_1669959

10.3. SelectionA - Gastric Adenocarcinoma Unique Domains

CHR	START	END	GENE_ID	FOLDCHANGE	ANNOTATION	FLAG
NC_000915	432871	432986	HP0418	9,40	hypothetical protein	Sel A unique
NC_000915	749410	749497	HP0696	8,29	N-methylhydantoinase	Sel A unique
NC_000915	880251	880485	HP0828	5,06	ATP synthase F0F1 subunit A	Sel A unique
NC_000915	555864	556076	HP0527	4,79	cag pathogenicity island protein cag7	Sel A unique
NC_000915	1051051	1051229	HP0988	4,51	IS605 transposase TnpA	Sel A unique
NC_000915	356709	356816	HP0349	4,41	CTP synthetase	Sel A unique
NC_000915	1194612	1194688	HP1132	3,45	ATP synthase F0F1 subunit beta	Sel A unique
NC_000915	70004	70177	HP0066	3,10	ATP-binding protein	Sel A unique
NC_000915	271337	271508	HP0262	2,93	hypothetical protein	Sel A unique
NC_000915	1061927	1062105	HP0998	2,89	IS605 transposase TnpA	Sel A unique
NC_000915	312869	313125	HP0295	2,55	flagellar hook-associated protein FlgL	Sel A unique
NC_000915	1563976	1564163	HP1490	2,50	hypothetical protein	Sel A unique
NC_000915	979186	979349	HP0921	2,49	glyceraldehyde-3-phosphate dehydrogenase	Sel A unique
NC_000915	320256	320492	HP0302	2,45	dipeptide ABC transporter ATP-binding protein DppF	Sel A unique
NC_000915	951212	951394	HP0898	2,35	hydrogenase expression/formation protein HypD	Sel A unique
NC_000915	846725	846835	HP0790	2,28	anti-codon nuclease masking agent PrrB	Sel A unique
NC_000915	1063457	1063650	HP1001	2,25	hypothetical protein	Sel A unique
NC_000915	588313	588527	HP0553	2,21	hypothetical protein	Sel A unique
NC_000915	106698	106847	HP0100	2,17	hypothetical protein	Sel A unique
NC_000915	5875	6023	HP0009	2,07	hypothetical protein	Sel A unique
NC_000915	1243612	1243805	HP1177	2,06	hypothetical protein	Sel A unique
NC_000915	999864	1000073	HP0939	1,91	amino acid ABC transporter permease	Sel A unique
NC_000915	176025	176197	HP0169	1,89	collagenase PrtC	Sel A unique
NC_000915	1047284	1047501	HP0983	1,88	hypothetical protein	Sel A unique
NC_000915	1372726	1372975	HP1301	1,78	50S ribosomal protein L15	Sel A unique

10.4. SelectionB - Autoimmune Gastritis Unique Domains

CHR	START	END	GENE_ID	FOLDCHANGE	ANNOTATION	FLAG
NC_000915	1004123	1004218	HP0943	8,98	D-amino acid dehydrogenase DadA	Sel B unique
NC_000915	864151	864245	HP0809	8,55	flagellar basal body protein FliL	Sel B unique
NC_000915	1491923	1492043	HP1422	8,18	isoleucyl-tRNA synthetase	Sel B unique
NC_000915	512074	512214	HP0487	5,27	hypothetical protein	Sel B unique

NC_000915	1073160	1073202	HP1010	4,71	polyphosphate kinase	Sel B unique
NC_000915	945751	945866	HP0892	4,57	hypothetical protein	Sel B unique
NC_000915	373124	373195	HP0363	3,99	protein-L-isoaspartate O-methyltransferase	Sel B unique
NC_000915	918695	918813	HP0867	3,76	ipid-A-disaccharide synthase	Sel B unique
NC_000915	1574664	1574835	HP1502	3,73	hypothetical protein	Sel B unique
NC_000915	1158327	1158505	HP1096	3,60	IS605 transposase TnpA	Sel B unique
NC_000915	1028615	1028881	HP0969	3,47	cation efflux system protein CzCA	Sel B unique
NC_000915	140452	141127	HP0130	3,34	hypothetical protein	Sel B unique
NC_000915	555284	556076	HP0527	3,12	cag pathogenicity island protein cag7	Sel B unique
NC_000915	475825	476089	HP0457	3,06	hypothetical protein	Sel B unique
NC_000915	749410	749655	HP0696	3,05	N-methylhydantoinase	Sel B unique
NC_000915	1564206	1564417	HP1490	3,01	hypothetical protein	Sel B unique
NC_000915	1558353	1558552	HP1486	2,96	hypothetical protein	Sel B unique
NC_000915	576113	576186	HP0544	2,87	cag pathogenicity island protein cag23	Sel B unique
NC_000915	106624	106910	HP0100	2,84	hypothetical protein	Sel B unique
NC_000915	721767	722123	HP0671	2,80	hypothetical protein	Sel B unique
NC_000915	406342	406513	HP0396	2,79	hypothetical protein	Sel B unique
NC_000915	811715	811947	HP0758	2,79	hypothetical protein	Sel B unique
NC_000915	1627113	1627477	HP1547	2,75	leucyl-tRNA synthetase	Sel B unique
NC_000915	882360	883007	HP0830	2,74	aspartyl/glutamyl-tRNA amidotransferase subunit A	Sel B unique
NC_000915	627603	628013	HP0593	2,69	adenine-specific DNA methyltransferase	Sel B unique
NC_000915	1399532	1399604	HP1340	2,67	biopolymer transport protein ExbD	Sel B unique
NC_000915	852233	852399	HP0795	2,63	trigger factor	Sel B unique
NC_000915	278658	278779	HP0269	2,60	(dimethylallyl)adenosine tRNA methylthiotransferase	Sel B unique
NC_000915	1338786	1338916	HP1266	2,58	NADH dehydrogenase subunit G	Sel B unique
NC_000915	146944	147190	HP0136	2,56	bacterioferritin comigratory protein Bcp	Sel B unique
NC_000915	143701	144124	HP0133	2,55	serine transporter SdaC	Sel B unique
NC_000915	477705	478199	HP0459	2,52	protein VirB4	Sel B unique
NC_000915	328396	328584	HP0312	2,49	ATP-binding protein	Sel B unique
NC_000915	1374888	1375034	HP1306	2,34	30S ribosomal protein S14	Sel B unique
NC_000915	902944	903171	HP0851	2,32	hypothetical protein	Sel B unique
NC_000915	519533	519950	HP0493	2,32	phospho-N-acetylmuramoyl-pentapeptide-transferase	Sel B unique
NC_000915	552880	553202	HP0525	2,31	virB11-like protein	Sel B unique
NC_000915	10215	11080	HP0012	2,28	DNA primase	Sel B unique
NC_000915	1583199	1583444	HP1509	2,27	glycerol-3-phosphate acyltransferase PlsY	Sel B unique
NC_000915	1382149	1382490	HP1321	2,27	hypothetical protein	Sel B unique

NC_000915	951075	951395	HP0898	2,27	hydrogenase expression/formation protein HypD	Sel B unique
NC_000915	907649	907838	HP0855	2,26	alginate O-acetylation protein AlgI	Sel B unique
NC_000915	60676	60803	HP0056	2,23	delta-1-pyrroline-5-carboxylate dehydrogenase	Sel B unique
NC_000915	1026447	1026625	HP0967	2,20	virulence associated protein D vapD	Sel B unique
NC_000915	497151	497374	HP0475	2,16	molybdenum ABC transporter ATP-binding protein ModD	Sel B unique
NC_000915	397190	397438	HP0387	2,12	primosome assembly protein PriA	Sel B unique
NC_000915	320109	320326	HP0302	2,09	dipeptide ABC transporter ATP-binding protein DppF	Sel B unique
NC_000915	1284797	1284992	HP1207	2,08	hypothetical protein	Sel B unique
NC_000915	1334949	1335478	HP1262	2,06	NADH dehydrogenase subunit C	Sel B unique
NC_000915	1545372	1545572	HP1472	2,03	type IIS restriction enzyme M protein	Sel B unique
NC_000915	1107752	1108034	HP1045	1,99	acetyl-CoA synthetase	Sel B unique
NC_000915	798735	798907	HP0743	1,99	rod shape-determining protein MreB	Sel B unique
NC_000915	1570699	1570993	HP1498	1,98	hypothetical protein	Sel B unique
NC_000915	1354475	1354543	HP1279	1,98	bifunctional indole-3-glycerol phosphate synthase isomerase	Sel B unique
NC_000915	835626	835696	HP0781	1,95	hypothetical protein	Sel B unique
NC_000915	972820	972958	HP0916	1,95	iron-regulated outer membrane protein FrpB	Sel B unique
NC_000915	1117318	1117518	HP1055	1,94	hypothetical protein	Sel B unique
NC_000915	511175	511376	HP0487	1,92	hypothetical protein	Sel B unique
NC_000915	1332564	1332994	HP1258	1,91	hypothetical protein	Sel B unique
NC_000915	880169	880334	HP0828	1,90	ATP synthase F0F1 subunit A	Sel B unique
NC_000915	536705	536926	HP0510	1,89	dihydrodipicolinate reductase	Sel B unique
NC_000915	298023	298185	HP0289	1,87	toxin-like outer membrane protein	Sel B unique
NC_000915	837963	838354	HP0783	1,87	hypothetical protein	Sel B unique
NC_000915	1656737	1657203	HP1577	1,86	ABC transporter permease	Sel B unique
NC_000915	709780	709958	HP0661	1,85	ribonuclease H	Sel B unique
NC_000915	170854	171136	HP0162	1,83	hypothetical protein	Sel B unique
NC_000915	329641	329835	HP0313	1,82	nitrite extrusion protein NarK	Sel B unique
NC_000915	1536901	1537293	HP1466	1,82	hypothetical protein	Sel B unique
NC_000915	803140	803421	HP0747	1,81	tRNA (guanine-N(7)-)-methyltransferase	Sel B unique
NC_000915	1176168	1176502	HP1114	1,77	excinuclease ABC subunit B	Sel B unique
NC_000915	16862	17194	HP0018	1,76	hypothetical protein	Sel B unique
NC_000915	1557886	1558193	HP1485	1,72	hypothetical protein	Sel B unique
NC_000915	1374255	1374313	HP1304	1,70	50S ribosomal protein L6	Sel B unique
NC_000915	21826	21995	HP0022	1,69	lipid A phosphoethanolamine transferase	Sel B unique
NC_000915	237376	237722	HP0228	1,68	hypothetical protein	Sel B unique
NC_000915	32679	32942	HP0033	1,64	ATP-dependent C1p protease	Sel B unique

					ClpA	
NC_000915	855368	855623	HP0799	1,62	molybdenum cofactor biosynthesis protein MogA	Sel B unique
NC_000915	78559	78755	HP0074	1,61	lipoprotein signal peptidase	Sel B unique
NC_000915	61634	61828	HP0057	1,58	hypothetical protein	Sel B unique
NC_000915	854493	854659	HP0797	1,57	flagellar sheath adhesin hpaA	Sel B unique
NC_000915	524327	524527	HP0498	1,56	sodium- and chloride-dependent transporter	Sel B unique
NC_000915	1544120	1544616	HP1471	1,53	type IIS restriction enzyme R protein	Sel B unique
NC_000915	84756	85321	HP0080	1,52	hypothetical protein	Sel B unique
NC_000915	983327	983850	HP0922	1,52	toxin-like outer membrane protein	Sel B unique
NC_000915	1172395	1172634	HP1111	1,51	ferrodoxin oxidoreductase subunit beta	Sel B unique

10.5. SelectionC - MALT lymphoma Unique Domains

CHR	START	END	GENE_ID	FOLDCHANGE	ANNOTATION	FLAG
NC_000915	556305	557066	HP0527	7,71	cag pathogenicity island protein cag7	Sel C unique
NC_000915	777672	777742	HP0724	5,24	anaerobic C4-dicarboxylate transporter	Sel C unique
NC_000915	1585441	1585851	HP1512	5,04	iron-regulated outer membrane protein FrpB	Sel C unique
NC_000915	204394	204591	HP0197	4,12	S-adenosylmethionine synthetase	Sel C unique
NC_000915	17965	18272	HP0018	3,45	hypothetical protein	Sel C unique
NC_000915	712341	712488	HP0665	3,18	coproporphyrinogen III oxidase	Sel C unique
NC_000915	1317869	1317975	HP1243	3,10	hypothetical protein	Sel C unique
NC_000915	274274	274470	HP0264	3,01	ATP-dependent protease binding subunit ClpB	Sel C unique
NC_000915	477730	477960	HP0459	2,91	protein VirB4	Sel C unique
NC_000915	203274	203353	HP0196	2,82	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	Sel C unique
NC_000915	1626372	1626445	HP1547	2,72	leucyl-tRNA synthetase	Sel C unique
NC_000915	1461955	1462126	HP1400	2,62	iron(III) dicitrate transport protein FecA	Sel C unique
NC_000915	983602	984010	HP0922	2,55	toxin-like outer membrane protein	Sel C unique
NC_000915	422464	422818	HP0409	2,35	GMP synthase	Sel C unique
NC_000915	1024311	1024404	HP0965	2,31	hypothetical protein	Sel C unique
NC_000915	1249755	1250123	HP1181	2,21	multidrug-efflux transporter	Sel C unique
NC_000915	221588	221643	HP0213	2,12	tRNA uridine 5-carboxymethylaminomethyl modification protein GidA	Sel C unique
NC_000915	484059	484302	HP0464	2,11	type I restriction enzyme R protein HsdR	Sel C unique
NC_000915	1332625	1332798	HP1258	2,04	hypothetical protein	Sel C unique
NC_000915	827888	827994	HP0775	1,98	penta-phosphate guanosine-3'-pyrophosphohydrolase SpoT	Sel C unique
NC_000915	1241938	1242043	HP1175	1,96	hypothetical protein	Sel C unique
NC_000915	411221	411357	HP0400	1,94	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	Sel C unique

NC_000915	1019626	1019879	HP0961	1,90	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase	Sel C unique
NC_000915	1206023	1206430	HP1143	1,85	hypothetical protein	Sel C unique
NC_000915	933394	933665	HP0884	1,79	hypothetical protein	Sel C unique
NC_000915	1428027	1428285	HP1366	1,77	type IIS restriction enzyme R protein	Sel C unique
NC_000915	795181	795399	HP0740	1,74	UDP-MurNac-pentapeptide presynthetase MurF	Sel C unique
NC_000915	1584941	1585092	HP1512	1,60	iron-regulated outer membrane protein FrpB	Sel C unique

10.6. Common Domains between Selection A-B-C

CHR	START	END	GENE_ID	FOLDCHANGE	ANNOTATION	FLAG
NC_000915	607582	607662	HP0577	5,33	bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase	Common between Sel A-B-C
NC_000915	1592925	1592985	HP1517	3,03	type IIS restriction enzyme R and M protein	Common between Sel A-B-C
NC_000915	241192	241296	HP0231	2,49	hypothetical protein	Common between Sel A-B-C
NC_000915	414784	414866	HP0402	2,21	phenylalanyl-tRNA synthetase subunit beta	Common between Sel A-B-C
NC_000915	1396381	1396472	HP1335	2,09	tRNA-specific 2-thiouridylase MnmA	Common between Sel A-B-C
NC_000915	577358	577581	HP0544	2,03	cag pathogenicity island protein cag23	Common between Sel A-B-C
NC_000915	1018926	1019099	HP0960	1,65	glycyl-tRNA synthetase subunit alpha	Common between Sel A-B-C
NC_000915	1642827	1643112	HP1561	1,64	iron(III) ABC transportersubstrate-binding protein CeuE	Common between Sel A-B-C

10.7. Common Domains between Selection A-C

CHR	START	END	GENE_ID	FOLDCHANGE	ANNOTATION	FLAG
NC_000915	437431	437598	HP0422	4,18	arginine decarboxylase	Common between Sel A-C
NC_000915	195103	195230	HP0189	2,65	hypothetical protein	Common between Sel A-C
NC_000915	1345616	1345804	HP1272	2,16	NADH dehydrogenase subunit M	Common between Sel A-C

10.8. Common Domains between Selection A-B

CHR	START	END	GENE_ID	FOLDCHANGE	ANNOTATION	FLAG
NC_000915	1031857	1031904	HP0971	9,41	hypothetical protein	Common between Sel A-B
NC_000915	1051051	1051229	HP0988	5,57	IS605 transposase TnpA	Common between Sel A-B

NC_000915	1555522	1555630	HP1481	5,40	hypothetical protein	Common between Sel A-B
NC_000915	168551	168654	HP0159	4,95	lipopolysaccharide 1,2-glucosyltransferase RfaJ	Common between Sel A-B
NC_000915	1614604	1614782	HP1535	4,63	IS605 transposase TnpA	Common between Sel A-B
NC_000915	1552411	1552496	HP1479	4,55	hypothetical protein	Common between Sel A-B
NC_000915	454329	454507	HP0437	4,28	IS605 transposase TnpA	Common between Sel A-B
NC_000915	769890	769985	HP0715	4,13	ABC transporter ATP-binding protein	Common between Sel A-B
NC_000915	1061927	1062105	HP0998	4,04	IS605 transposase TnpA	Common between Sel A-B
NC_000915	614191	614470	HP0583	3,61	hypothetical protein	Common between Sel A-B
NC_000915	951646	951852	HP0898	3,60	hydrogenase expression/formation protein HypD	Common between Sel A-B
NC_000915	1030389	1030578	HP0970	3,52	nickel-cobalt-cadmium resistance protein NccB	Common between Sel A-B
NC_000915	1380289	1380602	HP1319	3,47	50S ribosomal protein L3	Common between Sel A-B
NC_000915	1187149	1187265	HP1121	3,21	cytosine specific DNA methyltransferase	Common between Sel A-B
NC_000915	797809	798321	HP0743	3,09	rod shape-determining protein MreB	Common between Sel A-B
NC_000915	1579060	1579204	HP1506	2,86	glutamate permease GltS	Common between Sel A-B
NC_000915	512849	513141	HP0488	2,70	hypothetical protein	Common between Sel A-B
NC_000915	1278378	1278463	HP1201	2,61	50S ribosomal protein L1	Common between Sel A-B
NC_000915	1410757	1410897	HP1350	2,60	protease	Common between Sel A-B
NC_000915	731951	732085	HP0681	2,48	hypothetical protein	Common between Sel A-B
NC_000915	11805	12001	HP0013	2,33	hypothetical protein	Common between Sel A-B
NC_000915	51391	51625	HP0052	2,30	hypothetical protein	Common between Sel A-B
NC_000915	79748	80106	HP0075	2,10	phosphoglucosamine mutase	Common between Sel A-B
NC_000915	657675	658335	HP0613	1,76	ABC transporter ATP-binding protein	Common between Sel A-B
NC_000915	1295199	1295391	HP1218	1,71	glycinamide ribonucleotide synthetase PurD	Common between Sel A-B

10.9. Common Domains between Selection B-C

CHR	START	END	GENE_ID	FOLDCHANGE	ANNOTATION	FLAG
NC_000915	524080	524265	HP0498	3,72	sodium- and chloride-dependent transporter	Common between Sel B-C
NC_000915	890884	891033	HP0839	3,41	outer membrane protein P1 ompP1	Common between Sel B-C
NC_000915	1037115	1037287	HP0974	2,74	phosphoglyceromutase	Common between Sel B-C
NC_000915	1243612	1243988	HP1177	2,60	hypothetical protein	Common between Sel B-C
NC_000915	1034744	1034877	HP0973	2,07	hypothetical protein	Common between Sel B-C
NC_000915	1317869	1317975	HP1243	1,90	hypothetical protein	Common between Sel B-C
NC_000915	999625	1000073	HP0939	1,76	amino acid ABC transporter permease	Common between Sel B-C
NC_000915	1584941	1585092	HP1512	1,73	iron-regulated outer membrane protein FrpB	Common between Sel B-C
NC_000915	1184140	1184277	HP1118	1,54	gamma-glutamyltranspeptidase	Common between Sel B-C

9.10. List of strains used for comparative analysis

Strain Name	Strain ID
H. pylori strain ausabrJ05	CP011485
H. pylori strain PNG84A	CP011487
H. pylori strain ML3	AP014712
H. pylori strain ML3	AP014713
H. pylori strain ML2	AP014711
H. pylori strain ML1,	AP014710
H. pylori strain L7	CP011482
H. pylori strain K26A1	CP011486
H. pylori strain DU15	CP011483
H. pylori strain CC33C	CP011484
H. pylori strain BM013B	CP007606
H. pylori strain BM013A	CP007604
H. pylori strain BM012B	CP007605
H. pylori strain 29CaP	CP012907
H. pylori oki828	CP006826
H. pylori oki673	CP006825
H. pylori oki422	CP006824
H. pylori oki154	CP006823
H. pylori oki128	CP006822

H. pylori oki112	CP006821
H. pylori oki102	CP006820
H. pylori XZ274	CP003419
H. pylori UM299	CP005491
H. pylori UM298	CP006610
H. pylori UM066	CP005493
H. pylori UM037	
H. pylori UM032	CP005490
H. pylori SouthAfrica7	CP002336
H. pylori SouthAfrica20	CP006691
H. pylori Shi470	NC_010698
H. pylori Shi417	CP003472
H. pylori Shi169	CP003473
H. pylori Shi112	CP003474
H. pylori Sat464	CP002071
H. pylori SNT49	CP002983
H. pylori SJM180	NC_014560
H. pylori Rif2	CP003906
H. pylori Rif1	CP003905
H. pylori Puno135	CP002982
H. pylori Puno120	CP002980
H. pylori PeCan4	NC_014555
H. pylori PeCan18	CP003475
H. pylori P12	NC_011498
H. pylori OK310	AP012601
H. pylori OK113	AP012600
H. pylori NY40	AP014523
H. pylori Lithuania75	CP002334
H. pylori J99	NC_000921
H. pylori J166	CP007603
H. pylori India7	CP002331
H. pylori Hp238	CP010013
H. pylori HUP-B14	CP003486
H. pylori HPAG1	NC_008086
H. pylori Gambia94/24	CP002332
H. pylori G27	NC_011333
H. pylori F57	AP011945
H. pylori F32	AP011943
H. pylori F30	AP011941
H. pylori F16	AP011940
H. pylori ELS37	CP002953
H. pylori Cuz20	CP002076
H. pylori BM012S	CP006889
H. pylori BM012A	CP006888
H. pylori B8	NC_014256

H. pylori B38	NC_012973
H. pylori Aklavik86	CP003476
H. pylori Aklavik117	CP003483
H. pylori 908	CP002184
H. pylori 83	CP002605
H. pylori 52	CP001680
H. pylori 51	CP000012
H. pylori 35A	CP002096
H. pylori 26695-1MET	CP010436
H. pylori 26695-1CL	AP013356
H. pylori 26695	NC_000915

ACKNOWLEDGEMENTS

This work was supported by a grant from the Italian Ministry of Education and Univiersity (PRIN-2010P358BR_002, entitled: "Unraveling structural and functional determinants behind *Helicobacter pylori* pathogenesis and persistence")

Firstly, I would like to express my sincere gratitude to Dr. Clelia Peano for the continuous support of my Ph.D study and related research, for her patience, motivation, and knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better support for my Ph.D study.

I would like to thank the University of Milan, the Ph.D School in Molecular Medicine and Translational and Prof. Cristina Battaglia for the great opportunity to start my PhD and the training support.

My sincere thanks also goes to all the people that contributed to this project: Prof. Daniele Sblattero (Università degli Studi di Trieste), Prof. Mario Milco D'Elios (Università degli Studi di Firenze), Dr. Maria Felicia Soluri (Università degli Studi del Piemonte Orientale), Prof. Alberto Danielli, Dr. Eva Pinatel and Dr. Gianluca De Bellis (ITB-CNR). They provided a friendly and cooperative atmosphere and also useful feedback and insightful comments on my work.

I would like to thank the various members of the Genomics group of ITB-CNR with whom I had the opportunity to work and have not already mentioned: Dr. Roberta Bordoni, Dr. Clarissa Consolandi, Dr. Marco Severgnini, Dr. Luca Petiti, Dr. Ingrid Cifola, Dr. Eleonora Mangano, Dr. Camilla Ceccarani, Giada Caredda and Maria Vurchio.

PUBLICATIONS

1. **Puccio S**, Chu J, Praticò D - Involvement of 5 lipoxygenase in the corticosteroid-dependent amyloid beta formation: in vitro and in vivo evidence. PLoS ONE 6(1): e15163.

2. **Puccio S**. 5-Lipoxygenase and the corticosteroid-dependent abeta formation. Alzheimer's& Dementia: The Journal of the Alzheimer's Association 7.4 (2011): S402.

3. **Puccio S** & Soluri MF , Touati E, Pinatel E , DellaBella C, D'Elios MM, Sblattero D and Peano C. Identification of *H. pylori* epitopes responsible for host immuno-response modulation through ORF-filtered phage display libraries and Interactome- sequencing. Helicobacter Vol 19, Suppl. 1, pag. 154, n. P14.01. September 2014. ISSN 1083- 4389.

4. Gourlay,L.J, Peano,C., Deantonio,C., Perletti,L., Pietrelli,A., Villa,R., Matterazzo,E., Lassaux, P., Santoro, C., **Puccio, S**., Sblattero, D., Bolognesi, M. (2015) Selecting soluble/foldable protein domains through single-gene or genomic ORF filtering: structure of the head domain of *Burkholderia pseudomallei* antigen BPSL2063. Acta Cryst D.

5. Benedet, M., Falchi, F. A., **Puccio, S**., Di Benedetto, C., Peano, C., Polissi, A., & Dehò, G. (2016). The Lack of the Essential LptC Protein in the Trans-Envelope Lipopolysaccharide Transport Machine Is Circumvented by Suppressor Mutations in LptF, an Inner Membrane Component of the Escherichia coli Transporter. *PloS One*, *11*(8), e0161354.

6. Capra E. and Cremonesi P., Pietrelli A., **Puccio S**., Luini M., , Stella A., Castiglioni B. (2017) Genomic and transcriptomic comparison among *Staphylococcus aureus* strains associated with high and low within herd prevalence of intra-mammary infection. BMC Microbiology.

7. Pelliciari S, Pinatel S,Vannini A, Peano S, **Puccio S**, De Bellis G, Danielli A, Scarlato V and Roncarati D. (2017) Insight into the essential role of the *Helicobacter pylori* HP1043 orphan response regulator: genome-wide identification and characterization of the DNA-binding sites. Scientific Report.

8. Vannini A & Pinatel E, Costantini P, Pelliciari S, Roncarati D, **Puccio S**, De Bellis G, Peano C and Danielli A. The NikR regulon: a key regulatory network linking nickel homeostasis and persistence in Helicobacter pylori. (Submitted to Scientific Report).

9. **Puccio S** & Grillo G, Liciulli F, Severgnini M,Liuni S,Bicciato S,De Bellis G, Ferrari F & Peano C. WoPPER: webserver for position related data analysis of gene expression in prokariotes. (Submitted to Scientific Nucleid Acid Research - Web Server Issue 2017).

POSTERS

1. **Puccio S** & Soluri MF, Touati E, Pinatel E, Della Bella C, D'Elios MM, Sblattero D, Peano C. - Identification of H. pylori epitopes responsible for host immuno-response modulation through ORF-filtered phage display libraries and Interactome-Sequencing. "11th International Workshop on Pathogenesis and Host Response in Helicobacter Infections" Organized by: "European Study Group on Pathogenesis and Immunology in Helicobacter Infections - ESGPIHI" - Helsingor (Denmark) July 2-5 2014.

2. Pinatel E & Vannini A, Roncarati D, **Puccio S**, Scarlato V, Peano C and Danielli. “ChIP-seq analysis of the Helicobacter pylory Transcriptional Regulatory Network involved in metal homeostasis control”. 11th International Workshop on Pathogenesis and Host Response in Helicobacter Infections. "11th International Workshop on Pathogenesis and Host Response in Helicobacter Infections" Organized by: "European Study Group on Pathogenesis and Immunology in Helicobacter Infections - ESGPIHI" - Helsingor (Denmark) July 2-5 2014.

3. **Puccio S** & Soluri MF, Touati E, Pinatel E, Della Bella C, D'Elios MM, Sblattero D, Peano C. - Identification of H. pylori epitopes responsible for host immuno-response modulation through ORF-filtered phage display libraries and Interactome-Sequencing. " XXVIIth International Workshop on Helicobacter & Microbiota in Inflammation & Cancer" Organized by: "European Helicobacter and Microbiota Study Group - EHMSG" - Rome 11-13 September 11-13 2014.

4. Pinatel & Vannini A, Roncarati D, **Puccio S**, Scarlato V, De Bellis G, Danielli A and Peano C. “ChIP-seq and RNA-seq data analysis integration to re-define NikR role in the *Helicobacter pylori* Transcriptional Regulatory Network”. BACNET 2015 9-15 May 2015 San Feliu de Guixols

5. **Puccio S** & Soluri MF, Touati E, Pinatel E, Della Bella C, D'Elios MM, Sblattero D and Peano C. "Identification of *H. pylori* epitopes responsible for host immuno-response modulation through ORF-filtered phage display libraries and Interactome-sequencing" - "6th Congress of European Microbiologists" Organized by: "Federation of Europeans Microbiological Society FEMS" - Maastricht June 7-11 2015.

6. Benedet M, Falchi F, **Puccio S**, Di Benedetto C,Peano C, Polissi A, and Dehò G. - Characterization of the lipopolysaccharide transport machine in *Escherichia coli*: in search of a function for the elusive component LptC - "6th Congress of European Microbiologists" Organized by: "Federation of Europeans Microbiological Society FEMS" - Maastricht June 7-11 2015.

7. Falchi F, Sperandeo P, Maccagni E,Peano C, **Puccio S**, Polissi A, Dehò G. - Mutational analysis of LptA, an essential LPS-transport protein in *E. coli.* "6th Congress of European Microbiologists" Organized by: "Federation of Europeans Microbiological Society FEMS" - Maastricht June 7-11 2015.

8. Pinatel & Vannini A, Roncarati D, **Puccio S**, Scarlato V, De Bellis G, Danielli A and Peano C. - ChIP-seq and RNA-seq data integration to systematically (re)define the role of NikR in the *Helicobacter pylori* Transcriptional Regulatory Network. "6th Congress of European Microbiologists" Organized by: "Federation of Europeans Microbiological Society FEMS" - Maastricht June 7-11 2015.

9. Pinatel & Vannini A, Roncarati D, **Puccio S**, Scarlato V, De Bellis G, Danielli A and Peano C. "ChIP-seq and RNA-seq data analysis integration to re-define NikR role in the *Helicobacter pylori* Transcriptional Regulatory Network".31° Meeting della SIMGBM 23-26 settembre 2015 Ravenna.

10. **Puccio S** & Pinatel E, Grillo G, Licciulli F, Ferrari F, Liuni S, Severgnini M, Vannini A, De Bellis G, Danielli A and Peano C. HelicoMINE: an integrated DataBase and WebTools platform for *Helicobacter pylori* functional data analysis and visualization. "12th International Workshop on Pathogenesis and Host Response in Helicobacter Infections" Organized by: "European Study Group on Pathogenesis and Immunology in Helicobacter Infections" Helsingor 29 June2 July 2016.

ORAL PRESENTATION

1. **Puccio S,** Pinatel E, Pietrelli A, Rossi E, Consolandi C, Peano C - Comparison of two different computational methods for the prediction of new *E.coli* ncRNAs from directional RNA sequencing data.