Genomic characterisation and comparison of antibiotic resistant and sensitive *Helicobacter pylori* in New Zealand

Zoe King

A thesis submitted to Auckland University of Technology in partial fulfilment of the requirements for the degree of Master of Science (Research) (MSc(Res))

2020

School of Science

Abstract

Helicobacter pylori is an important and prevalent human pathogen present in approximately 50 % of the world's population; causing a range of clinical outcomes from chronic gastritis and gastric ulcers to gastric carcinoma and gastric mucosa-associated lymphoid tissue lymphoma (Hooi et al. 2017; Huang et al. 2002; Kuipers et al. 1995). Treatment of this pathogenic bacterium is becoming increasingly challenging due to the increasing levels of antibiotic resistance worldwide (Malfertheiner et al. 2017; Savoldi et al. 2018). To effectively treat this bacterium, local resistance rates must be known for clinicians to administer the appropriate antibiotics for the highest chance of eradication (Malfertheiner et al. 2017). Within New Zealand (NZ) there is a lack of data on antibiotic resistance rates and levels, as well as a lack of knowledge on the genomic mechanisms causing resistance. This study aimed to increase the understanding of the levels of antibiotic resistant strains of *H. pylori* in NZ, as well as elucidate the underlying mechanisms causing resistance.

Five *H. pylori* strains (A, B, C, D, and E), provided by Middlemore Hospital (Auckland, NZ) from patients undergoing gastroscopy were subject to antibiotic susceptibility testing, phylogenetic analysis, and comparative genomic analysis to determine their resistance profiles, genomic relationships, genomic characterisations and to identify known and novel mechanisms of resistance.

Antibiotic susceptibility testing was performed using two methods, in-house using disc diffusion, and by Middlemore Hospital using Epsilometer test (E-test) strips. In-house testing was unsuccessful. Results obtained from Middlemore Hospital were used to determine the resistance profiles of each isolates. Each isolate was tested against commonly used antibiotics in treatment of *H. pylori* infections clarithromycin, metronidazole, amoxicillin, and tetracycline. One isolate (A), was sensitive to all antibiotics tested, two isolates (B and E), were resistant to clarithromycin and metronidazole at > 256 µg/mL each, and two isolates (C and D), were resistant to only clarithromycin at 24 µg/mL and 1 µg/mL, respectively. Isolate D did not survive the experimental process and was not investigated further.

Sanger sequencing was initially used to sequence the 16S rRNA region to provide preliminary sequence-based identification of the four remaining isolates (A-C and E). Phylogenetic analysis was completed using the 16S rRNA gene, MLST genes, and PhyloPhlAn2 (400 genes) extracted from Illumina based sequencing of each isolate (A-C and E). The relationships of the isolates (A-C and E) were assessed with a dataset of 155 complete *H. pylori* genomes. The 16S rRNA gene and MLST genes highlighted the genomic diversity of the *H. pylori* species and identified each

isolate (A-C and E) as separate strains, likely arising from different sources. Also using the draft genomes, a phylogenetic analysis using PhyloPhlAn2 was employed using which increased the resolution of the trees but still lacked enough sequence information to decipher the interspecies relationships of the dataset. However, the closest relatives to each isolate (A-C and E) were identified from this dataset using the nucleotide sequence PhyloPhlAn2 tree and used for further analysis.

The draft genomes of isolates A-C, and E, were also used for comparative genomics with their respective closest references. Characterisation of the genomes showed similar characteristics to other *H. pylori* isolates. Comparison of the genomes to identify known mechanisms of resistance showed that the A2147G mutation within domain V of the 23S rRNA gene, conferring clarithromycin resistance, was only present in isolates B and C. The R16H amino acid substitution within RdxA, suggested to be involved with metronidazole resistance, was only identified in isolate B. Efflux pumps HefABC, HP1181, and HP1184 were identified in all isolates (A-C and E) and may be a common mechanism of resistance amongst all isolates, however, further work is required to establish their role in antibiotic resistance. The search for novel resistance mechanisms using CARD and OrthoVenn2 identified a major facilitator superfamily (MFS) efflux protein and an outer membrane protein (OMP) identified only in resistant isolates (B, C, and E), which are of further interest to elucidate their potential involvement in antibiotic resistance.

Overall, these results suggest that both clarithromycin and metronidazole resistance are present in NZ *H. pylori* strains. Antibiotic resistance appears to develop independently and through different methods based on the phylogenetic methods employed in this study. None of the resistant isolates (B, C, and E) shared one common mechanism of resistance, however, the presence of efflux pumps, in both sensitive (A) and resistant (B, C, and E) strains, suggests that this mechanism may play an important role in resistance. More work is required to fully elucidate the role of efflux pumps and the two novel proteins in antibiotic resistant *H. pylori*.

Table of Contents

List of Figures

List of Tables

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Zoe King

Date: 05/11/2020

Acknowledgements

The journey of completing my Masters was one that many people helped with and I am grateful to be able to offer my sincere thanks to all involved.

Firstly, I would like to thank my primary supervisor Dr Brent Seale. Thank you, Brent, for allowing me to work with you and for all your knowledge and encouragement throughout this time. This experience has been challenging, but you were always there with answers to my questions and reassurance, thank you. Secondly, I would like to thank my secondary supervisor, Dr Colleen Higgins. Thank you for your constant endeavour to teach your students, your support, expertise on molecular genetics and your kindness. You have helped push me to become the person and scientist I am today.

A huge thank you to Dr Susan Tylor and the team at Middlemore Hospital for supplying the *H. pylori* isolates used in this thesis and providing guidance through the challenging, initial stages of this project. This project was not doable without your contribution.

Thank you, Dr Kevin Lee, for your immense knowledge of bioinformatics and for always having suggestions on paths to follow for my project. I would also like to thank all of the AUT staff and lab technicians, especially Saeedah, for helping me with my experiments. To Dr Elizabeth Buckley, your constant support and knowledge has helped me immensely throughout this project, thank you. As well as this, thank you for being a friend. I cannot thank you enough for the times I needed to vent and the words of kindness and encouragement you gave me.

To the Molecular Genetics Research Group, I thank you all for your acceptance, listening to my progress and helping me troubleshoot my problems. To Toni, Lee and Shweta thank you all for the debriefing sessions after the meetings and providing connection when I felt lost.

To my Mum and Dad, Jean and Tim, my twin pillars without whom I could not stand, I thank you both for your boundless support and endless encouragement. To my Dad, thank you for all the times you listened to my problems, gave me practical solutions and when needed, told me to suck it up. To my Mum thank you for your constant encouragement, wise words and helping me believe in myself. I love you both.

To my Nana, Mary, I thank you for always being there to talk about my studies and asking me insightful questions about my project. To my other family and friends, of which there are too many to name, I thank you all for being there for me and giving me the time and space to vent about problems and enjoy myself away from my studies.

And finally, I would like to thank Amber, for always being there and supporting me, having patience with me, and always being there for a laugh, thank you.

Abbreviations

%	Percentage
°C	Degrees Celsius
g	Gravity
Gb	Giga base
kbp	Kilo base pair
M	Molar
m/s	Meters per second
Mbp	Mega base pair
MIC	Minimum Inhibitory Concentration
mg/L	Micro gram per litre
mL	Millilitre
mm	Millime meter
mM	Millimolar
nm	Nanometre
pH	Power of hydrogen
rpm	Revolutions per minute
V	Volts
v/v	Volume per volume
µg	Microgram
µm	Micrometre
ABC	ATP-binding cassette superfamily
ARO	Antibiotic resistance ontology
AUT	Auckland University of Technology
BHI	Brain heart infusion
BIC	Bayesian information criterion
BLAST	Basic local alignment search tool
CAC	Citric acid cycle
KOALA	KEGG ortholog and links annotation
LCB	Local collinear block
MALToma	Mucosa-associated lymphoid tissue lymphoma
MATE	Multi antimicrobial extrusion protein family
MCM	Mauve contig mover
MFP	Membrane fusion protein
MFS	Major facilitator Ssperfamily
MH	Mueller Hinton
MLST	Multi locus sequence typing
mRNA	Messenger ribonucleic acid
MUSCLE	MUlitple sequence comparison by log-expectation
NAD	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide + hydrogen
NADPH	Nicotine adenine dinucleotide phosphate
NGS	Next generation sequencing
nt	Nucleotide
NTC	No template control
NZ	New Zealand
OG	Orthologous group
OMP	Outer membrane protein
OPC	Orthologous protein cluster
ORF	Open reading frame
PacBio	Pacific Bioscience
PBP	Penicillin-binding protein
PE	Pair-end
PHASTER	PHAge search tool enhanced release
PP	Per protocol

CARD	Comprehensive Antibiotic Resistance Database	**PPI**	Proton pump inhibitor
Cas	CRISPR-associated	**RAST**	Rapid Annotations using Subsystems Technology
CDS	Coding sequence	**RCT**	Randomized clinical trial
CFU	Colony forming units	**RDP**	Ribosomal Database Project
CLSI	Clinical Laboratory Standards Institute	**RGI**	Resistance gene identifier
CoA	Acetyl-coenzyme A	**RNA**	Ribonucleic acid
COG	Clusters of orthologous groups/genes	**RND**	Resistance-nodulation-cell division superfamily
CRISPR	Clustered regularly interspaced short palindromic repeats	**RQ**	Research question
CSB	Columbia sheep blood	**rRNA**	Ribosomal ribonucleic acid
CTAB	Cetyltrimethylammonium bromide-polyvinylpyrrolidone	**SDS**	Sodium dodecyl sulphate
ddNTP	Dideoxynucletoide	**SMH**	Supplemented Mueller Hinton
DIAMOND	Double index alignment of next-generation sequencing data	**SMR**	Small multidrug resistance family
DNA	Deoxyribonucleic acid	**SMRT**	Single-molecule, real-time sequencing
dNTP	Deoxynucleotide	**SNP**	Single nucleotide polymorphism
DR	Direct repeat	**SOD**	Superoxide dismutase
EC	Enzyme commission	**SOLiD**	Sequencing by oligonucleotide ligation and detection
E-test	Epsilometer test	**ST**	Sequence type
EUCAST	European Committee on Antimicrobial Susceptibility Testing	**T4SS**	Type IV secretion system
FMN	Flavin mononucleotide	**TBE**	Tris-Borate ethylenediaminetetraacetic acid
gDNA	Genomic deoxyribonucleic acid	**tRNA**	Transfer ribonucleic acid
GO	Gene ontology	**VBNC**	Viable but non culturable
IS	Insertion sequence	**VFDB**	Virulence Factor Database
iTOL	Interactive tree of life	**WHO**	World health organisation
ITT	Intention-to-treat	**ZMW**	Zero-mode waveguide
KEGG	Kyto encyclopaedia of genes and genomes		

Chapter 1 General Introduction and Literature Review

1.1 General Introduction

Antibiotic resistance is an increasing worldwide issue affecting people from all socioeconomic backgrounds (Laxminarayan et al. 2016). Naturally occurring resistance is hastened by the continued overuse and misuse of therapeutic drugs and through the transfer and acquisition of resistance genes (Laxminarayan et al. 2016). The impact of antibiotic resistance extends beyond the inability to treat infections to the economic strain it puts on health care systems; and the effects on patients and their families (Naylor et al. 2019; Shrestha et al. 2018). It is therefore of utmost importance to understand and control the spread and development of antibiotic resistance.

H. pylori is found worldwide, and approximately 50 % of the world's population are infected with this pathogen (Hooi et al. 2017). Infection can lead to acute and chronic gastritis, peptic ulcers, non-ulcer dyspepsia and in severe cases, gastric carcinoma (Huang et al. 2002; Kuipers et al. 1995). Treatment options for these various clinical outcomes of infection vary in different countries (Chey and Wong 2007). The most common first-line treatment option is a triple therapy containing a proton pump inhibitor (PPI) and two antibiotics (e.g. amoxicillin, clarithromycin, metronidazole and levofloxacin). Other options include quadruple, contaminant, sequential and hybrid therapies with similar antibiotics (Chey and Wong 2007). Choice of treatment depends on a variety of factors but of increasing importance are the local resistance rates to first-line treatment antibiotics (Malfertheiner et al. 2017).

The World Health Organisation (WHO) has identified *H. pylori* as a high priority pathogen requiring new antibiotics for treatment (Tacconelli et al. 2018). Resistance rates vary globally but a severe increase has been observed in some areas (Savoldi et al. 2018). High resistance is seen with clarithromycin and metronidazole, where highest resistance is 34 % in Western Pacific regions and 91 % in African regions (Savoldi et al. 2018). Levofloxacin resistance is highest in Southeast Asia at 25 %, whereas amoxicillin and tetracycline have low resistance of under 15 %; except in Africa where resistance for Amoxicillin is 38 % (Savoldi et al. 2018). This global increase in resistance to crucial antibiotics has a direct impact on the successful eradication and treatment of *H. pylori* infections (Park et al. 2016).

Within New Zealand (NZ), *H. pylori* infection and antibiotic resistance rates are severely understudied, with only three publicly available studies on antibiotic resistant *H. pylori* in NZ (Ahmed et al. 2004; Fraser et al. 1999; Hsiang et al. 2013). This gap in knowledge provides an

essential research opportunity to further understand the rate of resistance to essential antibiotics and the underlying mechanisms of resistance within *H. pylori* isolates from NZ.

Understanding the mechanisms behind developed and acquired resistance in *H. pylori* is important to help develop alternative treatment regimens. Discovery and awareness of these mechanisms of resistance has been made easier through the rapid development of comparative genomics. Comparative genomics is the science of comparing whole genomes or genomic features of different organisms. This approach identifies similarities and differences of the examined genomes and may also be utilised to further the understanding of the biology of individual genomes (Wei et al. 2002). Previous studies on antibiotic resistant organisms such as *Acinetobacter baumannii* and *Klebsiella pneumoniae* have utilised comparative genomics to identify specific genes that confer antibiotic resistance but none have been carried out for *H. pylori* in NZ (Adams et al. 2008; Kumar et al. 2011).

Increasing antibiotic resistance worldwide is an important and on-going issue that requires urgent attention. With the high infection rate and potentially severe clinical outcomes of *H. pylori* infections, coupled with its increasing worldwide resistance and understudied nature in NZ, shows that this pathogen is an important organism to study. The purpose of this thesis is to increase the knowledge of mechanisms associated with the increasing antibiotic resistance of *H. pylori* in NZ using comparative genomic analysis. The aims of this project are listed below and can be viewed in Figure 1.1.

1.2 Research Questions and Aims

Research question (RQ)1: What antibiotics are *H. pylori* resistant to and at what concentrations?

Aim 1.1: To determine what antibiotics *H. pylori* are resistant to.

Objective 1.1.1: Disc assays and Epsilometer test (E-test) minimum inhibitory concentration (MIC) strips of different antibiotics will be placed onto agar containing different *H. pylori* strains. Zones of inhibition will be measured and recorded.

RQ2: What is the phylogenetic relationship between antibiotic resistant and sensitive *H. pylori* determined in research question 1?

Aim 2.1: To examine the 16S ribosomal ribonucleic acid (rRNA) phylogenetic relationship of antibiotic resistant and sensitive *H. pylori*.

Objective 2.1.1: Sequencing of the 16S rRNA gene will be used to create a phylogenetic tree of known isolates and related reference sequences.

Aim 2.2: To obtain the genome of the resistant and sensitive isolates and create a phylogenetic tree using PhyloPhlAn2.

Objective 2.2.1: Genome sequencing: Deoxyribonucleic acid (DNA) extractions of appropriate *H. pylori* isolates will be performed. Extractions will then be sent for whole genome sequencing (WGS).

Objective 2.2.2: To use the genome obtained in objective 2.2.1 to create a phylogenetic tree using PhyloPhlAn2 with other known complete genome reference sequences.

Aim 2.3: To determine the multi locus sequence type (MLST) profile for each isolate.

Objective 2.3.1: To use the genome obtained in objective 2.2.1 to identify relevant MLST genes and complete MLST analysis.

Aim 2.4: Comparison of 16S rRNA and whole genome phylogenies.

Objective 2.4.1: Phylogenetic trees created from aims 2.1 and 2.2 will be compared to observe relationships between resistant and sensitive isolates and reference sequences.

RQ3: What is the genetic basis of the resistance identified in research question 1?

Aim 3.2: To identify any genomic variations between sequenced genomes from aim 2.2.

Objective 3.2.1: Comparative genomics: WGS data will be analysed using comparative genomics of resistant and susceptible strains of *H. pylori* to identify differences in gene order, gene content and mutations.

Aim 3.3: To predict antibiotic resistance of complete *H. pylori* genomes using the resistant genes and identified in aim 3.2.

Objective 3.3.1: Search for genes identified in aim 3.2 among a compiled list of complete *H. pylori* genomes.

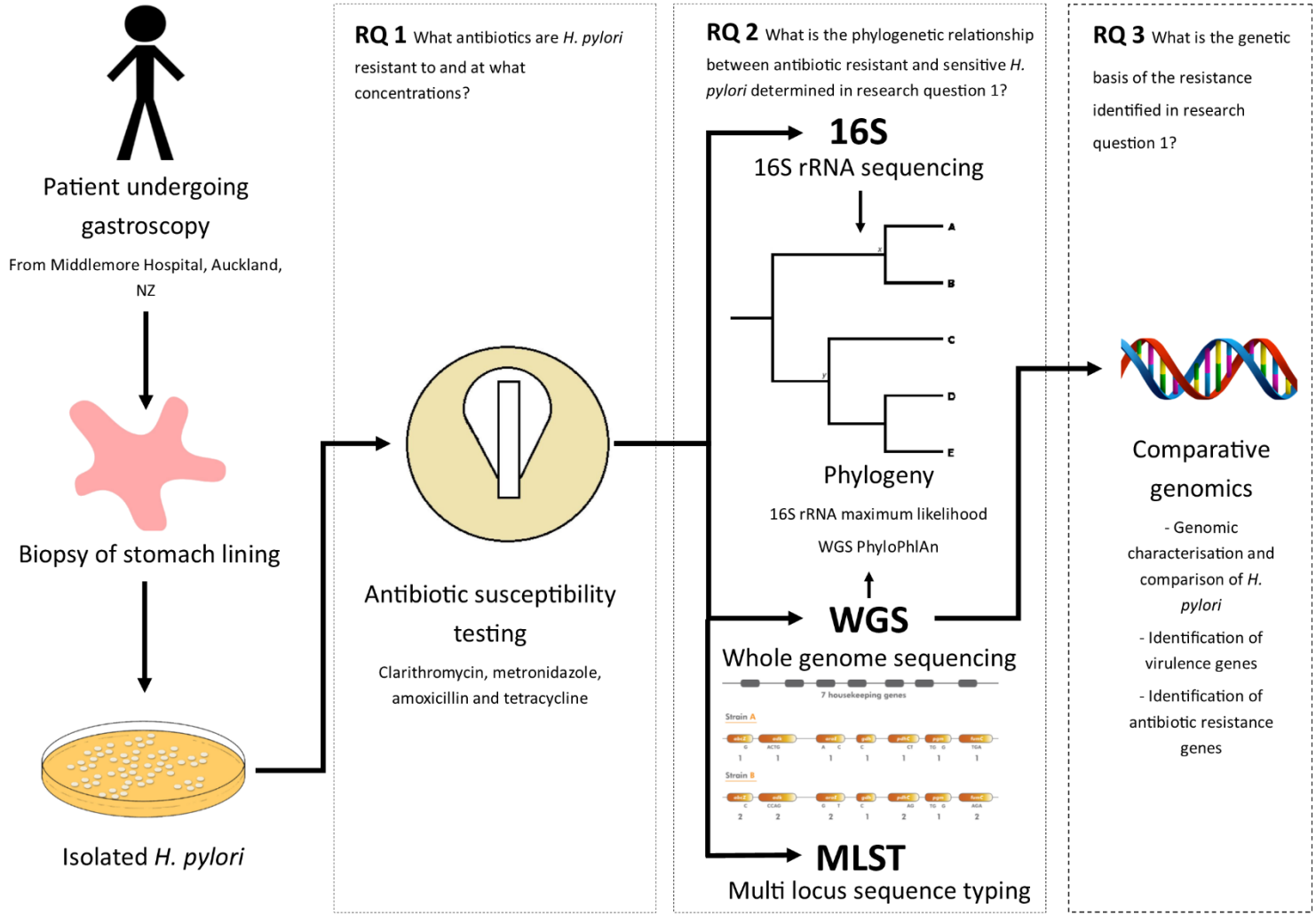

Figure 1.1: Schematic diagram of the RQs and aims of this

1.3 Literature Review

1.3.1 *Helicobacter pylori*

In 1983, Warren and Marshall (1983), published their paper on the discovery of an unidentified, curved bacilli present in patients with active chronic gastritis. A second paper published in 1984 detailed the classification of this unknown bacteria into the genus *Campylobacter* and showed that the bacteria was present in most patients with active chronic gastritis, duodenal ulcers and/or gastric ulcers (Marshall and Warren 1984). The epithet of *Campylobacter pyloridis*, to describe this spiral shaped bacterium, was revised to *Campylobacter pylori* in 1987 (Marshall and Goodwin 1987).

Although *C. pylori* was observed in most patients with gastritis and peptic ulcer disease, confirmation of the aetiological role of this organism was still required. A volunteer, Dr Barry Marshall, who showed normal signs of gastric mucosa and gastric pH, ingested a live culture of *C. pylori*. After three days symptoms of gastritis ensued and after eleven days antral and fundal biopsies revealed histological gastritis and *C. pylori* was cultured (Sobala et al. 1991). This experiment established the aetiological role of acute upper gastrointestinal illness and acute antral gastritis to *C. pylori* (Marshall 2008; Morris and Nicholson 1987; Sobala et al. 1991).

Taxonomic classification originally placed this bacterium within the *Campylobacter* genus, due to similarities in morphology, atmospheric requirements and DNA base composition (Marshall and Warren 1984). However, it was noted that the presence of four sheathed flagella was outside the norm of Campylobacters. Analysis of the 16S rRNA gene demonstrated that *C. pylori* was not related to the *Campylobacter* genus due to its differential grouping from other *Campylobacter* species, but was more closely related to *Wolinella succinogenes* (Romaniuk et al. 1987). This analysis, along with further chemotaxonomic properties that differed from the genus *Wolinella* and closely related taxa, resulted in the reclassification of this organism into a new genus: *Helicobacter*; with *H. pylori* as the type species (Goodwin et al. 1989).

1.3.1.1 Morphology

Morphologically, this Gram-negative bacterium has two forms; a spiral shaped rod (spiral) and coccoid (Wang et al. 1997). The bacteria were first noted as curved rods or S-shaped profiles with two to three curves, approximately 0.5 micrometres (µm) wide and 3 µm long (Figure 1.2 A and B) (Geis et al. 1989; Goodwin et al. 1985b; Jones et al. 1985). During lag and exponential phase of growth, *H. pylori* are primarily spiral shaped, however, the death phase shows primarily coccoid morphology (Figure 1.2 C and D) (Worku et al. 1999).

Although spiral shaped bacilli are the more predominant morphological form of *H. pylori* found in the stomach, coccoid forms have also been identified (Chan et al. 1994; Ogata et al. 1998). The coccoid forms of the bacteria are unable to be cultured *in vitro* but are able to be induced through placing *H. pylori* under physical or chemical stresses such as the introduction of antibiotics, an aerobic environment, temperature changes, pH changes, extended incubation and nutrient availability (Bode et al. 1993; Kusters et al. 1997). Morphological transformation into coccoid shapes has been proposed as the manifestation of the cell's death (Kusters et al. 1997). However, other studies have shown that the coccoid form is associated with cell adaptation in sub-optimal conditions and allows the bacteria to enter a viable but not culturable (VBNC) state (Azevedo et al. 2007; Cellini et al. 1994; Willén et al. 2000). It has therefore been proposed that this highly adaptable bacteria has three main cellular types: the spiral form that is virulent, able to replicate and induce inflammation; the VBNC coccoid form that is unable to be cultivated on solid media and are able to survive stressful environments; and finally the degenerative and dying spiral and coccoid forms of the bacteria (Saito et al. 2003).

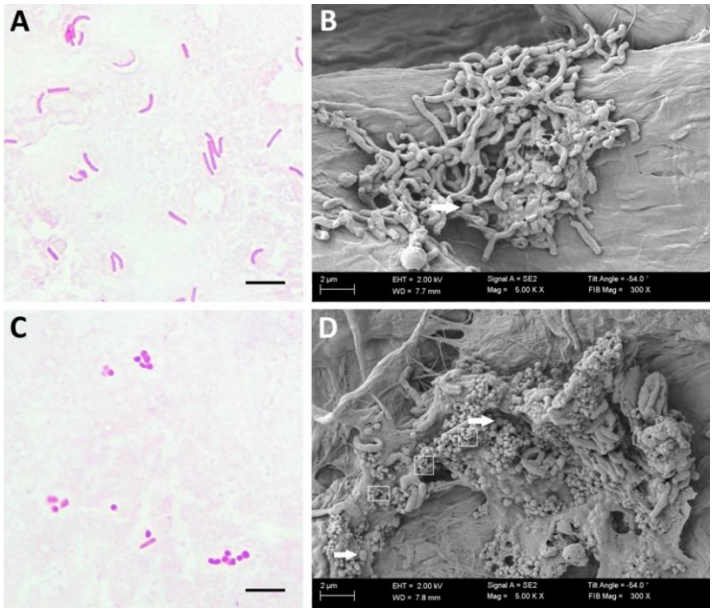

Figure 1.2: Gram stain and electron micrograph of *H. pylori* morphology. A) gram stain of spiral form of *H. pylori*, B) electron microscope image of spiral form of *H. pylori*, C) gram stain of coccoid form of *H. pylori*, D) electron microscope image of coccoid forms of *H. pylori*. Images retrieved from Krzyżek et al. (2019).

1.3.1.2 Flagella

H. pylori can produce both sheathed and unsheathed flagella, with the latter being less prevalent (Kostrzynska et al. 1991; Qin et al. 2017). Four to six sheathed flagella are produced from a single pole of the cell and are associated with pathogenesis of the bacterium (Figure 1.3) (Celli et al. 2009; Kostrzynska et al. 1991; Suerbaum 1995). The flagella are 3 – 5 micrometres (µm) in length and 30 – 35 nanometres (nm) wide. The sheath narrows towards

the end of the flagellum and then extends beyond to create a club-shaped thickening at the end of the sheath (Geis et al. 1989; Kostrzynska et al. 1991). It has been well documented that the flagella of *H. pylori* are essential to its ability to infiltrate the gastric epithelium and cause infection within the host (Eaton et al. 1989; Eaton et al. 1996; Ottemann and Lowenthal 2002). The flagella produced are similar to other known enteric bacteria such as *Escherichia coli* and *Salmonella* species (Geis et al. 1989).

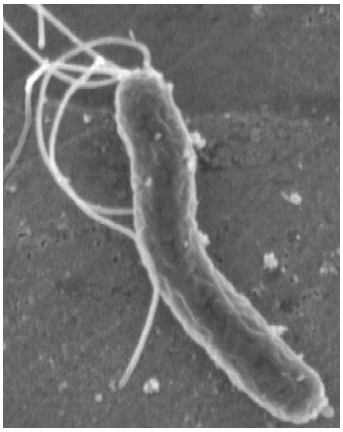

Figure 1.3: Electron micrograph of *H. pylori*, showing their flagella. Image retrieved from https://www.cab.unimelb.edu.au/cab_mucosal.htm.

1.3.1.3 Genomic Characteristics

H. pylori have an average genome size of approximately 1.6 mega base pairs (Mbp) which are arranged in a circular genome (Alm et al. 1999; Tomb et al. 1997). They are AT rich, with an approximate G+C content of 39% and contain approximately 1,500 open reading frames (ORFs) (Alm et al. 1999; Baltrus et al. 2009; Tomb et al. 1997). They contain two copies of the 16S, 23S and 5S rRNA genes (Alm et al. 1999; Tomb et al. 1997). Plasmids are found in around half of sequenced *H. pylori* genomes and contribute to the introduction and shuffling of new chromosomal DNA into *H. pylori* strains (Ali et al. 2015; Alm et al. 1999; Lamichhane et al. 2019; Tomb et al. 1997).

An *H. pylori* genome contains approximately 1,537 protein coding genes (Ali et al. 2015). Salama et al. (2000) showed that within their dataset of 15 clinical isolates, *H. pylori* contained a conserved core of 1,281 genes and 12 - 18 % of their genomes were strain specific. These 1,281 conserved genes encoded for metabolic, cellular, regulatory and biosynthetic functions, whereas the strain specific genes encoded for proteins with no known function or genes that regulate DNA exchange which may promote the high level of genetic diversity seen among *H. pylori*.

H. pylori was the first bacterial species to have two complete genome sequences publicly available (Alm et al. 1999). Comparative genomics of these two strains, J99 and 26995, identified a region dubbed the 'plasticity zone' (Alm et al. 1999). This zone had a G+C content of 35 % (lower than the rest of the genome, 39%), contained approximately 47% of each strain's specific genes and was located in the same chromosomal position. The plasticity zone term was used to describe these hyper variable regions between *H. pylori* strains. As more complete *H. pylori* genomes and sequencing data became available these plasticity zones were found to harbour clusters of complete gene sets including type IV secretion system genes (for pathogenicity) and genes involved in DNA-processing (Kersulyte et al. 2009; Kersulyte et al. 2003). These plasticity zones have since been reclassified as conjugative transposons or integrating conjugative elements and have been found in *H. pylori* strains worldwide (Fischer et al. 2014; Kersulyte et al. 2009).

H. pylori are known for their high level of genetic diversity (Akopyanz et al. 1992). The high diversity observed among *H. pylori* isolates may be due to the genomic macrodiversity recognised in the lack of conserved gene order (Jiang et al. 1996). Horizontal gene transfer and gene shuffling caused by plasmids, free recombination and a high mutation rate contribute to the macrodiversity of these genomes and classify this genome as panmictic; random mating within a population leading to the genes at a particular locus being independent to that of others (Björkholm et al. 2001; Hofreuter and Haas 2002; Smith et al. 1993; Suerbaum et al. 1998).

Mutation frequency analysis, based on the *rpoB* (β subunit of RNA polymerase) gene in *H. pylori* as an indicator have provided various frequencies ranging from 3×10^{-5} to 4×10^{-8} per site per year (Björkholm et al. 2001). Through genome wide analysis of mutation events, mutation rates of 2.5×10^{-5} per year and site and recombination rates of 5.5×10^{-5} per initiation site and year have been established for *H. pylori* (Kennemann et al. 2011). Recombination frequency drops when *H. pylori* are not part of a mixed infection (Kennemann et al. 2011). The mutation frequency of *H. pylori* is similar to *Neisseria meningitidis* but higher than *E. coli*, *Streptococcus pneumoniae* and *Borrelia burgdorferi* (Suerbaum 2000). These high mutation rates suggest that *H. pylori* has a high frequency of spontaneous mutations (Hänninen and Hannula 2007).

1.3.1.4 Growth in the Lab

Cells of *H. pylori* are slow growing and fastidious (Marshall and Warren 1984). The bacterium has displayed facultative acidophilic behaviour with growth in a pH range of 4.5 to 6.0 with an optimum pH of 5.5 (Kangatharalingam and Amy 1994). It has also been established that this bacterium is an obligate microaerophile, growing best with 8-10 % carbon dioxide levels and being unable to grow in completely anaerobic conditions (Goodwin et al. 1985a;

Kangatharalingam and Amy 1994; Rollason et al. 1984). *H. pylori* has an optimum growth temperature of 37 °C, with a growth range of 30 – 37 °C (Jiang and Doyle 1998). Initial isolation of the bacteria is achieved using selective medias such as Skirrow's or Dent's media due to the presence of antibiotics that *H. pylori* are intrinsically resistant to such as vancomycin and trimethoprim that prevent the growth of other contaminants (Mégraud 1997; Tee et al. 1991).

Laboratory growth of *H. pylori* can be achieved using both solid and liquid media. On solid media, *H. pylori* requires a nutrient rich media containing blood or blood products such as Columbia blood agar base, Mueller Hinton (MH) Agar base or Blood Agar Base supplemented with horse, ox or sheep's blood (Blanchard and Nedrud 2012). Liquid media such as Brain Heart Infusion (BHI) or Brucella broth supplemented with 10 % heat-inactivated foetal bovine serum or 7 % horse serum may also be used to grow *H. pylori* (Blanchard and Nedrud 2012; Jiang and Doyle 2000).

1.3.2 Epidemiology

Global infection rates of *H. pylori* show that infections are highest in Africa, Latin America, the Caribbean and Asia, whereas Northern America and Oceania have low infection rates (Hooi et al. 2017). Many of those affected by this pathogen acquire it during childhood and infection is generally lifelong (Kuipers et al. 1995). The infection may lead to gastritis, peptic ulcers or gastric cancer but the majority of infected individuals are asymptomatic (Mattsson et al. 1998a). Patients with duodenal ulcers, gastric ulcers or non-ulcer dyspepsia who also carry the *H. pylori* antibody have been shown to be infected with *H. pylori* in 85.7 %, 62.5 % and 22.2 % of cases respectively; showing the high correlation of gastric diseases with *H. pylori* (Lee et al. 1993). In 2012, approximately 15 % of cancer cases worldwide were associated with infectious agents; of these, *H. pylori* was the most prevalent (Plummer et al. 2016). High global rates of infection leading to gastric ulcers, dyspepsia and in some cases neoplasia, demonstrates the significance of this pathogen to the health and wellbeing of all individuals.

1.3.2.1 Global Spread of *H. pylori*

H. pylori have inhabited the human stomach for thousands of years, with anatomically modern humans being infected by the pathogen prior to migrations from Africa, and in turn globally spreading along with its host (Linz et al. 2007; Moodley et al. 2012). Due to the extreme sequence diversity of *H. pylori*, this bacterium has been used as a reliable marker for human migration (Moodley and Linz 2009). Through MLST sequences of seven housekeeping genes (ATP synthase, F1 alpha (*atpA*), elongation factor EF-P (*efp*), A/G-specific adenine glycosylase (*mutY*), Inorganic pyrophosphatase (*ppa*), anthranilate isomerase (*trpC*), urease accessory protein (*urel*) and GTPase (*yphC*)) and one virulence gene (*vacA*), using STRUCTURE software, *H. pylori* have

shown groupings based on their continent of origin (Table 1.1) (Achtman et al. 1999; Falush et al. 2003; Moodley and Linz 2009; Moodley et al. 2012; Nell et al. 2013). Further analysis using whole genome data has provided insight into a number of different subpopulations due to additional migrations of modern humans (Table 1.1) (Thorell et al. 2017).

Along with geographical groupings of similar *H. pylori* strains, there is also a distinctive difference in the virulence factors from different geographic areas. Markedly, the difference between East Asian and Western type *cag* pathogenicity islands (*cag*PAI) that are associated with pathogen virulence and pathogenesis of several diseases (Nguyen et al. 2010; Yuan et al. 2017). It has been suggested that the East Asian type are more virulent due to an association with IL-8 secretion and gastric mucosal inflammatory cell infiltration (Yuan et al. 2017).

Table 1.1: Inferred population structure of *H. pylori* isolates based on MLST and whole genome analysis.

Continent of origin	Population	Subpopulation	Reference
Africa	hpNEAfrica	hspEastNEAfrica	(Moodley et al. 2012)
		hspCentralNEAfrica	(Nell et al. 2013)
	hpAfrica1	hspSAfrica	(Nell et al. 2013)
		hspWAfrica	(Nell et al. 2013)
		hspCAfrica	(Nell et al. 2013)
	hpAfrica2	hspNorthSan	(Moodley et al. 2012)
		hspSouthSan	(Moodley et al. 2012)
Europe	hpEurope	hspEuropeN	(Thorell et al. 2017)
		hspEuropeS	(Thorell et al. 2017)
Africa/Europe	hpAfrica1/hpEurope hybrid	hspAfrica1NAmerica	(Thorell et al. 2017)
		hspAfrica1Nicaragua	(Thorell et al. 2017)
		hspMiscAmerica	(Thorell et al. 2017)
		hspEuropeColombia	(Thorell et al. 2017)
Asia	hpEeastAsia	hspAmerind	(Falush et al. 2003)
		hspEAsia	(Falush et al. 2003)
		hspMaori	(Moodley et al. 2009)
	hpAsia2	hspLadakh	(Tay et al. 2009)
		hspIndia	(Tay et al. 2009)
	hpSahul	hspAustralia	(Moodley et al. 2009)
		hspNGuinea	(Moodley et al. 2009)

1.3.2.2 Natural Reservoirs

The natural environment of *H. pylori* is not fully understood. The only known natural reservoir of *H. pylori* is the human stomach (Marshall and Warren 1984; Warren and Marshall 1983). Recently, most research has focused on the oral cavity of humans as another reservoir for *H. pylori*. However, there have been many other suggestions for the potential reservoirs of this pathogenic bacterium, including sheep and natural water sources (Momtaz et al. 2014; Sasaki et al. 1999).

The human oral cavity has been the main focus of research into additional reservoirs of *H. pylori,* with results being inconsistent. Within the oral cavity, dental plaque and saliva have been the

prominent sites where *H. pylori* has been identified (Desai et al. 1991; Gebara et al. 2004; Liu et al. 2009). *H. pylori* are more frequently isolated from patients presenting with periodontitis (infection of the gums) and it is suggested that the periodontal pocketing and inflammation caused by the disease harbours *H. pylori* (Gebara et al. 2004; Souto and Colombo 2008; Umeda et al. 2003). The presence of *H. pylori* in the oral cavity is associated with presence of bacterium in the stomach and is suggested to be involved with re-infection as oral *H. pylori* are not effected by antibiotic treatment (Asqah et al. 2009; Desai et al. 1991; Liu et al. 2009). Although many studies have described the presence of *H. pylori* within the oral cavity and suggested it as a permanent reservoir for *H. pylori*, other researchers have questioned this idea. Silva Rossi-Aguiar et al. (2009) were unable to identify *H. pylori* in the oral cavity of patients who had *H. pylori* present within the stomach and suggested that *H. pylori* transitions through the oral cavity as opposed to residing within it. Olivier et al. (2006), have also proposed that the oral cavity is not a permanent reservoir for *H. pylori* as they were unable to detect the pathogen in dental samples. The discrepancy observed between the various studies may be due to the various methods of identification used and the various demographics studied.

Other potential natural reservoirs of the bacterium include sheep and natural water sources. Through both standard culturing and molecular identification methods, *H. pylori* has been isolated and identified from sheep stomach and sheep milk (Dore et al. 2001; Momtaz et al. 2014). It has been hypothesised that sheep are the ancestral host of *H. pylori* and are responsible for the original transmission to humans (Dore et al. 2001; Momtaz et al. 2014). In water sources *H. pylori* DNA has been identified in municipal water, wells, river and pond water (Hultén et al. 1996; Sasaki et al. 1999). Biofilms may also play a role in the spread and survival of this pathogenic bacterium in water sources. Biofilms comprise of matrix-enclosed bacterial populations that are able to adhere to other bacteria, surfaces and interfaces (Costerton et al. 1995). Cole et al. (2004) have demonstrated that *H. pylori* possess a number of genes involved in biofilm production that allows them to adhere to abiotic surfaces. *H. pylori* biofilms have been identified in water distribution systems even after washing (Mackay et al. 1998; Park et al. 2001; Watson et al. 2004). The presence of *H. pylori* within biofilms allows the bacterium to survive in a stable and potentially microaerophilic environment (Gião et al. 2008). Although these studies have demonstrated the presence of *H. pylori* in water samples, it is unknown whether water is a reservoir or transmission route of the pathogen. This is likely due to the fact that *H. pylori* are believed to enter a VBNC state in the natural environment, preventing them from multiplying but allowing them to survive until they enter a suitable host (Adams et al. 2003; Azevedo et al. 2007). There are two main studies supporting the idea of water as a reservoir for *H. pylori*. Goodman et al. (1996) have demonstrated that Columbian children who swam in rivers, streams,

and swimming pools were at an increased risk of developing an *H. pylori* infection. Whereas in Peru, municipal drinking water was identified as the source for an increased risk of obtaining *H. pylori* (Klein et al. 1991). These other potential sources all require additional research to complete the understanding of their role in natural reservoirs of *H. pylori*.

1.3.2.3 Transmission

A route of transmission for *H. pylori* remains undefined. There are multiple routes of transmission that have been proposed, with person-to-person and water-borne transmission being important factors.

Person-person transmission can be divided into two main categories: vertical transmission (within the same family) or horizontal transmission (outside of the same family). Vertical transmission or intrafamilial transmission has been suggested as one of the main routes of *H. pylori* infection due to the identification of the same or similar strains of *H. pylori* present within the same families (Bamford et al. 1993; Kivi et al. 2003; Raymond et al. 2004). Sibling-to-sibling transmission has been observed in Colombian children from older to younger siblings that are close in age (Goodman and Correa 2000). Mother-child transmission is also recognised as a major transmission route for the spread of *H. pylori,* as this pathogen is frequently acquired during childhood, usually within the first five years of life (Mitchell et al. 1992; Rothenbacher et al. 1999). Due to the infection of children in the first five years of life, where the mother usually has the most contact, the spread of infection may occur due to *H. pylori* being present in the mouth of mothers and through the shared contact of spoons, pacifiers, feeding bottles and food with the child, they become infected (Escobar and Kawakami 2004; Rothenbacher et al. 1999). Horizontal transmission, transmission outside of the main family, may be due to an increased amount of contact between children and non-parental caretakers (Schwarz et al. 2008). Person-person transmission routes also include gastro-oral through vomitus, where *H. pylori* has been isolated from the fresh vomitus of children (Leung et al. 1999); oral-oral through ethno-specific food practices such as chopstick charing (Chow et al. 1995); and through faecal-oral routes, as *H. pylori* have been isolated from human faeces (Thomas et al. 1992). Person-to-person transmission, primarily within families, is an important factor contributing to the spread of *H. pylori* infections. However, this route is still not completely understood as a major transmission route is still yet to be defined.

In lower socioeconomic areas where clean water is not readily available, waterborne transmission of *H. pylori* is important to consider. As mentioned above (Section 0), *H. pylori* DNA has been isolated from a variety of water sources. Many studies indicate that access to clean water and proper wastewater sanitation may reduce the risk of acquiring *H. pylori* (Baker and

Hegarty 2001; Horiuchi et al. 2001; Hultén et al. 1996; Nurgalieva et al. 2002; Travis et al. 2009). Although these studies have identified *H. pylori* DNA within water samples, it is unknown whether it comes from viable cells, VBNC cells or dead cells. Studies have demonstrated that *H. pylori* is able to convert into a coccoid state when exposed to a natural environment (Adams et al. 2003; Azevedo et al. 2007; Queralt and Araujo 2007). However, there is conflicting evidence on whether these coccoid forms of the bacteria are in a VBNC state or the manifestation of cell death and whether they are able to cause infection from transmission via water sources (Boehnke et al. 2017; Cellini et al. 1994; Queralt and Araujo 2007; Sen et al. 2011; She et al. 2003). Overall, there is still a lack of data to confirm that viable *H. pylori* are present within water, and that they are able to cause infection if ingested.

It has also been suggested that *H. pylori* may be a foodborne pathogen as it has been identified in a variety of foodstuffs. Primarily using molecular-based methods, *H. pylori* have been identified in ready-to-eat meals, minced-meat, hamburger meat, vegetables, salads and various dairy products (Atapoor et al. 2014; Berroteran et al. 2002; Fujimura et al. 2002; Gilani et al. 2017; Hemmatinezhad et al. 2016; Saeidi and Sheikhshahrokh 2016; Yahaghi et al. 2014). The prevalence of *H. pylori* in food is likely associated with a faecal-oral transmission route through poor hygiene conditions of food handlers or through contamination of the food itself through the utilisation of faeces as fertiliser (Begue et al. 1998; Hopkins et al. 1993). However, as noted with waterborne transmission, there is a lack of direct evidence for the presence of *H. pylori* in foods and the viability of these cells to cause infection in a human host.

A prominent factor discussed with transmission routes is the socioeconomic factor of the patients. Low socioeconomic status in both childhood and adult life is associated with an increased risk of acquiring *H. pylori* and developing an infection (Malaty and Graham 1994; Ozaydin et al. 2013). Lower age range males have also shown an increased prevalence of *H. pylori* infection (de Martel and Parsonnet 2006; Graham et al. 1991; Rowland et al. 2006). These factors are important to consider when identifying and discussing the transmission of this pathogenic bacterium.

1.3.2.4 Vectors

Houseflies (*Musca domestica*) have been described as a potential vector of *H. pylori*. Grübel et al. (1997) show that viable H. *pylori* may be found on the external surfaces of the housefly for up to 12 hours, after initial exposure, and found in the gut of the housefly up to 30 hours after initial digestion of pure cultures of *H. pylori*. However, Osato et al. (1998), describe that houseflies are unable to obtain *H. pylori* from human faecal samples, therefore making them an unlikely vector of this pathogen. As well as houseflies, cockroaches (*Periplaneta fuliginosa*) have

been suggested as a possible vector for *H. pylori*, once again due to their unsanitary living environments and potential to contaminate food for human consumption (Imamura et al. 2003).

1.3.3 Infection

Worldwide infection rates of *H. pylori* indicate that, in 2015, more than half of the world's population were positive for carrying *H. pylori* (Hooi et al. 2017). Although many individuals are carriers of *H. pylori*, most will not develop symptoms. Those that do develop symptoms related to *H. pylori* infection will have varying clinical outcomes (Figure 1.4) which are dependent on both host and bacterial factors (Suerbaum and Michetti 2002). Host factors associated with the clinical outcome of infection include polymorphisms in genes encoding pro-inflammatory cytokines, genes involved in initiation and upregulation of immune response and the secretor status of the host (Dunne et al. 2014). Bacterial factors include antibiotic resistance and virulence factors (Uotani et al. 2015).

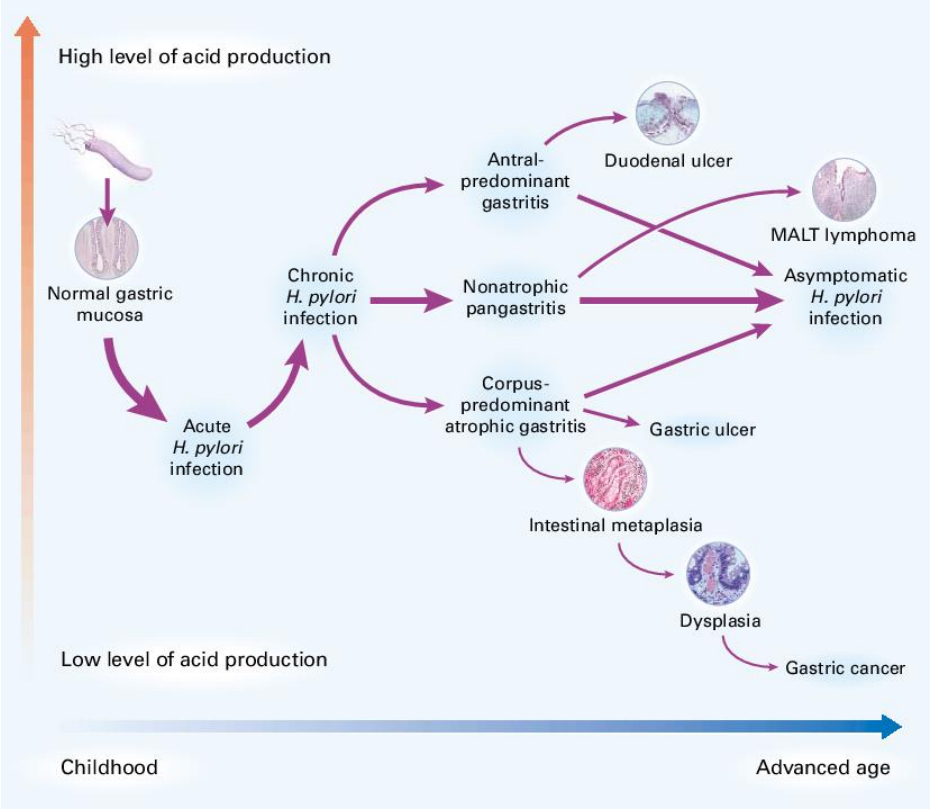

Figure 1.4: Development of an *H. pylori* infection. Image retrieved from (Suerbaum and Michetti 2002).

1.3.3.1 Clinical Outcomes

Clinical outcomes of an *H. pylori* infection include the development of gastritis, ulcer diseases, and in severe cases, gastric carcinoma and mucosa-associated lymphoid tissue lymphoma (MALToma) (Nomura et al. 1994; Parsonnet et al. 1991; Veldhuyzen van Zanten and Sherman 1994). Many individuals who carry *H. pylori* will remain symptomless throughout their lifetime

(Mungazi et al. 2018). However, 10 to 20 % will develop gastritis and/or peptic ulcer disease and 1 to 2 % will develop gastric cancers (Kusters et al. 2006). Infection of normal gastric mucosa can lead to gastritis (inflammation of the stomach lining), which can be either acute or chronic (Dixon et al. 1996). *H. pylori* associated acute gastritis is characterised as an asymptomatic or mild self-limited dyspepsia illness which is usually not recognised by patients (Dixon et al. 1996; Egan et al. 2007; Sobala et al. 1991). Histologically, acute gastritis is characterised by neutrophilic leukocytes infiltrating the stomach (Feldman and Lee 2016). Untreated acute gastritis will almost always develop into active chronic gastritis (Figure 1.5 B), however, it may also spontaneously resolve (Jensen et al. 2019). *H. pylori* is recognised as the most common cause of active chronic gastritis, a progressive, multistep, lifelong inflammation (Ihamauki et al. 1985; Kuipers et al. 1995; Morgan and Crowe 2016; Valle et al. 1996). Active chronic gastritis is histologically characterised by the presence of mononuclear cells (mostly lymphocytes), plasma cells and macrophages (Kuipers et al. 1995; Watari et al. 2014). There are three main types of chronic gastritis identified: antral-predominant gastritis, nonatrophic pangastritis and corpus-predominant atrophic gastritis (Figure 1.4) (Suerbaum and Michetti 2002; Watari et al. 2014). Clinical sequelae can lead active chronic gastritis into peptic ulcers; either duodenal ulcers (Figure 1.5 C) or gastric ulcers (Figure 1.5 D); mucosal atrophy; and in severe cases gastric carcinoma (Figure 1.5 E) and gastric MALToma (Figure 1.5 F) (Jensen et al. 2019). Peptic ulcers may also be caused by non-steroidal anti-inflammatory drugs (NSAIDs) (Drini 2017). Use of these drugs are the major cause of infection-free peptic ulcers (Vergara et al. 2005). Interaction between NSAIDs users and *H. pylori* infections is unclear, with data suggesting *H. pylori* may increase, decrease or have no effect on ulcer risk in NSAID uses (Chan et al. 2002; Huang et al. 2002; Labenz et al. 2002; Papatheodoridis et al. 2006).

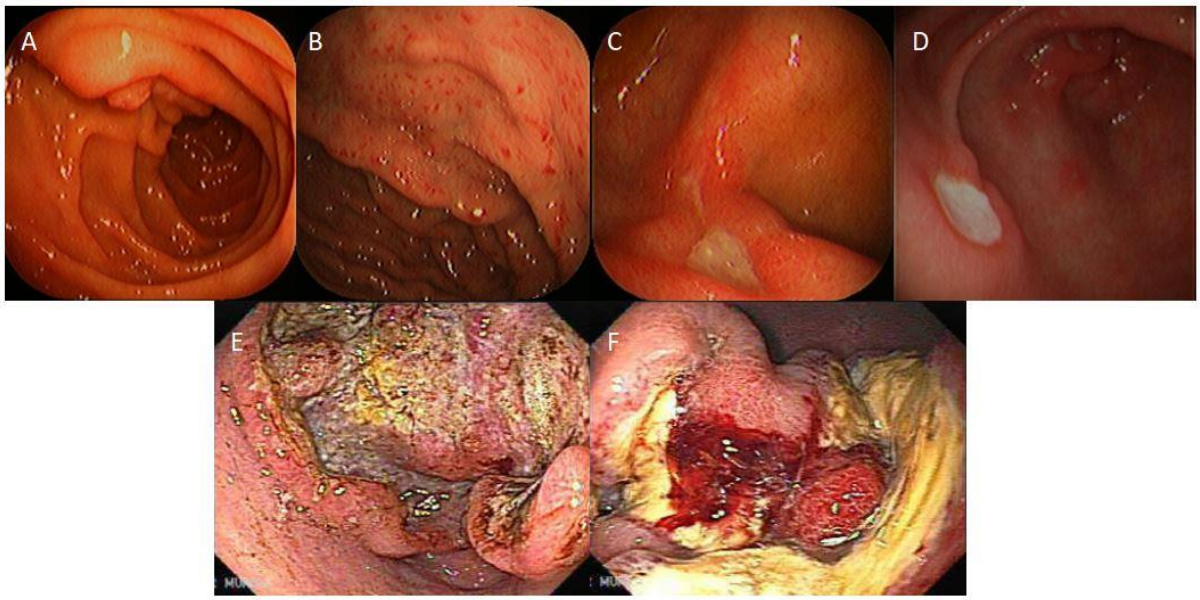

Figure 1.5: Endoscopic images of healthy and infected stomach linings. (A) Normal, healthy duodenum; (B) gastritis infection; (C) duodenum ulcer; (D) stomach ulcer; (E) gastric adenocarcinoma and (F) gastric MALToma all associated with an *H. pylori* infection. Images A-D retrieved from https://www.laendo.net/endoscopic-images/stomach/350-gastric-ulcers-related-to-hpylori, image E retrieved from https://www.gastrointestinalatlas.com/english/gastric_cancer.html and image F retrieved from https://www.gastrointestinalatlas.com/english/gastric_lymphoma.html.

1.3.3.2 Virulence

There are many contributing components towards the virulence of *H. pylori*. The virulence of this pathogen leading to colonization in the host and infection can be viewed as a four-step process involving multiple virulence factors, summarised in Figure 1.6.

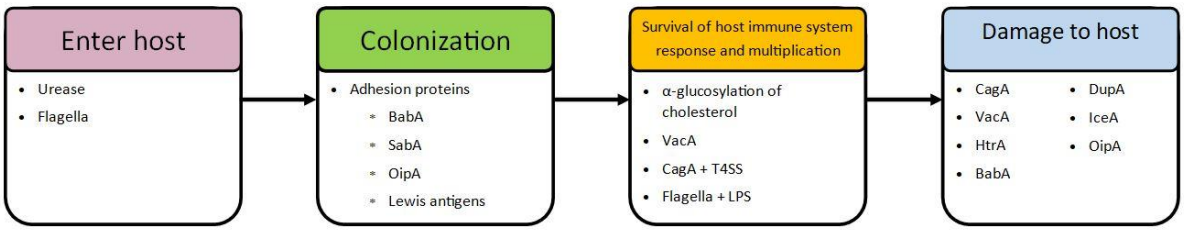

Figure 1.6: Flow chart of major virulence factors involved in an *H. pylori* infection. Image adapted from (Chang et al. 2018). T4SS = type IV secretion system, LPS = lipopolysaccharide.

These virulence factors can be broken down into four main categories: entry into host, colonization, immune escape and disease induction (Chang et al. 2018). All isolates express the common virulence factors urease, flagella and adhesion proteins to aid in colonization and survival within the human stomach (Clyne et al. 1995). Urease produced by the bacterium hydrolyses urea present in the stomach to form ammonia and carbon dioxide, which act as a buffer against the harmful gastric acid (Clyne et al. 1995). By neutralizing the pH of the gastric environment, the viscoelasticity in the mucin gel is reduced and the bacterium is able to use its

flagella to move freely into the mucus layer of the stomach and attach to the epithelial cells for colonisation (Celli et al. 2009). Once established inside the epithelial cells, adhesion ensues through different adhesion proteins, as without adhesion to gastric epithelial cells *H. pylori* will be washed away through clearance systems from the host (Magalhães and Reis 2010). *H. pylori* adhesion proteins have been identified as outer membrane proteins (OMPs) and include: Bab A (blood group antigen-binding adhesin, *babA*), that attaches to Lewis b histo-blood group that are present in the gastric mucosa (Boren et al. 1993; Ilver et al. 1998); SabA (sialic acid-binding adhesin, *sabA*) which bind to associated sialylated/fucosylated glycans (e.g. sialylated Lewis antigens, Le^x) which are crucial for *H. pylori* to develop chronic infection in the host (Aspholm et al. 2006; Mahdavi et al. 2002) and; OipA (outer inflammatory protein, *oipA*), which promote adherence to gastric epithelia but to what specific ligands is still to be determined (Dossumbekova et al. 2006; Yamaoka et al. 2000). These adhesion factors are strain specific and usually occur with other virulence factors to produce chronic infection in the host (Hennig et al. 2006; Ilver et al. 1998; Yamaoka et al. 2000). Adhesion can allow for the efficient delivery of other bacterial virulence factors (Gerhard et al. 1999).

Cells of *H. pylori* must then overcome the host's harsh immune system. The host's immune response to *H. pylori* includes the development of specific antibodies such as immunoglobulin g and a (IgG and IgA) that target *H. pylori* antigens (e.g. membrane proteins, urease, flagella and lipopolysaccharides) (Mattsson et al. 1998b; Mattsson et al. 1998c). As *H. pylori* are known for their life-long chronic infections, host immune responses are either not suitable to eradicate the bacteria or this pathogen is able to overcome the response in a variety of ways.

To avoid phagocytosis by the host, *H. pylori* use α-glucosylation of cholesterol, from host epithelial cells, to avoid T cell responses (Wunder et al. 2006). This pathogen is also able to modify T cell responses by altering the hosts $CD4^+$ T cell response to an *H. pylori* infection (Lundgren et al. 2003).

The vacuolating toxin protein (VacA), encoded by *vacA*, is an important part of *H. pylori*'s resistance against the host's immune responses. This cytotoxin is able to down-regulate the immune response by inducing vacuolation in the epithelial cells leading to apoptosis and impeding the signalling pathway for T cell activation (Gebert et al. 2003; Yahiro et al. 2015).

The cytotoxin-associated protein (CagA), encoded by *cagA*, is another important component of *H. pylori*'s ability to maintain a chronic infection. This protein is able to translocate into gastric epithelial cells through a type IV secretion system (T4SS) (Odenbreit et al. 2000). This T4SS is encoded by a 40-kilo-base pair (kbp) pathogenicity island (PAI), found within certain strains of *H. pylori* (Censini et al. 1996). Once CagA is in the epithelial cells it triggers cell scattering, actin-

cytoskeletal rearrangements and production of interleukin 8 (IL8) that contributes to inflammation (Brandt et al. 2005). CagA also contributes to the blocking of T cells, further aiding in this pathogens ability for chronic infection (Paziak-Domańska et al. 2000).

From molecular markers within the *H. pylori* genome, clinical isolates can be divided into two broad groups, type I and type II based on the presence or absence of *cagA* and *vacA*. Type I isolates possess the *cagA* gene to code for CagA and VacA. Type II isolates lack the *cagA* gene and therefore do not produce CagA or VacA (Xiang et al. 1995). The *cag*PAI, as mentioned above, encodes for a T4SS aiding in evading host immune responses; it is also involved in the mutagenesis of gastric epithelial cells contributing to gastric adenocarcinoma and MALToma (Sepulveda et al. 2010; Umehara et al. 2003).

Damage to the host is facilitated by a number of virulence factors. Major virulence factors include CagA, VacA, high temperature requirement gene A (*htrA*), duodenal ulcer promoting gene A (*dupA*), induced by contact with epithelium gene A (*iceA*) and OipA. Within the context of damaging the host, cagA is an oncoprotein that is able to produce malignant neoplasms in mammals through disrupting host signalling pathways and leading to gastric adenocarcinoma and MALToma (Hatakeyama 2014; Shmuely et al. 2001; Umehara et al. 2003). The VacA protein is also associated with development of gastric cancer; variation between strains produce different toxin activity as not all *H. pylori* strains cause gastric cancers (Lee et al. 2008; McClain et al. 2017). Allelic varieties containing s1, m1 or i1 *vacA* alleles have been associated with a higher risk of gastric adenocarcinoma when compared to those with the s2, m2 or i2 allelic varieties (Miehlke et al. 2000; Rhead et al. 2007). HtrA, encoded by *htrA*, is a protease secreted by *H. pylori* that is important for growth under stressful conditions and proteolytic activity against E-cadherins allowing the bacterium entry to intracellular space, and is also associated with development of gastric adenocarcinoma (Hoy et al. 2010; Schmidt et al. 2016; Zarzecka et al. 2019). DupA, encoded by *dupA*, is associated with the development of duodenal ulcers and is a marker for the increased risk of duodenal ulcer, but reduced risk of gastric cancer (Lu et al. 2005; Wang et al. 2015b). Peptic ulcers, specifically duodenal ulcers, have also been associated with *iceA* presence, specifically the allelic variant *iceA1*, but the specific mechanism is still unclear (Huang et al. 2016b; Peek et al. 1998; Shiota et al. 2012). As well as adhering *H. pylori* to host epithelial cells, OipA can also induce apoptotic cascade in host cells and is associated with more severe clinical outcomes such as gastric cancer (Braga et al. 2019; Teymournejad et al. 2017). This range of virulence factors make this pathogen a successful inhabitant of human hosts and with extreme clinical outcomes such as gastric cancer, make it important to control.

1.3.4 Treatment

1.3.4.1 Diagnosis

Treatment of *H. pylori* infection first involves positive identification of the bacterium. Diagnostic tests include: enzyme-linked immunosorbent assay (ELISA), salivary antibody test, urea breath test, biopsy, histology, and culture (Howden and Hunt 1998). Within NZ, patients presenting with dyspepsia or peptic ulcers undergo faecal antigen testing as the primary way to test for the presence of *H. pylori* (New Zealand Guidelines Group 2004). Oesophago-gastro-duodenoscopy may be used as an invasive method of testing and collecting biopsy material for further testing as well as monitoring treatment progress (New Zealand Guidelines Group 2004).

1.3.4.2 Antibiotic Characteristics and Cellular Targets

The original definition of an antibiotic was "a substance that is produced by one microorganism that inhibits the growth of another microorganism" (Russell 2004, p152). Through advancements in synthetic techniques, the definition of an antibiotic has been updated to "a substance or similar substance produced by a microorganism or wholly or partly by chemical synthesis, which in low concentrations can inhibit the growth of other microorganisms" (Russell 2004, p152). There are three main sources of antibiotics: microorganisms, chemical synthesis and semi-synthesis (Russell 2004).

Treatment of *H. pylori* infections generally involve a PPI and two to three antibiotics that are taken for one to two weeks (Crowe 2019). The most common PPI and antibiotics used to treat an *H. pylori* infection are discussed below. The chemical structures of the most common antibiotics are presented in Figure 1.7.

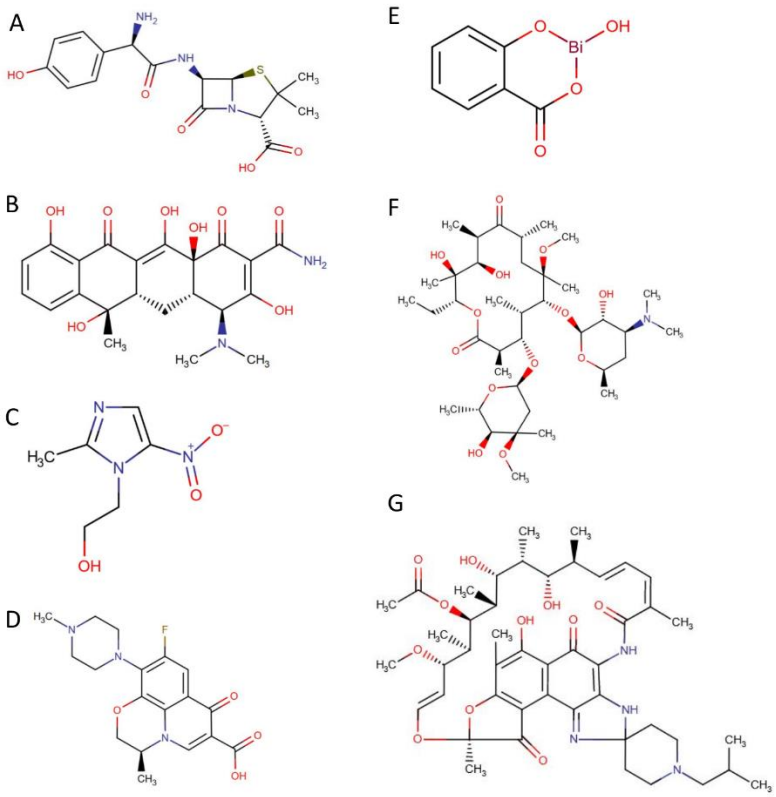

Figure 1.7: Chemical structures of common antibiotics involved in treatment of *H. pylori*: Amoxicillin (A), tetracycline (B), metronidazole (C), levofloxacin (D), bismuth subsalicylate (E), clarithromycin (F), and rifabutin (G). Images retrieved from https://www.drugbank.ca/.

Proton Pump Inhibitors

PPIs are widely used to treat gastrointestinal diseases and are therefore an important part of the treatment of an *H. pylori* infection. PPIs are used in gastric associated infections and diseases as they increase the pH levels in the stomach (Huang and Hunt 1996). Low pH levels can cause *H. pylori* to enter their coccoid form, making them resistant to antibiotics (Axon 1994). Antibiotics used in treatment rely on the bacteria to be in a viable, growing state for effective eradication (Marcus et al. 2012). The PPIs used therefore allow *H. pylori* to maintain a viable, growing state and the antibiotics are able to eradicate the actively growing cells (Huang and Hunt 1996). Common PPIs used in treatments include omeprazole, lansoprazole, pantoprazole, rabeprazole and esomeprazole (New Zealand Formulary 2020b; Vergara et al. 2003).

Clarithromycin

Clarithromycin falls under the macrolide class of antibiotics and is a semi-synthetic derivative of erythromycin (Guay et al. 2001). Its chemical structure includes a 14-memebered lactone ring with a methylated hydroxyl group in position 6, as seen in Figure 1.7 F. The methoxy group in position 6 increases the bioavailability of the antibiotic through decreased decomposition in acidic conditions such as the gastric environment (Kirst and Sides 1989). The mechanism of action of clarithromycin is the same as other macrolides; it involves the reversible binding to

domain V of the 23S rRNA component of the 50S ribosomal subunit (Kanoh and Rubin 2010). This binding causes the early release of peptidyl-tRNA from the ribosome, preventing elongation and inhibiting protein synthesis (Brisson-Noël et al. 1988).

Metronidazole

Metronidazole belongs to the nitroimidazole class of antibiotics and has a chemical structure of 1-β-hydroxyethyl-2-methyl-5-nitromidazole (Figure 1.7 C). It is a broad spectrum antibiotic, used to treat anaerobic or microaerophilic, gram-positive or gram negative bacteria and protozoa (Edwards 1980). The mechanism of action of this pro-drug involves a four-step process. Firstly, the antibiotic enters the cell through diffusion across the cell membrane (Ings et al. 1974). Secondly, pyruvate-ferredoxin oxidoreductase reductively activates the nitro-group of metronidazole, causing a change in the drug's chemical structure. The reduction of metronidazole causes a concentration gradient to form that increases the uptake of the drug into the cell (Edwards 1986). Thirdly, the reduced intermediate particles exert a cytotoxic effect of the cell by interacting with the cell's DNA leading to DNA strand breakage, unwinding and helix destabilisation, all leading to the cell's death (Tocher and Edwards 1992). Lastly, the cytotoxic intermediate products are broken down to their inactive end products (Goldman et al. 1986). Specifically within *H. pylori* cells, metronidazole reacts with two nicotine adenine dinucleotide phosphate (NADPH) nitroreductases, oxygen-insensitive NAPDH nitroreductase (RdxA) and NADPH flavinoxidoreductase (FrxA), as well as pyruvate-ferredoxin oxidoreductase (PFOR), to catalyse the transformation of metronidazole into a bactericidal agent (Sisson et al. 2002).

Amoxicillin

Amoxicillin belongs to the aminopenicillins subcategory of the penicillin drug class. It is a broad spectrum, semisynthetic, bactericidal antibiotic commonly used to treat chest, ear, throat, sinus and dental infections, as well as *H. pylori* infections (Mutters and Meyle 2013; New Zealand Formulary 2020a). Amoxicillin is extremely similar to ampicillin, the other antibiotic within the aminopenicillins subcategory and the pioneer aminopenicillin. The difference between the two is the addition of a hydroxyl group to amoxicillin, thereby changing the physiochemical properties of the molecule and its pharmacokinetics allowing this drug to be administered orally (Figure 1.7 A) (Mutters and Meyle 2013).

The mode of action for amoxicillin involves a similar mechanism observed in other penicillin's and β-lactam antibiotics. Bactericidal activity is exerted onto the bacterial cell through the inhibition of cell wall synthesis (Cho et al. 2014). Amoxicillin binds to the penicillin-binding proteins (PBP) that cross-link two linear peptidoglycan chains, the last step of cell wall synthesis.

When binding to PBPs, the β-lactam ring of the antibiotic is opened and irreversibly acetylates the C-terminal transpeptidase-domain, inactivating the enzyme and causing cell lysis (Cho et al. 2014; Mutters and Meyle 2013; Neu 1982).

Bismuth

Bismuth compounds have been used to treat gastrointestinal diseases for many years. Bismuth is a natural compound, and bismuth salts are used in the treatment of *H. pylori* and other gastrointestinal diseases (Alkim et al. 2017; Gorbach 1990). The specific compounds used in the treatment of *H. pylori* are bismuth subsalicylate, colloidal bismuth subcitrate and ranitidine bismuth citrate (Lambert and Midolo 1997). The chemical structure of bismuth subsalicylate is shown in Figure 1.7 E and is one of the most common bismuth compounds used to treat *H. pylori* (Bland et al. 2004). The mechanism of action for bismuth salts on *H. pylori* is still not greatly understood but they include the disruption of the glycocalyx–cell wall that surrounds the cell, and inhibition of urease activity, thereby preventing the continued colonisation of *H. pylori* (Marcus et al. 2015; Stratton et al. 1999; Zhang et al. 2006).

Tetracycline

Tetracycline is part of the tetracyclines class of antibiotics. They are broad range antibiotics as they exhibit activity against a wide range of both Gram positive and Gram negative bacteria, as well as protozoan parasites (Chopra and Roberts 2001). The structure of tetracyclines includes a linear fused tetracycline nucleus where different functional groups may attach (Chopra and Roberts 2001). Commercially, tetracycline is marketed as achromycin which has a chemical structure of $C_{22}H_{24}N_2O_8$, as seen in Figure 1.7 B (Chopra and Roberts 2001).The mode of action for tetracyclines is the inhibition of protein synthesis by blocking the elongation stage (Chopra and Roberts 2001). Tetracycline blocks the attachment of aminoacyl-tRNA to the ribosomal acceptor (A) site, thereby inhibiting the addition of amino acids to the developing polypeptide (Suarez and Nathans 1965).

Levofloxacin

Levofloxacin belongs within the fluroquinolone class of antibiotics as a third-generation fluroquinolone. It is another broad-spectrum antibiotic exerting its effect on both gram-positive and gram-negative bacteria (Anderson and Perry 2008). Structurally, levofloxacin is the optically active L-isomer of ofloxacin (Figure 1.7 D) (Fish and Chow 1997). Levofloxacin's mode of action involves the interaction with, and inhibition of, type II topoisomerases such as DNA gyrase and topoisomerase IV. These two topoisomerases have important roles in bacterial DNA replication. DNA gyrase, encoded by the *gyrA* and *gyrB* genes, supports the beginning of chain elongation by reducing the strain on the double-stranded DNA as it is being unwound; topoisomerase IV is

involved in the final stages of DNA replication where it separates the topologically linked chromosomes (Levine et al. 1998). Levofloxacin binds to these two important enzymes in a way that forms drug/enzyme/DNA complexes disrupting normal bacterial DNA replication and resulting in bacterial death (Drlica 1999; Marians and Hiasa 1997). Although both enzymes are targets for levofloxacin, DNA gyrase is the primary target for gram-negative bacteria and topoisomerase IV is the primary target of gram-positive bacteria (Hooper 1999).

Rifabutin

Rifabutin is a broad-spectrum antibiotic belonging to the rifamycin class of antibiotics (Della Bruna et al. 1983). It is a derivative of rifamycin and has a molecular formula of $C_{46}H_{62}N_4O_{11}$ which is structurally similar to rifamycin (Figure 1.7 G). It is primarily used to treat *Mycobacterium avium* complex (MAC) in patients with HIV infections (Nightingale et al. 1993). However, it has also been used as a second-line or salvage treatment for *H. pylori* infections (Gisbert et al. 2003; Perri et al. 2000; Van Der Poorten and Katelaris 2007). The target of this antibiotic is the DNA-dependent RNA polymerase. Rifabutin inhibits this polymerase at the β-subunit preventing transcription initiation leading to inhibition of protein synthesis (Kunin 1996).

1.3.4.3 Treatment Regimens

There are many available treatments for infection with *H. pylori*, and these treatments are continuously evolving. The correct treatment regimen for a patient must take into account factors such as: the local antibiotic resistance patterns (if available), previous exposure to antibiotics, allergies, side effects, cost and ease of administration (Crowe 2019). Due to these varying factors there are many different therapies available worldwide, the most common are summarised in Table 1.2.

Table 1.2: Summary of treatment regimens for *H. pylori* infection.

Therapy	Components	Dosing frequency	Duration	Reference
Triple therapies	PPI[1] + amoxicillin (1 g) + clarithromycin (500 mg)	BID	7 – 14 days	(Chey and Wong 2007)
	PPI[1] + amoxicillin (1 g) + metronidazole (500 mg)	BID	7 – 14 days	(McColl 2010)
	PPI[1] + metronidazole (500 mg) + clarithromycin (500 mg)	BID	7 days	(Gisbert et al. 2005)
	PPI[1] + levofloxacin (500 mg) + amoxicillin (1 g)	BID	10 days	(Gisbert et al. 2006)
	PPI[1] + rifabutin (150 mg) + amoxicillin (1 g)	BID	10 days	(Van Der Poorten and Katelaris 2007)
Bismuth quadruple therapy	PPI[1] + bismuth (120 – 300 mg) + nitroimidazole[2] (500 mg) + tetracycline (500 mg)	BID (PPI) and QID	10 – 14 days	(McColl 2010)
Concomitant therapy	PPI[1] + clarithromycin (500 mg) + amoxicillin (1 g) + nitroimidazole[2] (500 mg)	BID	10 – 14 days	(Gisbert and Calvet 2012)
Sequential therapy	PPI[1] + amoxicillin (1 g), followed by PPI[1] + clarithromycin (500 mg) + nitroimidazole[2] (500 mg)	BID	5 – 7 days 5 – 7 days	(McColl 2010)
Hybrid therapy	PPI[1] + amoxicillin (1 g), followed by PPI + amoxicillin (1 g) + clarithromycin (500 mg) + metronidazole (500 mg)	BID BID	7 days 7 days	(Hsu et al. 2011)
High-dose dual therapy	PPI[1] + amoxicillin (1 g or 750 mg)	TID or QID	14 days	(Chey and Wong 2007)

[1] Dosage of PPI depends on type used, [2] metronidazole or tinidazole, BID = twice daily, TID = three times daily, QID = four times daily.

1.3.4.4 First-line Treatments

Treatment of *H. pylori* begins with first-line treatment; the regimen that has the best chance of eradication. This first-line treatment is usually a triple therapy that consists of a PPI and two antibiotics (Chuah et al. 2017). The particular antibiotics used depend on the antibiotic resistance rates of the surrounding area and the patient's previous exposure to antibiotics and penicillin allergies (Crowe 2019). The standard triple therapy regimen is PPI + amoxicillin + clarithromycin, however, clarithromycin resistance rates are growing rapidly and this treatment should not be used when the area's clarithromycin resistance is more than 15 % (Chey and Wong 2007; Malfertheiner et al. 2017; Savoldi et al. 2018). With previous exposure to macrolides the alternative triple therapy is PPI + amoxicillin + metronidazole (Chey and Wong 2007). In the instance of penicillin resistance, the alternative triple therapy is PPI + clarithromycin + metronidazole.

Because of increasing resistance and previous exposure to macrolides other regimens have been proposed. Bismuth quadruple therapy has been recommended as an alternative to increasing clarithromycin resistance (Randel 2018; Tursi et al. 2017). This therapy consists of PPI + bismuth + tetracycline + nitroimidazole (either metronidazole or tinidazole) taken for ten days (Tursi et al. 2017; Wang et al. 2017). Eradication rates of bismuth quadruple therapy vs clarithromycin triple therapy show 77.6 % and 68.9 % respectively in randomized clinical trial (RCT) patients, and bismuth quadruple therapy was not affected by metronidazole resistance (Venerito et al. 2013).

Concomitant and sequential therapies are also alternative first-line treatments. Sequential therapy involves taking a PPI + amoxicillin for five to seven days, followed by a PPI + clarithromycin and a nitroimidazole for another five to seven days (Zullo et al. 2013). Concomitant therapy contains the same antibiotics as sequential therapy but they are all taken at once for ten to fourteen days (McColl 2010; Zullo et al. 2013). Concomitant therapy is also an alternative quadruple therapy that does not contain bismuth, as bismuth-salts are not available in all countries, including NZ and Australia (Löscher and Alberer 2013; McColl 2010). Sequential and concomitant therapy have equal rates of eradication of *H. pylori*, with Korean and Taiwanese patients having successful eradication at 92.3 % (sequential) and 93 % (concomitant); and 75.6 % (sequential) and 80.8 % (concomitant) (Lim et al. 2013; Wu et al. 2010). The eradication rates of sequential and concomitant therapy make it a suitable option as a first-line treatment therapy in areas of high clarithromycin resistance (Lim et al. 2013).

The hybrid therapy is a combination of sequential and concomitant therapies; it involves a standard dose of a PPI (esomeprazole suggested) and 1 g amoxicillin, taken twice daily for seven

days then for another seven days a standard dose of a PPI, 1 g amoxicillin, 500 g clarithromycin and 500 g metronidazole are taken twice daily (Hsu et al. 2011). Eradication rates of hybrid therapies in Korea show an acceptable level of 85.9 % but also show no significant difference between sequential and concomitant therapy as a first-line treatment therapy (He et al. 2015; Oh et al. 2014; Song and Zhou 2016).

As well as using different combinations of antibiotics and acid suppressors, the duration of treatment is an important factor for sufficient eradication rates of first-line treatment therapies. Triple therapy regimens containing a PPI + amoxicillin and clarithromycin have shown higher eradication rates when taken for 14 days compared to seven days (Arama et al. 2016; New Zealand Formulary 2020b).

1.3.4.5 Second-line Treatments

Failure of first-line treatments is usually due to resistance to one or more of the antibiotics used in initial treatment, it is therefore recommended to not include antibiotics the patient has had previous exposure to and to run an antibiotic susceptibility test on the patient's specific *H. pylori* strain (Graham and Fischbach 2010). To treat an *H. pylori* infection, there should usually be two treatment options that have an acceptable eradication rate in accordance with the local antibiotic resistance rates. The treatment with the highest eradication rate should be used first and if fails, the second best option should be used (Graham and Calvet 2012). Second-line therapies include an alternative triple therapy or bismuth quadruple therapy than that used for first-line treatment (Gomollón et al. 1999; McColl 2010).

1.3.4.6 Salvage Treatments

When two treatment regimens have failed, a salvage regimen is used as a last effort to eradicate the infection. Fluroquinolones are not readily used in first-line treatments and thus have been suggested as possible salvage treatments for failed eradication of *H. pylori* using the treatments described above (Randel 2018). Levofloxacin, triple therapy includes a PPI + amoxicillin (1 g) + levofloxacin (500 mg) administered twice a day for ten days and has an eradication rate of 66 % when used as a second-line or salvage therapy (Gisbert et al. 2006). Comparing the typical bismuth quadruple therapy and 10-day levofloxacin-based triple therapy as a salvage therapy, showed that the levofloxacin triple therapy had less adverse side effects and was more effective at treating *H. pylori* than the bismuth quadruple therapy (Saad et al. 2006). However, fluroquinolone resistance has been steadily increasing globally (Gemilyan et al. 2019; Glocker et al. 2007; Kim et al. 2005; Mégraud et al. 2013). This increasing resistance is of concern with continued use of treatments containing this class of antibiotics.

Clarithromycin-based concomitant therapy may be used as a salvage therapy but only when patients have no previous macrolide exposure and no risk for macrolide resistance, and where local clarithromycin resistance rates are below 15 % (Chey and Wong 2007; Crowe 2019). A high-dose dual therapy is a suitable option for salvage therapy, especially in countries where bismuth is not available. This regimen involves a PPI + amoxicillin (1 g or 750 mg); the PPI is administered three times daily and amoxicillin is taken four times daily at a 750 mg dose and three times daily at 1 g for 14 days (Chey and Wong 2007; Yang et al. 2019). This regimen is an important salvage option in patients or areas of proven or suspected high resistance to clarithromycin, macrolides and or fluroquinolones (Crowe 2019). Eradication rates of this therapy as a salvage treatment is 89.3 % for both intention-to-treat (ITT) and per protocol (PP) analysis (Yang et al. 2015).

The final salvage therapy is the rifabutin-based therapy, commonly a modified triple therapy involving rifabutin. This salvage therapy consists of a PPI + amoxicillin (1 g) + rifabutin (150 mg), taken twice daily for 10 days has had an eradication rate as a salvage therapy of 72 % and 76 % for ITT and PP analysis respectively (Van Der Poorten and Katelaris 2007).

As discussed above, there are many different treatment regimens available for treating an *H. pylori* infection. It is important to remember that clinicians treating this infection should be aware of antibiotic resistance rates of the local area, as well as patient allergies and previous exposures when deciding on an appropriate regimen, and not a global consensus.

1.3.5 Antibiotic Resistance

1.3.5.1 Resistance Rates

Antibiotic resistance is a worldwide issue affecting people from all socioeconomic backgrounds due to the overuse and misuse of these drugs (Laxminarayan et al. 2016). In 2017, the WHO published a priority list of antibiotic resistant bacteria based on ten criteria to assess where the future direction of antibiotic development should focus (Tacconelli et al. 2018). The priority list had three tiers; critical, high and medium priority, of which clarithromycin resistant *H. pylori* was placed in the high category.

A meta-analysis of WHO regions (Africa, Americas, Eastern Mediterranean, European, South East Asia and Western Pacific), spanning a 10-year period (2006 – 2016) showed a high level of resistance to treatment antibiotics (Savoldi et al. 2018). Overall resistance rates to clarithromycin, metronidazole, levofloxacin, amoxicillin, tetracycline and a combination of clarithromycin and metronidazole are as follows. Clarithromycin resistance was over 15 % in all regions except the Americas at 14 % (95 % CI, 9 % - 19 %). The rang of resistance was from 15 % in the Africa region (95 % CI, 0% - 30 %) to 34 % in the Western Pacific region (95 % CI, 30 % - 38

%). Metronidazole resistance was above 15 % in all regions, ranging from 27 % in the America region (95% CI, 14 % - 39 %) to 91 % in the Africa region (95 % CI, 87% - 94 %). Resistance to levofloxacin ranged from 14 % in the Africa (95% CI, 12 % - 28 %), Americas (95% CI, 12 % - 28 %) and European regions (95% CI, 12 % - 16 %) to 25 % in the South East Asia region (95 % CI, 13 % - 28 %). Amoxicillin resistance was below 15 % for all regions except the Africa region where it was 38 % (95% CI, 32 – 45 %). Tetracycline resistance was below 15 % in all regions. The combination of clarithromycin and metronidazole had a resistance rate of ≤ 15 % in all regions, except the Africa region where no data was available.

Resistance to clarithromycin, metronidazole and levofloxacin by country can be viewed in Figure 1.8. For these three antibiotics, resistance rates vary greatly geographically, it also shows one of the limitations of this study where many countries were not included in the meta-analysis. Small sample sizes from included countries, such as Africa, also poses a limitation of the analysis as this may not be a fair representation of the general population (Savoldi et al. 2018). A previous meta-analysis study in 2010 also found an increase in resistance to first-line antibiotics world-wide based on continental areas (De Francesco et al. 2010). Clarithromycin resistance rates ranged from 11 % in Europe to 29 % in Africa. Metronidazole resistance ranged from 17 % in Europe to 92 % in Africa. Levofloxacin resistance was 11 % in Asia, 24 % in Europe and 0 % in Africa with America having no available data. Tetracycline and amoxicillin resistance were under 10 % in Europe, Asia and America but 44 % and 66 % in Africa respectively. Comparing these two meta-analyses is somewhat challenging due to the different geographical organisation used but an overall increase in resistance can be seen in clarithromycin, metronidazole and levofloxacin.

Resistant rate studies in NZ and lacking with only three studies publicly available from 1999, 2004 and 2013. In the first study, only resistance towards clarithromycin and metronidazole were tested (Fraser et al. 1999). This study identified metronidazole resistant strains of *H. pylori* as the most prevalent, with 32 % of isolates tested showing resistance. Clarithromycin resistance was much lower, with only 7 % of isolates showing resistance (Fraser et al. 1999). The second study showed a drop in resistance for both metronidazole and clarithromycin, with 20 % and 0 % of isolates showing resistance, respectively (Ahmed et al. 2004). They also tested for amoxicillin and tetracycline resistance but did not identify any resistant strains (Ahmed et al. 2004). Nine years later, Hsiang et al. (2013) showed an increase in metronidazole, clarithromycin and amoxicillin resistance to 49.3%, 16.4 % and 5.5 %, respectively. Once again, no tetracycline resistance was identified. This shows a general increase in metronidazole, clarithromycin and amoxicillin resistance in NZ over a fourteen-year period.

Overall, resistance rates vary in different regions all over the world. This reinforces the idea that treatment of *H. pylori* infections must take in to account local resistance rates and not rely on a global consensus for effective eradication.

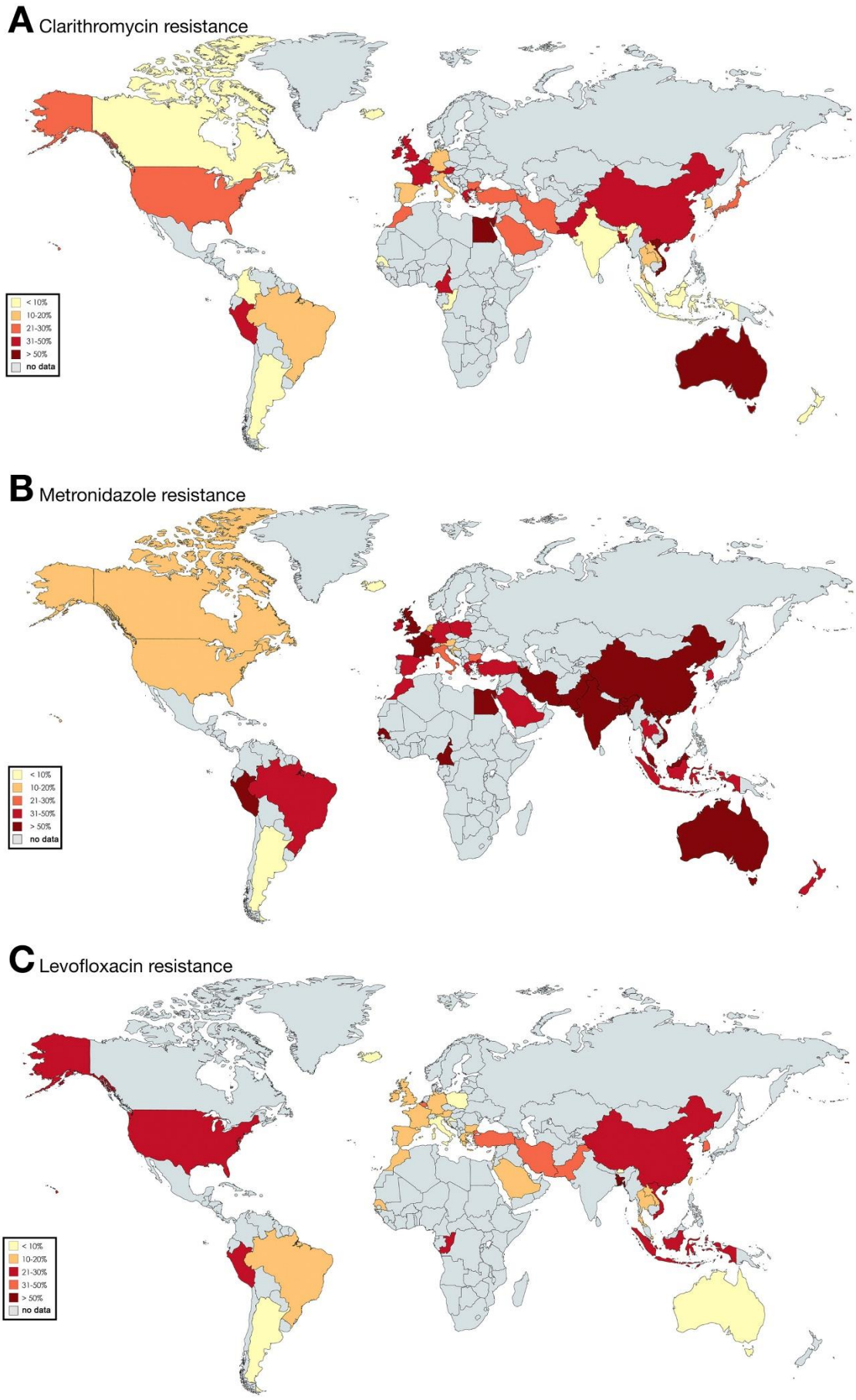

Figure 1.8: Resistance rates of clarithromycin (A), metronidazole (B) and levofloxacin (C) by country. Image retrieved from (Savoldi et al. 2018).

The increasing rate of resistance to first-line antibiotics has an adverse effect on the performance of treatment regimens (Savoldi et al. 2018). Eradication rates of triple therapies have decreased below 90 % in some areas such as Korea (Kim et al. 2007). Eradication rates of treatment regimens ≤80 % have been proposed as unacceptable and an alternative regimen must be found that has either a 95 % - 100 % eradication rate, or if this is unattainable, an eradication rate of 90 % - 95 % (Graham et al. 2007).

1.3.5.2 Mechanisms of Antibiotic Resistance

Bacteria can be intrinsically resistant to antibiotics or they can acquire resistance through chromosomal mutations or horizontal gene transfer (Blair et al. 2014; Hawkey 1998). Intrinsic resistance occurs naturally and is species and genus specific. Comparatively, acquired or developed resistance is only present in certain strains of a species through the acquisition of plasmid DNA or mutations (Hawkey 1998).

There are four main mechanisms of antibiotic resistance, drug inactivation through chemical alteration or destruction of the antibiotic, changes in target site through target protection or modification of the target site, active efflux, and reduced membrane permeability (Figure 1.9) (Blair et al. 2014; Hawkey 1998).

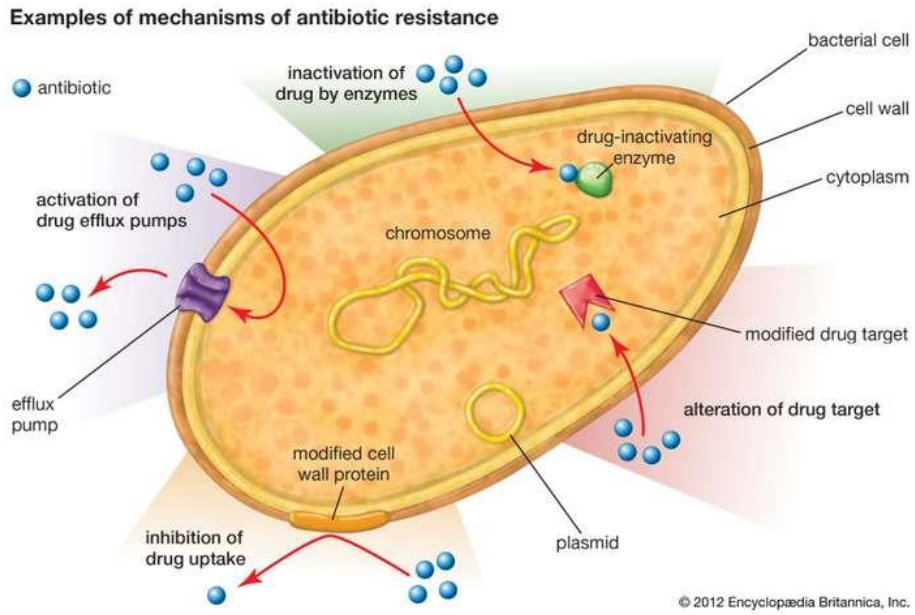

Figure 1.9: Mechanisms of antibiotic resistance. Image retrieved from https://www.britannica.com/science/antibiotic-resistance.

Within NZ, levofloxacin, rifabutin, and bismuth are not utilised in the treatment of *H. pylori* infections (New Zealand Formulary 2020b). Thus, the following antibiotics and associated mechanisms of resistance discussed are only those associated with *H. pylori* treatment in NZ.

1.3.5.3 Clarithromycin Resistance

Decreased eradication rates to commonly used triple therapies is largely associated with the rise in clarithromycin resistance (Horiki et al. 2009). The most well documented mechanism of resistance towards clarithromycin by *H. pylori* are mutations within domain V of the 23S rRNA gene. These point mutations decrease the affinity of the drug to the ribosome, thereby preventing it from interfering with protein synthesis (Stone et al. 1997; Versalovic et al. 1996; Vester and Douthwaite 2001). The first documented mutations conferring clarithromycin resistance were identified in 1996; an adenine-to-guanine transition at positions 2146 or 2147 (A2146G and A2147G) of the 23S rRNA gene (Table 1.3) (Versalovic et al. 1996). It is important to note that previous studies of the point mutations at nucleotides 2146 and 2147 have used different position numbers (e.g. 2058, 2059, 2514, 2515 and 2143, 2142). Through recent re-sequencing of the 23S rRNA gene the length of the *H. pylori* 26695 (NC_000915) gene has been re-evaluated and positions 2146 and 2147 have been proposed as the correct nucleotide numbers (Gong et al. 2020). Further studies have identified other mutations within domain V, including adenine-to-cytosine at positions 2146 and 2147 (A2146C, A2147C), and adenine-to-threonine at position 2146 (A2146T) (Table 1.3) (Khan et al. 2004; Kim et al. 2002; van Doorn et al. 2001). Many studies have validated the presence of the A2147G, A2146G and A2146C mutations, with more than 90% of clarithromycin resistant *H. pylori* strains having a combination of these mutations present (Hansomburana et al. 2012; Mégraud 2004; Tamayo et al. 2017; van Doorn et al. 2001; Versalovic et al. 1996). Other mutations present within the 23S rRNA gene have been identified, however, their direct association with clarithromycin resistance has not been fully evaluated (Table 1.3). Of particular interest is the threonine-to-cytosine transversion at position 2186 (T21826). This mutation has been identified in clarithromycin resistant strains and is believed to be associated with high levels of clarithromycin resistance (Kim et al. 2002). However, Burucoa et al. (2005) have shown that T2186C is also present in clarithromycin sensitive strains, suggesting that this mutation is not directly related with clarithromycin resistance. Mutations have also been identified outside of domain V in resistant strains. A threonine-to-cytosine transition at position 2715 (T2715C) was identified in domain IV of the 23S rRNA gene conferring a low level of resistance (Fontana et al. 2002). These mutations may arise from spontaneous mutation, at an estimated rate of 8×10^{-10}, or through horizontal gene transfer at a transformation frequency of 2.2×10^{-6} per viable cell (Taylor et al. 1997; Wang et al. 2001).

Once again, the geographic spread of this bacterium has led to geo-specific mutations with differing MICs. Namely, the A2147G mutation has predominantly been found in European countries and Iran, while the A2146G mutation is commonly identified in Brazil but not identified

at all in Korean isolates (Agudo et al. 2010; De Francesco et al. 2007; Kargar et al. 2011; Kim et al. 2008b; Ribeiro et al. 2003; Wueppenhorst et al. 2009). The A2146G mutation has shown a higher MIC than the A2147G when assessed *in vitro* (De Francesco et al. 2009b; Versalovic et al. 1997; Versalovic et al. 1996). However, in therapeutic practice, eradication rates have been much lower in patients with *H. pylori* strains containing the A2147G mutation, showing contrary results to *in vitro* studies (De Francesco et al. 2009a). This difference in MIC and eradication rate has been suggested as a more severe conformational change in the gene specific to the point mutations observed in *H. pylori* or an overestimation of the MIC by the method used to determine the MIC (De Francesco et al. 2009b; Stone et al. 1997).

Other genes associated with clarithromycin resistance include *infB*, encoding translation initiation factor IF-2, and *rpl22*, encoding ribosomal protein L22 (Table 1.3) (Binh et al. 2014). A 9 base pair (bp) insertion or 3 bp deletion within the *rpl22* gene and a single point mutation within the *infB* gene causing a guanine-to-adenine mutation at position 160 (G160A) have been identified as being involved with increased resistance towards clarithromycin (Binh et al. 2014). Both of these genes are associated with the large ribosomal subunit and have been identified as conferring resistance to erythromycin in *E. coli* (La Teana et al. 2001; Zaman et al. 2007).

Smiley et al. (2013) have identified several different OMPs with varying characteristics that were differentially expressed in clarithromycin resistant strains of *H.* pylori. They suggested that synergistically, these OMPs may provide an advantage for clarithromycin resistant *H. pylori*. The identification and further characterisation of OMPs in antibiotic resistance in *H. pylori* is an area that requires further study.

Efflux pumps, HefABC, HP1181, HP1184, have been associated with clarithromycin resistance and are discussed further in Section 1.3.5.7 (Falsafi et al. 2016; Hirata et al. 2010).

Table 1.3: Mechanisms of clarithromycin resistance in *H. pylori*.

Mechanism of resistance	Mutation associated with antibiotic resistance	Reference
23S rRNA mutation	A2147G	(Versalovic et al. 1996)
	A2146G	(Versalovic et al. 1996)
	A2146C	(van Doorn et al. 2001)
	A2148T	(Toracchio et al. 2004; Wang and Taylor 1998)
	G2145A	(Kocazeybek et al. 2019)
	T2715C	(Fontana et al. 2002; Hao et al. 2004)
	T2186C	(Khan et al. 2004; Kim et al. 2002)
	A2148G	(Hansomburana et al. 2012)
	A2120G	(Hultén et al. 1997)
	A2119G	(Hansomburana et al. 2012; van Doorn et al. 2001)
	G2115A	(Hansomburana et al. 2012)
	C2151G	(Garrido and Toledo 2007)
	G1943A	(Garrido and Toledo 2007)
	T1946C	(Garrido and Toledo 2007)
	T2293C	(Hao et al. 2004)
	G2228A	(Hao et al. 2004)
	C2249T	(Hao et al. 2004)
	T2194C	(Park et al. 2018)
	C2199T	(Park et al. 2018)
	T2248C	(Khademi et al. 2014)
	C2698A	(Rimbara et al. 2008b)
infB	G160A	(Binh et al. 2014)
rpl22	9 bp insertion or 3 bp deletion	(Binh et al. 2014)
Efflux pumps (HefABC, HP1181, HP1184)	HefA (N177T)	(Falsafi et al. 2016; Hirata et al. 2010)
Outer membrane proteins	NA	(Smiley et al. 2013)

NA = not applicable.

1.3.5.4 Metronidazole Resistance

In standard triple therapies, metronidazole resistance is a major contributor to eradication failure and is the leading resistance phenotype in NZ strains of *H. pylori* (Graham et al. 1992; Hsiang et al. 2013; Mégraud 2004). There are four possible mechanisms of resistance described for metronidazole resistant strains of *H. pylori*: reduced activity of nitroreductases, reduction in antibiotic uptake, increased efflux, increased DNA repair activity and alterations in the oxygen-radical scavenger system (Alba et al. 2017).

The natural occurrence of metronidazole resistance is due to the *de novo* mutational inactivation of the oxygen-insensitive NADPH nitroreductase (*rdxA*), RdxA, responsible for the reduction of hydroxylamine which damages the helical structure of DNA (Goodwin et al. 1998). The inactivation of NADPH flavin oxidoreductase (FrxA), encoded by *frxA*, in combination with RdxA inactivation has also shown metronidazole resistance (Marais et al. 2003). Inactivation of these genes through mutations renders these genes unable to encode the reductases required to reduce the prodrug metronidazole into its bactericidal agent through premature truncation (Goodwin et al. 1998; Kwon et al. 2001; Marais et al. 2003; Tanih et al. 2011). The combination of inactivation of the two reductases can increase the level of resistance but inactivation of FrxA by itself does not cause metronidazole resistance (Jeong et al. 2000; Yang et al. 2004). Truncated FrxA has been identified in metronidazole susceptible cells with intact RdxA (Yang et al. 2004).

There are a number of different mutations, frameshift, missense, nonsense, insertions and deletions, within *rdxA* and *frxA* have been identified in metronidazole resistant strains of *H. pylori* (Jenks et al. 1999; Kwon et al. 2000c; Marais et al. 2003; Tanih et al. 2011). Mutations in these genes leading to premature truncation and amino acid substitutions are found in many metronidazole resistant strains (Chua et al. 2019; Tanih et al. 2011; Yang et al. 2004). The *rdxA* mutations present in metronidazole resistant isolates differ. Some studies have identified a 200 bp deletion within the gene whereas others have not found this mutation (Abdollahi et al. 2011; Kargar et al. 2010). Many other studies have identified a range of missense mutations (Chua et al. 2019; Mirzaei et al. 2014; Tanih et al. 2011). Insertion of transposable elements within or adjacent to the *rdxA* gene has also been linked with metronidazole resistance (Debets-Ossenkopp et al. 1999; Tankovic et al. 2000). Due to the wide array of mutations within the *rdxA* gene and protein, it has been suggested that multiple mechanisms of resistance may be present in metronidazole resistant strains (Chisholm and Owen 2003).

Other nitroreductases present in *H. pylori*, ferredoxin, flavodoxin, 2-oxoglutarate oxidoreductase and pyruvate oxidoreductase, have been studied for their possible involvement in metronidazole resistance but inactivation of the genes encoding them led to cell death (Kwon

et al. 2000a). Ferredoxin-like protein, encoded by *fdxB* (HP1508), is a nitroreductase that may be involved in low rate metronidazole resistance in *H. pylori* through inactivation of the gene but is not well characterised (Kwon et al. 2000a). A single insertion located upstream of the ribosomal binding site, required for transcription or translation of the *rdxA* gene may be involved in metronidazole resistance where mutations are not observed directly within the gene (Moore and Salama 2005). Insertions and/or deletions or substitutions have also been identified in the upstream regions of *frxA* genes (Han et al. 2007).

Due to this prodrug requiring a reduction of its nitro-group to exert its bactericidal effect, the antimicrobial effect is only exerted on anaerobic or microaerophilic bacteria (Edwards 1993). Alterations in the redox potential at the site of reduction is an area of interest for mechanisms of resistance. *H. pylori* have exhibited a loss of resistance towards metronidazole when incubated in strictly anaerobic conditions (Cederbrant et al. 1992; Smith and Edwards 1995). The enzyme, nicotinamide adenine dinucleotide (NADH) oxidase, is believed to act as an "oxygen scavenger" that reduces intracellular oxygen tension, creating an ideal redox state and allowing metronidazole to be reduced (Smith and Edwards 1997). Low levels of NADH oxidase (an enzyme that reacts with oxygen to produce water or hydrogen peroxide) activity may reduce the oxygen-scavenger system efficiency (Smith and Edwards 1997). The excess oxygen present at the site of reduction may outcompete metronidazole from flavodoxins preventing the reductive activation of metronidazole (Jorgensen et al. 1998; Smith and Edwards 1997).

Mutations within the *fur* (ferric uptake regulator) regulatory gene can also have an effect on the cell's redox potential, potentially causing resistance towards metronidazole (Choi et al. 2011). The superoxide dismutase (SOD) enzyme encoded by the *sodB* gene is a virulence factor associated with *H. pylori* that protects the bacterium from superoxide attack by converting superoxide radicals to hydrogen peroxide or dioxygen (Pesci and Pickett 1994). This enzyme is regulated by the *fur* gene (Tsugawa et al. 2011b). Three main amino acid mutations have been identified within the *fur* regulator gene of metronidazole resistant *H. pylori* strains, arginine-to-isoleucine at position 3 (R3I), cysteine-to-tyrosine at position 78 (C78Y) and proline-to-serine at position 114 (P114S) (Albert et al. 2005; Tsugawa et al. 2011b). These mutations stop the repression of SOD and have been linked with the development of metronidazole resistance (Tsugawa et al. 2011b). The mutations described above have all been studied *in vitro* and their clinical relevance is still to be determined.

A less characterised mechanism of metronidazole resistance is the increased activity of the DNA repair mechanism RecA, encoded by *recA* (Chang et al. 1997; Thompson and Blaser 1995). The *recA* gene is found ubiquitously in bacteria and is involved in the bacterial SOS response for DNA

repair and natural transformation (Bergé et al. 2003; Cox 1991; Gaasbeek et al. 2009). Within *H. pylori*, mutations, tyrosine-to-histidine at position 103 (Y103H) and serine-to-aspartic acid at position 121 (S121D), in RdxA have been suggested as being involved in metronidazole resistance (Chang et al. 1997). Mutations leading to premature truncation of RecA have also been associated with metronidazole resistant strains of *H. pylori* (Thompson and Blaser 1995). However, other studies have not reported that mutations within *recA* do not contribute to high-level metronidazole resistance at all (Debets-Ossenkopp et al. 1999). This mechanism of resistance is still insufficiently described and further work is required to fully understand to role of *recA* mutations in metronidazole resistance within *H. pylori*.

Once again, the HefABC, HP1181, HP1184, efflux pumps have been suggested to be involved in metronidazole resistance and are discussed further in Section 1.3.5.7 (Falsafi et al. 2016; Hirata et al. 2010).

1.3.5.5 Amoxicillin Resistance

Amoxicillin resistance is low when compared to other antibiotics used in first-line treatment, with resistant MICs ranging from 1-8 mg/L (De Francesco et al. 2011). The reliance on this antibiotic for the successful eradication of the pathogen is high, and thus, monitoring and understanding its resistance mechanisms is crucial. Mechanisms of resistance involve target alteration and decreased permeability.

Commonly, gram-negative bacteria become resistant to β-lactams through the acquisition and production of β-lactamases that break open the β-lactam ring of the antibiotic making it unable to exert its bactericidal effect (Therrien and Levesque 2000). The presence of β-lactamases in *H. pylori* is not clear. Some studies indicate that there is no production of these enzymes in amoxicillin resistant isolates, however, Tseng et al. (2009) have shown the presence of the β-lactamase gene bla_{TEM-1} in a high-level resistant strain of *H.* pylori (Dore et al. 1999; Wu et al. 2000). As well as the identification of bla_{TEM-1}, treatment with β-lactamase inhibitors have shown an improved eradication rate compared to amoxicillin on its own (Ojetti et al. 2004).

Mutations in PBPs are also associated with amoxicillin resistance. As previously discussed, amoxicillin binds to PBPs which are required for the biosynthesis of the peptidoglycan layer of the cell wall. There are nine putative PBPs identified in *H. pylori*, three high-molecular-weight and six low-molecular-weight (Harris et al. 2000). Within all nine of these putative PBPs are three conserved domains: SXXK, SXN and KTG (X represents a variable amino acid) (Harris et al. 2000). Amoxicillin resistant strains exhibit a reduced affinity to PBP 1, 2 and 3 (Gerrits et al. 2006; Kwon et al. 2003; Okamoto et al. 2002; Rimbara et al. 2007; 2008a). Significant mutations conferring amoxicillin resistance include amino acid substitutions within or adjacent to the second and third

conserved domains of PBP1 (*pbp1A*) (SXN402-404 and KTG555-557) reducing the affinity of PBP (Gerrits et al. 2006). These mutations include substitutions, nonsense mutations, and insertions; mutations have also been identified in PBP2 and PBP3 (Table 1.4) (Rimbara et al. 2008a). Multiple mutations in PBPs are required to confer amoxicillin resistance. The serine-to-arginine substitution at position 414 (S414R) and an asparagine-to-tyrosine substitution at position 562 (N562Y) are the most common mutations identified in amoxicillin resistant strains of *H. pylori* (Gerrits et al. 2006; Kim and Kim 2013; Rimbara et al. 2007). Mutations within PBPs only confer low-level (MIC <8 μg/ml) amoxicillin resistance; additional mechanisms are required for a higher-level of amoxicillin resistance (Co and Schiller 2006; DeLoney and Schiller 2000; Gerrits et al. 2006; Kwon et al. 2003).

Decreased membrane permeability has been identified as an additional mechanism of resistance towards amoxicillin (Kwon et al. 2003). *H. pylori* possess five porin proteins: HopA to HopE (Doig et al. 1995). Of these six proteins, mutations within HopB and HopC are associated with increased amoxicillin resistance (Co and Schiller 2006). These mutations include amino acid changes within the 116 to 201 region of the *hopB* gene and a nonsense mutation leading to a stop codon at the 211th position of the *hopC* gene (Table 1.4) (Co and Schiller 2006). Other OMPs have also been associated with increased amoxicillin resistance, these include: Omp25 and Omp32 (Table 1.4) (Godoy et al. 2007). Analysis of gene expression when *H. pylori* becomes exposed to amoxicillin showed an upregulation in *omp25* and a downregulation in *omp32* (Godoy et al. 2007). In other Gram-negative bacteria Omp25 has been suggested as a porin, whereas Omp32 has been characterised as an anion-selective porin, suggesting that these proteins may be involved with increasing drug release (Boigegrain et al. 2004; Contreras et al. 2003; Godoy et al. 2007; Siroy et al. 2005).

The HefABC efflux pump has been involved in amoxicillin resistance and is discussed further in Section 1.3.5.7 (Hirata et al. 2010).

Table 1.4: Mechanisms of amoxicillin resistance in *H. pylori*.

Mechanism of resistance	Mutation associated with antibiotic resistance	Reference
Mutations within PBP1	N562Y (nonsense mutation)	(Kim and Kim 2013)
	A369T	(Rimbara et al. 2008a)
	V374L	(Rimbara et al. 2008a)
	S414R	(Gerrits et al. 2006)
	L423F	(Rimbara et al. 2008a)
	V16I	(Kim and Kim 2013)
	V45I	(Kim and Kim 2013)
	T593A	(Rimbara et al. 2008a)
	G595S	(Kim and Kim 2013)
	A599T	(Kim and Kim 2013)
	D535N	(Rasheed et al. 2014)
	S543R	(Rasheed et al. 2014)
	464E (insertion)	(Okamoto et al. 2002)
	Y637* (nonsense mutation)	(Matteo et al. 2008)
	T556S	(Tseng et al. 2009)
	T593A	(Kim and Kim 2013)
Mutations within PBP2	A296V	(Rimbara et al. 2008a)
	S494H	(Rimbara et al. 2008a)
	A541M	(Rimbara et al. 2008a)
	E572G	(Rimbara et al. 2008a)
Mutations within PBP3	A499V	(Rimbara et al. 2008a)
	E536K	(Rimbara et al. 2008a)
Decreased membrane permeability	HopB (amino acid substitutions within 116 to 201 region)	(Co and Schiller 2006)
	HopC (nonsense mutation at position 211)	(Co and Schiller 2006)
Efflux pump	HefC (L378F and D131E)	(Qureshi et al. 2014)
Other OMPs (Omp25, Omp32)	NA	(Godoy et al. 2007)

NA = not applicable. *= stop codon.

1.3.5.6 Tetracycline Resistance

Tetracycline is an essential part of quadruple treatment regimens. Tetracycline resistance is not common and MICs of resistant strains are ≥4 mg/L of tetracycline and mechanisms of resistance involve target alteration and efflux pumps (Kwon et al. 2000b).

The primary mechanism of resistance for tetracycline are mutations within the binding site of the 16S rRNA gene. Single mutations at binding site C of domain III (AGA965-967) confer tetracycline resistance by reducing the antibiotic binding to the ribosome (Wu et al. 2005). These single mutations include adenine-to-guanine at position 965 (A965G), adenine-to-thymine at position 965 (A965T), adenine-to-cytosine at position 965 (A965C), adenine-to-cytosine at position 967 (A967C) and adenine-to-thymine at position 967 (A967T) (Wu et al. 2005). Another single mutation has been identified in domain I, a guanine-to-adenine change at position 360 (G360A), however these were also associated with a triple mutation in positions 965 to 967 (Trieber and Taylor 2002). This triple mutation, AGA to TTC at position 965 to 967, were identified in tetracycline resistant *H. pylori* strains and were shown to induce tetracycline resistance through transformation to a previously susceptible strain (Trieber and Taylor 2002). This triple mutation is associated with high levels of tetracycline resistance and have been identified in both copies of the 16S rRNA gene present in *H. pylori* (Gerrits et al. 2003; Ribeiro et al. 2004). Finally, two deletions have also been observed in resistant strains, a guanine at position 771 in domain II and a guanine at position 942 in domain II. The deletion at G942, within the Tet-4 binding site (domain II) of the 30S ribosomal subunit may reduce the affinity of tetracycline thereby aiding the strain in tetracycline resistance, although this is not believed to be a main mechanism (Trieber and Taylor 2002).

The HP1184 and HP1181 efflux pumps are suggested to be involved in tetracycline resistance and are discussed further in Section 1.3.5.7 (Falsafi et al. 2016).

1.3.5.7 Efflux Pumps

Efflux pumps are found in both Gram-positive and Gram-negative bacteria and are involved in the extrusion of toxic compounds from within the cell to the external environment (Pearson et al. 1999; Truong-Bolduc et al. 2005). They are involved in bacterial virulence, resistance to nonclinical toxic components such as heavy metals and organic solvents, cell-to-cell signalling, and most importantly, antibiotic resistance (Kieboom et al. 1998; Kulathila et al. 2011; Lin et al. 2003; Okusu et al. 1996; Pearson et al. 1999). There are five main superfamilies of efflux proteins (Figure 1.10), the major facilitator superfamily (MFS), the ATP-binding cassette superfamily (ABC), the small multidrug resistance family (SMR), the resistance-nodulation-cell division superfamily (RND), and the multi antimicrobial extrusion protein family (MATE) (Delmar et al.

2014). These families are characterised based on transport function, substrate specificity, energy coupling and sequence similarity (Saier 1998).

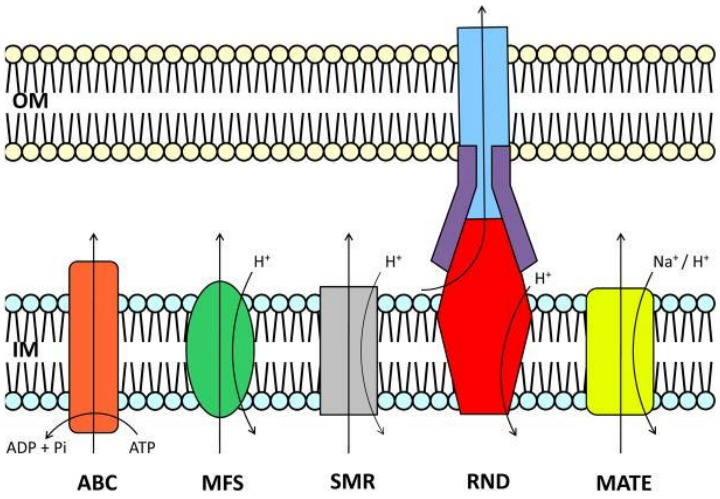

Figure 1.10: Efflux pump superfamilies. ATP-binding cassette superfamily (ABC), major facilitator superfamily (MFS), small multidrug resistance family (SMR), resistance-nodulation-cell division superfamily, and multi antimicrobial extrusion protein family (MATE). OM = outer membrane, IM = inner membrane. Image retrieved from (Delmar et al. 2014)

Efflux pumps contribute to the antibiotic resistance of many clinically relevant Gram-positive and Gram-negative bacteria (Gibbons et al. 2003; Okusu et al. 1996). The RND superfamily of efflux pumps are an important family that are associated with multidrug efflux in many Gram-negative bacteria (Chuanchuen et al. 2002; Ma et al. 1995; Magnet et al. 2001). This superfamily is unique from the others due to the association of the pump protein (RND) with two other protein classes, an OMP channel and a membrane fusion protein (MFP;

Figure 1.11). The most well characterised examples of this efflux family are the AcrAB-TolC efflux system from *E. coli* and the MexAB-oprM efflux system from *Pseudomonas aeruginosa* (Li et al. 1995; Okusu et al. 1996).

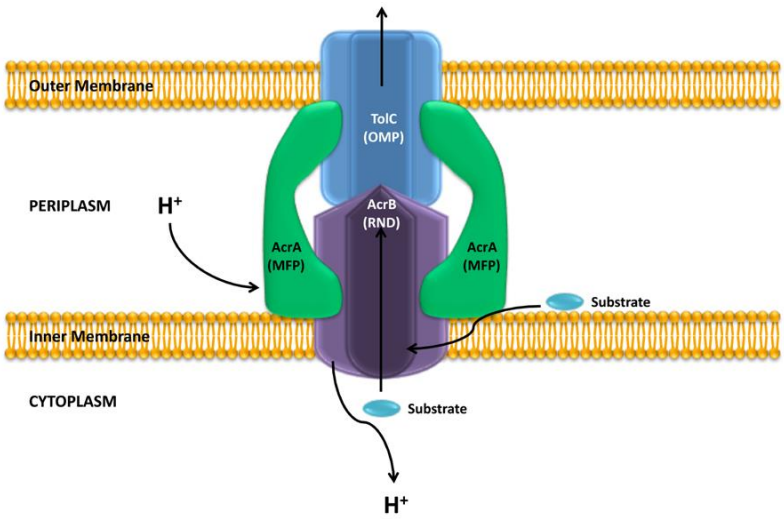

Figure 1.11: Structure of RND-type efflux pump, based on the AcrAB-TolC system from *E. coli*. Image retrieved from (Alvarez-Ortega et al. 2013).

Within *H. pylori*, four putative RND efflux systems have been identified, HP1489-HP1487, HefABC (HP0605-HP0607), HefFED (HP0969-HP0971) also known as CznABC; and HefIHG (HP1329-HP1327), also known as CzcA (HP1329), CzcB (HP1328) and CrdB (HP1327) (Bina et al. 2000; Johnson and Church 1999). The HefABC efflux pump is the most well characterised of these four efflux systems and is encoded by three genes, *hefA*, *hefB*, and *hefC*. The *hefA* gene encodes a protein homologue of the TolC OMP (HefA), the *hefB* gene encodes a protein homologue of the membrane fusion protein AcrA (HefB), and the *hefC* gene encodes a protein homologue of the RND pump, AcrB (HefC) (Bina et al. 2000). The other three systems, HP1489-HP1487, HefFDE and HefIHG are putative OMP, RND pumps and MFPs (Bina et al. 2000; Johnson and Church 1999; van Amsterdam et al. 2005). Initially it was believed that these efflux pumps did not play a role in intrinsic resistance towards antibiotics in *H. pylori* (Bina et al. 2000). However, their involvement in antibiotic resistance has since been shown (Liu et al. 2008; van Amsterdam et al. 2005).

Clarithromycin, metronidazole and amoxicillin resistance have been associated with the HefABC efflux system. The association of HefABC with clarithromycin resistance was suggested by Hirata et al. (2010), who showed that messenger RNA (mRNA) from the efflux system was present in resistant isolates; and the introduction of an efflux pump inhibitor decreased the MIC of clarithromycin. An asparagine-to-threonine change at position 177 (N177T) in HefA has also been linked to clarithromycin resistant isolates, however, this mutation was not identified in all resistant isolates (Iwamoto et al. 2014). HefA is directly involved in metronidazole resistance, shown through the knockout of this gene in a resistant strain leading to metronidazole sensitivity (Lee et al. 2018; van Amsterdam et al. 2005). The presence of metronidazole has also shown to increase the expression of TolC homologue HefA, suggesting that exposure to metronidazole

leads to metronidazole resistance (Mehrabadi et al. 2011; Tsugawa et al. 2011a). Amoxicillin resistance has been linked with the following mutations in HefC aspartic acid-to-glutamic acid at position 131 (D131E) and a leucine-to-phenylalanine change at position 378 (L378F) (Qureshi et al. 2014). Strains with both mutations have shown a 16x increase in MIC towards amoxicillin, whereas, strains with only the L378F mutation have shown an 8x increase in MIC (Qureshi et al. 2014).

Twenty-three other putative outer membrane efflux proteins belonging to the ABC, MFS and MATE superfamilies have been identified in *H. pylori* isolates (van Amsterdam et al. 2005). The *hp1184* and *hp1181* genes, encoding efflux pumps belonging to the MATE and MFS superfamiles, respectively, have been associated with active efflux in multidrug resistant strains of *H. pylori* (Falsafi et al. 2016). The *hp1165* gene, encoding an MFS efflux pump is associated with tetracycline resistance (Li and Dannelly 2006). Of the drug classes relevant to this study, HP1184 and HP1181 have been associated with macrolide, nitroimidazole and tetracycline resistance, however, they are also associated with the active efflux of other compounds and antibiotics (Falsafi et al. 2016; van Amsterdam et al. 2005). All of these efflux proteins have been found in susceptible and resistant strains of *H. pylori*, suggesting that post-transcriptional regulation is involved in the expression of resistant phenotypes (Falsafi et al. 2016; Li and Dannelly 2006).

1.3.5.8 Factors that can affect Antibiotic Resistance

In addition to antibiotic resistance mechanisms acquired or developed by the bacterium, there are other contributing factors to the failing eradication rates of treatments. These other factors fall under three main categories: bacterial strain related factors, host related factors and environmental factors.

Bacterial strain related factors aiding the bacterium in treatment survival have both been previously discussed, biofilm formation and coccoid formation (Section 0 and 1.3.1.1, respectively). Biofilm formation allows *H. pylori* to from a protective layer and this coupled with a high colony density can aid in the survival of antibiotic treatment (Stark et al. 1999). However, it has also been suggested that although *H. pylori* have the ability to form biofilms, they prefer a planktonic state of growth in the presence of gastric mucosa (Cole et al. 2004). Likewise, coccoid formation allows the bacterium to enter a self-protective state where they can survive antibiotic treatment and revert back to their helical forms when beneficial (Brenciaglia et al. 2000).

Host related factors include patient compliance, gene polymorphisms, smoking, age and gender. Poor compliance from patients has shown an association with a decreased eradication rate but

is not a sole reason as patients with good compliance have also shown a decreased eradication rate (Wermeille et al. 2002). Factors that influence patient compliance with successful *H. pylori* eradication include: complexity of treatment, side effects of treatment, treatment efficiency, motivation of physician, patient information and therapy duration (O'Connor et al. 2009). Genomic polymorphisms in patients such as the cytochrome P450 2C19 (CYP2C19) variant in exons four and five of the gene can have an unfavourable effect on the efficiency of PPIs during treatment due to the enzymatic pathway used to metabolise PPIs (Kuo et al. 2014). Several studies have shown smoking while undergoing eradication therapy decreases the chances of a successful eradication (Camargo et al. 2007; Itskoviz et al. 2017; Kamada et al. 1999; Suzuki et al. 2006). This is due to a variety of factors that include decreased blood flow in the gastric mucosa hindering the delivery of eradication drugs and an increased acidic environment that encourages more non-replicative bacteria that eradication drugs cannot target (Itskoviz et al. 2017). Age and gender are also factors affecting successful eradication of *H. pylori* with females and patients under 50 having a lower eradication rates (Chang et al. 2019; Mamori et al. 2010).

Environmental factors principally involve the occurrence of reinfection. After successful completion of the treatment plan, patients have a four-week wait before confirmation of eradication. Within this period reinfection may occur from previously mentioned sources (Section 1.3.2.3) such as the patient's oral cavity or through contaminated water sources, especially in regions where water sanitation and hygiene are lacking (Klein et al. 1991; Nurgalieva et al. 2002; Song and Li 2013).

1.3.6 Genomics

1.3.6.1 16S rRNA

For many years, the 16S rRNA gene has been used as an important tool for helping understand the phylogenetic relationships between bacteria (Woese and Fox 1977). This gene encodes the RNA component of the 30S ribosomal subunit and is ubiquitously found in all bacterial species in singular or multiple copies (Wang et al. 2015c). Using the 16S rRNA gene as an accurate chronometer is attributed to a variety of factors, including; its functional consistency, its ubiquity, its length, and the conserved and variable regions within the gene (Johnson et al. 2019; Woese 1987; Woese et al. 1975).

As well as helping understand bacterial phylogeny, the 16S rRNA gene has been used as a means to identify and classify bacteria at a genus and species level, especially in the absence of defined or recognised biochemical characterisations or rare isolation of the bacteria (Johnson et al. 2019; Mignard and Flandrois 2006; Woo et al. 2003). Accurate identification at the genus level is

attainable in most instances, however accurate species identification is considered less achievable and reliable (Woo et al. 2003).

The use of the 16s rRNA gene in identification of bacteria is particularly useful in the clinical laboratory setting. Through conventional methods, many bacterial pathogens are able to be identified from an isolated colony within 24 hours. However, in the case of slow growing, fastidious, and poorly identified isolates though conventional techniques, molecular identification is a useful tool (Patel 2001).

1.3.6.2 Sanger Sequencing

The Sanger sequence method was first described in 1977 as the chain-termination or dideoxy technique. It comprises of dideoxynucletoides (ddNTP) and a DNA polymerase that prevents chain elongation within *in vitro* DNA replication (Sanger et al. 1977). Since its discovery, the technique has become more streamlined and economical. The modern method, displayed in Figure 1.12, first involves polymerase chain reaction (PCR) amplification of the target DNA; then sequencing follows by incorporation of the amplified target DNA, primers, DNA polymerase, deoxynucleotides (dNTP) and a small amount of fluorescently labelled ddNTPs. The mixture then undergoes another PCR; the sample is denatured, primers are annealed, and the chain elongates. This process is repeated numerous times allowing the complementary addition of dNTPs and fluorescently labelled ddNTPs to the template strand. Random incorporation of the ddNTPs induces chain termination due to the lack of the 3' hydroxyl group preventing DNA polymerase from binding another dNTP onto the growing chain; producing chains of varying lengths (Sanger et al. 1977; Smith et al. 1986). Once this process is complete and there is ample DNA at varying lengths, it is separated by size using capillary gel electrophoresis and the fluorescently labelled ddNTPs are read by a laser and detector which is recorded by a computer for sequence analysis (Swerdlow and Gesteland 1990). The produced sequence is the complementary strand to the template DNA, with the same 5' end but differential 3' ends depending on which ddNTP was incorporated at that position. Due to the fluorescently labelled ddNTPs, this method is commonly referred to as dye-terminator sequencing.

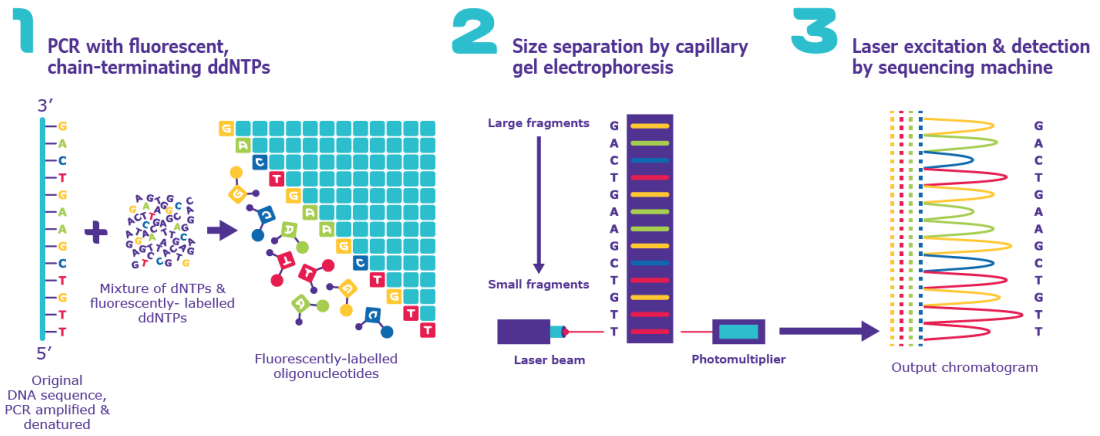

Figure 1.12: Outline of the modern process of Sanger sequencing. Image retrieved from https://www.sigmaaldrich.com/technical-documents/articles/biology/sanger-sequencing.html.

Sanger sequencing produces read lengths in a range of approximately 500-800 bp (Mardis 2017; Morozova and Marra 2008; Venter et al. 2001). The method is error prone within the first and last 15-45 nucleotides (nt) of the sequences producing poor quality data due to primer binding (Binladen et al. 2007; Ebili et al. 2017). These factors make Sanger sequencing an ideal choice for small sequence projects, such as the relatively short length gene, 16S rRNA.

1.3.6.3 Next-generation Sequencing

For large scale sequencing projects, focus has shifted towards new sequencing technologies known as next generation sequencing (NGS). This new range of sequencing technologies was established to reduce time and costs associated with Sanger sequencing and other related first-generation sequencing techniques (Schuster 2008). Since its development, NGS has become a widely utilised tool within the field of life-sciences (Schuster 2008).

Pyrosequencing

The first NGS technology developed was an improved Pyrosequencing method, based on the sequencing-by-synthesis principal, involving the denoted inclusion of a nucleotide by the release of a pyrophosphate and subsequent generation of photons (Margulies et al. 2005; Ronaghi et al. 1996). Pyrosequencing found its place with its superior detection limit of 5%, making it the preferred choice for short sequences, where Sanger sequencing struggles, and solving ambiguities within Sanger sequences (Harrington et al. 2013). Even with these superiorities over Sanger sequencing, Roche closed its 454-Pyrosequencing platform in 2013 when it became non-competitive against other NGS services (Bio IT World 2013).

Illumina Sequencing

Illumina sequencing is another sequence-by-synthesis approach; however, this technology differs from 454-Pyrosequecning in two main ways: the use of flow cells instead of microwells with beads and it uses a dye-termination approach instead of pyrosequencing (Illumina 2017). The workflow of Illumina sequencing follows a basic three step process: library preparation, amplification and sequencing (Figure 1.13). Library preparation begins with the fragmentation of the sample and then subsequent addition of adaptors onto both the 5' and 3' ends of the DNA (Craig et al. 2008; Hodkinson and Grice 2015). The amplification or cluster generation step involves the tagged library being placed onto a flow cell where oligonucleotides, complementary to the ligated adaptors, hold the single-stranded fragments in place. The fragments are simultaneously amplified from a primer on the 3' end of the attached fragment via bridge PCR to create groupings of single-stranded clonal fragments (Bentley et al. 2008). From the clustered clonal fragments, sequencing may begin. Illumina sequencing uses a reversible terminator method that is similar to Sanger sequencing in principle (Canard and Sarfati 1994; Sanger et al. 1977). Many fluorescently labelled nucleotides are simultaneously added to the flow cell allowing one base addition before chain termination. Excess nucleotides are washed away, and the flow cell is imaged. The terminators are reversed, cleaving the fluorescent label and blocker, allowing chain elongation to continue and fluorescently labelled nucleotides are once again added. This cycle continues until sequencing is complete (Bentley et al. 2008; Canard and Sarfati 1994; Hodkinson and Grice 2015).

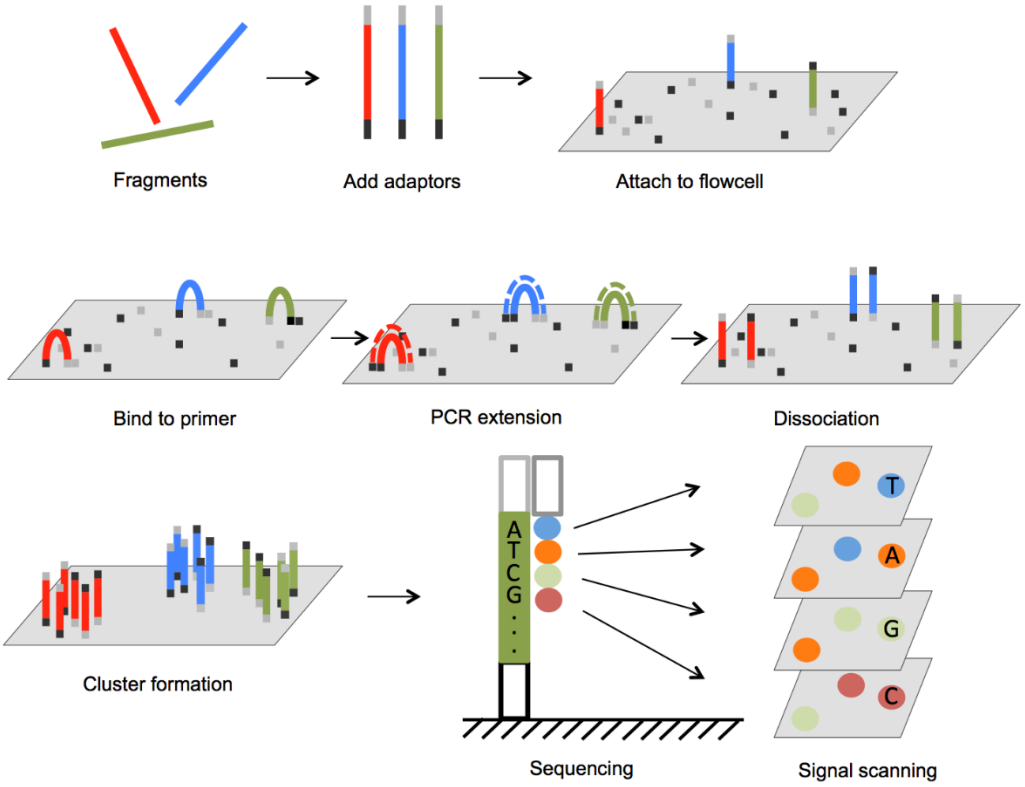

Figure 1.13: Workflow of Illumina sequencing. Image retrieved from (Lu et al. 2016).

Illumina MiSeq™ sequencing produces a large number of sequences, approximately 1 giga byte (Gb), consisting of pair-end (PE) reads up to 150 bp in length; Illumina HiSeq™ can produce up to 600 Gb of 150 bp PE reads (Caporaso et al. 2012; Illumina 2020; Morozova and Marra 2008; Quail et al. 2012). This technology has a low raw error rate of 0.8%, its limitations lie within its short sequence reads making it difficult to resolve short sequence repeats and assembling *de novo* sequences; and inaccuracies within GC rich areas (Morozova and Marra 2008; Quail et al. 2012; Wold and Myers 2008). However, due to the low cost and decreased run time, this sequencing method has become widely applied in the fields of WGS and functional genomics (Morozova and Marra 2008; Yohe and Thyagarajan 2017). In the scope of antibiotic resistance, WGS has become a widely used tool to discover and monitor the fast-paced changes of resistance due to the low cost, deep coverage and potential ease of discovering resistance in slow growing and fastidious microorganisms (McDermott et al. 2016; Schürch and van Schaik 2017; Su et al. 2019b).

SOLiD Sequencing

Sequencing by Oligonucleotide Ligation and Detection (SOLiD) sequencing uses a sequencing-by-ligation approach. This sequencing produces highly accurate read lengths of between 25 – 35 bp (Liu et al. 2012; Mardis 2017; Margulies et al. 2005). While SOLiD sequencing has found its place within whole genome resequencing, targeted resequencing, transcriptome research and

epigenome research, it has not reached the same level of success as other NGS technologies (Hodkinson and Grice 2015; Liu et al. 2012).

Ion Torrent Sequencing

Ion torrent sequencing is another sequencing-by-synthesis approach. Library preparation is similar to 454-Pyrosequencing, where DNA fragments are attached to beads within a microwell (Hodkinson and Grice 2015). Unmodified nucleotides, are sequentially added to the microwells; incorporation of a base to a growing DNA strand releases a proton which increases the surrounding pH that is detected by a pH meter under the microwell (Hodkinson and Grice 2015; Liu et al. 2012). Ion torrent sequencing has a raw error rate of 1.71%, and has issues with the accuracy of homopolymer sequences (Hodkinson and Grice 2015). It can produce read lengths approximately 200 bp in length (Quail et al. 2012). Due to the low cost and fast runtime of this sequencing platform, it has become a popular choice for medium-sized projects such as microbial genomes and microbiome research (Milani et al. 2013; Rusk 2011).

Pacific Bioscience Sequencing

Pacific Biosciences' (PacBio) release of single-molecule, real-time sequencing (SMRT) technology was the start of the third generation of sequencing technology (Eid et al. 2009). The method differs significantly from the previously described methods. Library preparation involves the fragmentation of the DNA into multi-kbp pieces and end repair with A/T overhang or blunt SMRT loop adaptor ligation (Travers et al. 2010). A major difference from other NGS technologies is the absence of an amplification step, instead it moves straight onto sequencing. Sequencing is executed on a chip with zero-mode waveguide (ZMW) wells; DNA polymerase is attached to the ZMW wells and fluorescently labelled nucleotides are added. Incorporation of a nucleotide emits light which is imaged in real time and with no termination groups present, chain elongation is continuous (Eid et al. 2009).

Initial release of this sequencing platform generated concern around the high error rate. The error rate has since been improved with the incorporation of circular consensus sequencing. This allows repeated sequencing of fragments thereby reducing the error rate to approximately 12.86 %, with errors throughout the read and not localised towards the ends (Buermans and den Dunnen 2014; Hodkinson and Grice 2015; Quail et al. 2012).

An advantage of PacBio sequencing is the long-read lengths, initial release of the platform saw read lengths of 1500 bp, whereas updated technology produces read-lengths of around 10 – 15 kbp (AllSeq 2020; Quail et al. 2012). PacBio sequencing is a strong sequencing technology for *de novo* assembly of reference genomes, especially for closing gaps, for the sequencing of long repetitive regions and the identification of mutations (Rhoads and Au 2015). However, its high

error rate and high cost render this technology still out of reach for many researchers and laboratories (Quail et al. 2012).

Oxford Nanopore Sequencing

Nanopore sequencing is another third generation sequencing approach that uses a single molecule of DNA or ribonucleic acid (RNA) with no amplification step (Buermans and den Dunnen 2014). Library preparation involves the ligation of adaptors to the fragmented DNA or complementary DNA (cDNA), used to capture and load a processive enzyme at the 5' end of the strand. Sequencing works by enzymes unwinding the DNA and passing it through a pore with a distinct current (Jain et al. 2016). Changes in ionic current due to a changing nucleotide in the DNA strand are detected (Jain et al. 2016). This technology can produce read-lengths ranging from 60 – 300 kbp (Jain et al. 2018; Oxford Nanopore 2020). Error rates of this method have recently dropped to 5 – 10 % (Tyson et al. 2017). Application of this sequencing technology has focused on smaller genomes such as bacteria, viruses and parasites commonly found in clinical settings for fast identification (Greninger et al. 2015; Kilianski et al. 2015; Quick et al. 2016).

1.3.6.4 Bioinformatics

Bioinformatics is the culmination of many disciplines, using computer-based analysis to understand biological questions. This fast-changing, computer-based discipline aims to efficiently and effectively store and understand the large amounts of data obtained from an increasingly sequenced based future (Baxevanis 2001). Bioinformatics use a workflow or pipeline to elucidate the identification and function of genes, in a rapid timeframe when compared to doing it by hand (Leipzig 2016; Overbeek et al. 2013). A basic pipeline of bioinformatics follows three main analysis stages: primary analysis, involving the detection and analysis of raw sequence data; secondary analysis, the alignment of sequence reads against a reference genome or *de novo* assembly; and tertiary analysis, variant annotation, filtering, and visualisation to provide context and links between obtained data and phenotypic observations (Pereira et al. 2020).

Primary Analysis: Analysis of Raw Reads

Initial analysis starts with assessing the quality and errors of the raw reads, denoted by the Phred score as well as the trimming of the adaptors (Pereira et al. 2020). As described above, each sequencing technology differs in its type and amount of errors. It is therefore important to search and analyse the raw reads for the common errors of that particular technology.

Analysis of raw reads is conducted to ensure that there is minimal contamination and enough sequence coverage of the target template, as well as summary statistics for assessing the overall quality of the sequencing (Pfeifer 2017). Quality of sequence data is primarily measured by a

Phred score, where each nucleotide is given a probability of it being erroneous (Bokulich et al. 2013; Ewing and Green 1998). This measurement is important in Illumina sequencing where the main error is miscalled bases (Hoffmann et al. 2009). The trimming of adaptors, used during library preparation, is an essential step to improve the alignment of raw reads against a reference and reduce the chance of erroneous variant calls (Pfeifer 2017).

Secondary Analysis: Genomic Assembly

As previously mentioned, once raw sequence reads have been obtained the next step involves the assembly into a complete genome. Illumina sequencing produces reads approximately 150 bp in length. These raw reads are overlapped to form contiguous sequences (contigs) that form a consensus sequence of DNA in that section (Ekblom and Wolf 2014). There are many programs available for microbial genome assembly, however, those used in this thesis are as follows: SPAdes, a bacterial *de novo* assembler (Bankevich et al. 2012); Unicycler, a tool used to assemble bacterial genomes from both long and short reads, in conjunction with SPAdes, to produce accurate assemblies (Wick et al. 2017); Bandage, an interactive visualisation tool used to view *de novo* assemblies (Wick et al. 2015); and Nullarbor, a pipeline designed for producing complete public health microbiology reports (Seemann et al. n.d.).

Nullarbor combines a number of programmes to assemble, annotate and analyse Illumina paired-end sequencing data to produce a final report including: sequence data, species identification, MLST, resistome, virulome, assembly and annotation, core genome analysis, core SNP phylogeny, pairwise core SNP distances, core SNP density and pan genome analysis (Seemann et al. n.d.). A list of the included programmes is presented in Table 1.5.

Table 1.5: Programmes included in Nullarbor to produce the final report.

Programme	Reference
Clean reads	
Trimmomatic	(Bolger et al. 2014)
Species identification	
Kraken + Centrifuge	(Kim et al. 2016; Wood and Salzberg 2014)
***De novo* assembly**	
SPAdes	(Bankevich et al. 2012)
MegaHit	(Li et al. 2015)
Annotation	
Prokka	(Seemann 2014)
MLST	
MLST + *pubMLST*	(https://github.com/tseemann/mlst) (Jolley et al. 2018)
Resistome	
Abricate + *Resfinder*	(https://github.com/tseemann/abricate) (Zankari et al. 2012)
Virulome	
Abricate + *VFDB*	(https://github.com/tseemann/abricate) (Chen et al. 2016)
Variants and core genome-SNPS	
Snippy + Snippy-core	(https://github.com/tseemann/snippy)
Infer core SNP phylogeny	
IQTree + FastTree	(Nguyen et al. 2015; Price et al. 2010)
Snp-dist	(https://github.com/tseemann/snp-dists)
Pan genome	
Roary	(Page et al. 2015)

Italics indicate a database.

Tertiary Analysis: Gene Prediction and Annotation

Gene prediction and annotation creates the link between the sequence and the biology of the organism. Applying layers of analysis and interpretation to a DNA sequence provides biological understanding and function to an otherwise arbitrary sequence (Stein 2001). Characterisation and analysis of genome structure may be understood at a range of levels; the overall nucleotide

structure at the DNA and gene level, including relevant statistics; at the coding and theoretical protein level; and at the non-coding and regulatory elements level (Wei et al. 2002). Application of this technology can provide information on evolution, conserved regions, insertions and/or deletions, tandem repeats and single nucleotide polymorphisms (Read et al. 2002; Wei et al. 2002).

Using automated annotation pipelines has hastened the process of gene prediction and annotation by using homologueous genes from previously annotated close-relative genomes against the newly sequenced DNA (Richardson and Watson 2012). This process can identify protein-coding genes, RNAs, transfer RNAs (tRNA), small RNAs, insertion sequences (IS), transposons, mobile elements and other genomic features previously annotated in other genomes (Stein 2001; Tatusova et al. 2013). Limitations of this process are seen within the contradiction of the use of a reference annotation and the aim of sequencing a new genome, to find a difference between close relatives (Richardson and Watson 2012). Genes or other genomic features not present in the reference genome may not show up in a new sequence annotation and be labelled as a hypothetical protein (Richardson and Watson 2012).

There are many tools available for gene prediction and annotation but in this study two main programmes were used: Rapid Annotations using Subsystems Technology (RAST) (Aziz et al. 2008) and Prokka (Seemann 2014). RAST and Prokka have both been widely used in combination with each other for gene prediction and annotation of bacterial genomes (Adelskov and Patel 2015; Busch et al. 2017; Huang et al. 2016a; Mostafa et al. 2019; Noda et al. 2018; Yuki et al. 2017). RAST is an online, fully automated annotation service that can produce annotations for complete or near-complete archaeal or bacterial genomes in 12 – 24 hours (Aziz et al. 2008). This software uses a multistep process that utilises GLIMMER3 for initial gene prediction followed by subsequent basic local alignment search tool (BLAST) searches and similarity within a growing library of manually curated subsystems to predict genes and assign functions (Aziz et al. 2008; Overbeek et al. 2013). Prokka is a command line operated software that is capable of producing annotations for bacterial genomes in 10 minutes (Seemann 2014). Similar to RAST, Prokka also uses a hierarchal, multistep process to identify genes and functions using BLAST+; utilising smaller trustworthy databases before moving onto domain-specific databases such as UniProt and RefSeq, with final searches within Hidden Markov Model (HMM) databases such as Pfam and TIGRFAM (Seemann 2014). If no matches are found, the label 'hypothetical protein' is assigned (Kilianski et al. 2015).

1.3.7 Comparative Genomics

Comparative genomics is broadly defined as the comparison of genomic features that helps determine biological understanding (Xia 2013). This branch of genomics aims to identify similarities and differences of compared genomes with the underlying assumption that the genomes share a common ancestor, to understand how evolutionary forces govern genomic changes, and to discern how this evolution can help improve personal and environmental health and other worldly issues (Ureta-Vidal et al. 2003; Xia 2013).

1.3.7.1 Ortholog Analysis

Orthologues are homologueous genes that are found in different species that encoded proteins with biologically equivalent functions that derive from a common ancestral gene through a speciation event (Primrose and Twyman 2009). Identifying and understanding these homologueous genes is an important first step in comparative genomics used to accurately describe similarities and differences and help discern phylogeny (Gabaldón 2008; Gabaldón and Koonin 2013; Sennblad and Lagergren 2009).

Clusters of Orthologous Groups/Genes of proteins (COG) was a database that was designed to classify proteins from bacteria, archaea and eukaryotes based on the orthology concept of phylogeny (Tatusov et al. 2000). The basic principal of the COG database is: if any two proteins from varying lineages fall within the same COG assignment, they are considered orthologs (Tatusov et al. 2000; Tatusov et al. 1997). COGs were identified using an all-against-all sequence comparison using a BLAST algorithm that encompasses both slow-evolving and fast-evolving genes (Tatusov et al. 2000). The COG database is still a popular tool used in comparative genomics, although the original database (https://www.ncbi.nlm.nih.gov/COG/) is not maintained. Many tools have become available to classify newly sequenced genes into COG categories; of such tools the following have been used in this thesis: Kyto Encyclopaedia of Genes and Genomes (KEGG) (https://www.kegg.jp/) and The evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG) v4.5.1 (http://eggnogdb.embl.de/).

KEGG

The KEGG database is an integrated resource used for the biological interpretation of genomic sequences and high-throughput data to help establish links from collections of genes within a genome to the functionality of the cell or organism (Kanehisa et al. 2015). Orthologous groups (OG) are stored within the KEGG Orthology (KO) database and can be viewed through KEGG pathway maps, BRITE hierarchies and KEGG modules; all manually curated databases (Kanehisa et al. 2015). The KEGG pathway database contains maps of cellular networks that link genes to gene products and can be viewed as enzyme names or enzyme commission (EC) numbers

(Kanehisa and Goto 2000). Other annotation servers such as RAST incorporate the KEGG database into their annotation outputs, however, this external database is not maintained and can display inaccuracies from the original database when enzymes are updated or added (Aziz et al. 2008).

The BlastKOALA (KEGG Ortholog and Links Annotation) programme is KEGG's internal web-based tool (https://www.kegg.jp/blastkoala/) used to automatically annotate genomes by assigning a KO number to characterise genes and reconstruct KEGG pathways (Kanehisa et al. 2016). The annotation is achieved through an initial BLASTp search of query amino acid sequences against a non-redundant, pangenome dataset (KEGG GENES database) at either the family, genus or species level before KO assignment is performed using the KOALA algorithm involving GFIT (Gene Function Identification Tool) to evaluate sequence similarity scores, best-hit relations, protein domains and taxonomy groups for each gene (Kanehisa et al. 2013; Kanehisa et al. 2015; Kanehisa et al. 2016). The annotated genome is able to be reconstructed into KEGG pathway maps, BRITE hierarchies and KEGG modules based on the assigned K numbers of each gene using the KEGG Mapper Reconstruct Pathway tool (https://www.kegg.jp/kegg/tool/map_pathway.html) to visualise and interpret high-level functions encoded by the genome (Kanehisa et al. 2016).

EggNOG

The eggNOG database is a collection of orthologous relationships, gene evolutionary histories and functional annotations that are publicly available on a web-based interface (http://eggnog5.embl.de) (Huerta-Cepas et al. 2018; Jensen et al. 2008). eggNOG v5.0.0 is the most recent update of the database with an upgraded core organism set now consisting of 5,090 organisms (previously 2,031) (Huerta-Cepas et al. 2018). The eggNOG database contains and identifies OGs established from an all-against-all approach using Smith-Waterman alignments, reciprocal best matches, and triangular linkage clustering (Jensen et al. 2008).

eggNOG-Mapper (http://eggnog-mapper.embl.de/) is an online tool used for the functional annotation of sequences based on the precomputed ortholog clusters and phylogenies determined in the eggNOG v5.0.0 database (Huerta-Cepas et al. 2017; Huerta-Cepas et al. 2018). Protein query sequences are uploaded into the online interface and undergo a HMMER3 search against the eggNOG v5.0.0 HMM database to provide an initial orthologous group match and functional annotation. Functional annotation is completed by assigning the taxonomically closest ortholog and functional descriptions including: Gene Ontology (GO) terms (Gene Ontology 2015), KEGG pathways (Kanehisa et al. 2013) and COG functional categories (Galperin et al. 2015; Huerta-Cepas et al. 2017).

1.3.7.2 Whole Genome Analysis

Mauve

Through genome evolution, recombination, horizontal gene transfer and deletions can all alter the order and contents of genomes. Mauve is a tool that is used to create multi-sequence alignments to observe large-scale evolutionary events (Darling et al. 2004). Mauve uses local collinear blocks (LCB) to identify homologueous regions shared by two or more genomes that are being aligned that contain no genomic rearrangements. LCBs are defined by providing a minimum weight, a measure of confidence that the genome rearrangement is true. A higher minimum weight has greater confidence but lower sensitivity and a lower minimum weight has greater sensitivity but lower specificity (Darling et al. 2004).

The traditional Mauve algorithm uses a multi-step process to create an alignment, briefly described as follows. First, anchors, multiple Maximal Unique Matches (multiple-MUMs), are identified for a rapid-alignment approach; the anchors may be rearranged to indicate potential genomic rearrangements. Secondly, random anchors that are small and potentially spurious are removed by defining the minimum weight for LCBs. Thirdly, the parameters of the anchored regions are defined, and Mauve completes the gapped global alignment (Darling et al. 2004).

Advancements in the original algorithm have led to the emergence of the progressiveMauve algorithm (Darling et al. 2010). These improvements focus around aligning larger numbers of genomes and more divergent genomes as well as the default parameters and the overall alignment being more accurate (Darling et al. 2010). Other advancements have led to the development of the Mauve contig mover (MCM), an algorithm used to reorder contigs in draft genomes against a reference genome (Rissman et al. 2009).

PhyloPhlAn2

To help taxonomically identify and discern evolutionary relationships, PhyloPhlAn was developed to assign phylogeny and putative taxonomy for microbial genomes (Segata et al. 2013). PhyloPhlAn2 is built off of the original software and is able to resolve both closely related strains as well as deep-branching candidate divisions from 400 ubiquitous proteins, identified and optimised from 3,737 genomes (Segata et al. 2013).

Complete or draft genome assemblies may be used for input data into the software. The input data is searched for the 400 conserved proteins and the nucleic acid or amino acid sequences of the proteins are concatenated and aligned using MUlitple Sequence Comparison by Log-Expectation (MUSCLE) (Edgar 2004). Phylogenetic trees are then built using an external tree builder such as RAxML or FastTree (Price et al. 2010; Segata et al. 2013; Stamatakis 2006).

1.3.7.3 Identification of Antibiotic Resistance Determinants

CARD

The Comprehensive Antibiotic Resistance Database (CARD) is an online, manually curated database on the molecular basis of antimicrobial resistance (https://card.mcmaster.ca/home) (McArthur et al. 2013). The Antibiotic Resistance Ontology (ARO) was developed by CARD to classify integrated sequence and molecular data of antibiotic resistance genes (McArthur et al. 2013). Entries are only recorded within CARD if the sequences are available in GenBank and are associated with a peer-reviewed publication. Molecular data entries include all annotations, NCBI accession numbers, taxonomy identification numbers of pathogen hosts and associated PubMed publications. The database is continuously updated and is widely used to screen bacterial genomes for well-known and characterised resistance determinants (Alcock et al. 2020; Happel et al. 2020; Kim et al. 2018).

The Resistance Gene Identifier (RGI) is an in-built tool within CARD used primarily to annotate query genome sequences against data within CARD (https://card.mcmaster.ca/analyze/rgi) (Jia et al. 2017). Antibiotic resistance genes and mutations are predicted in query sequences using ORF prediction using Prodigal (Hyatt et al. 2010) and sequence alignment using Double index alignment of next-generation sequencing data (DIAMOND) (Buchfink et al. 2015) or BLAST (Camacho et al. 2009). Currently, the RGI is only able to detect non-synonymous substitutions. Frameshift mutations, insertions and deletions are not identified, the RGI is therefore limited in its abilities (Alcock et al. 2020). Searches within the RGI may be conducted within three paradigms: perfect, strict or loose (Jia et al. 2017). Combining a loose search using RGI and phenotypic screening may enhance the identification of novel antibiotic resistance genes (Alcock et al. 2020). The RGI has been applied to pathogenic bacteria to understand their antibiotic resistance genes (Chi et al. 2020; Ghosh et al. 2020; Grevskott et al. 2020).

OrthoVenn2

OrthoVenn is an online tool used to annotated, visualise and compare orthologous clusters within different species (https://orthovenn2.bioinfotoolkits.net/home) (Wang et al. 2015d). OrthoVenn2 builds on the original version through an updated platform, allowing comparison of up to 12 species at once and is ten times faster than the original OrthoVenn system (Xu et al. 2019). OrthoVenn2 uses a modified OrthoMCL algorithm (Li et al. 2003) for orthology cluster prediction (Figure 1.14). An all-against-all alignment is performed using DIAMOND (Buchfink et al. 2015) to identify regions of similarity within query genomes; orthAgogue is then used to identify presumptive orthology (Xu et al. 2019), and the output is visualised in a cluster table or Venn diagram (Figure 1.14). After identification of an ortholog cluster, the putative function of

the ortholog is determined through another DIAMOND (Buchfink et al. 2015) search through the non-redundant protein database within UniProt (The UniProt Consortium 2014) and may be displayed as GO terms and a protein network for each cluster identified (Figure 1.14). OrthoVenn has been used in comparative genomic studies to identify orthologous protein clusters potentially responsible for multidrug resistance in *A. baumannii* (Si-Tuan et al. 2020).

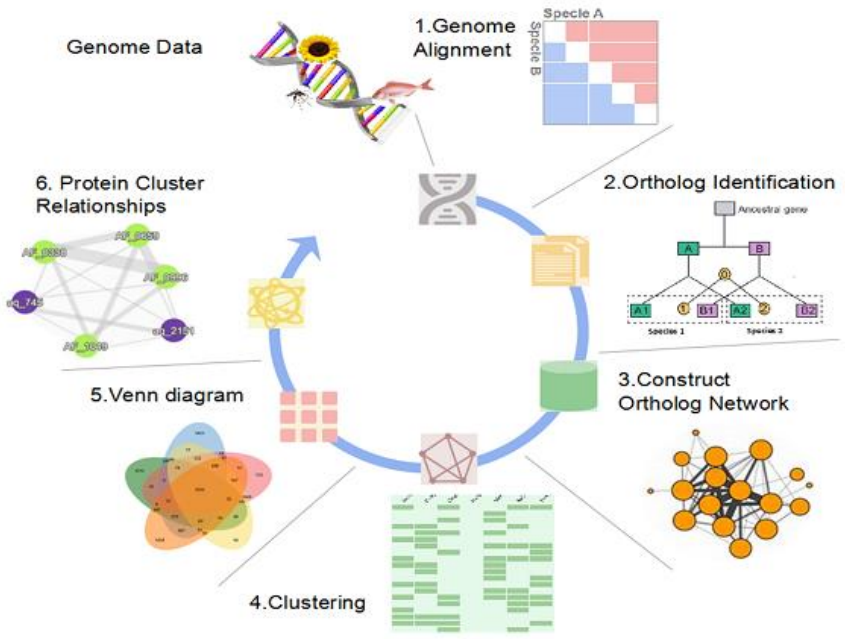

Figure 1.14: Workflow of OrthoVenn2. Image retrieved from https://orthovenn2.bioinfotoolkits.net/home.

1.3.8 Summary of Literature Review

Overall, this literature review has discussed many aspects of the human pathogen, *H. pylori*, as well as tools used for genomic sequencing and comparative analyses. It has highlighted the major morphological and genomic characteristics that define this bacterium and aid in its survival in its harsh environment. Requirements for growth have been described that will aid in RQ one of this thesis (Section 0). The vast global spread of *H. pylori* has been presented as well as the still unknown nature of the natural reservoirs and transmission of the pathogen. The range of clinical outcomes from an *H. pylori* infection underlines the potential severity of this bacterium on human health. These outcomes coupled with the rise of antibiotic resistance exhibited worldwide by this pathogen is creating challenges for clinicians to assign effective treatment regimens to eradicate *H. pylori*. The lack of data around NZ *H. pylori* isolates as well as resistance rates and mechanisms of resistance highlights the need for further research in this area. Finally, an overview of sequencing techniques and a range of comparative genomic tools were described that will be used in this thesis. These tools will help answer RQs two and three of this thesis (Section 0).

Chapter 2 Growth and Antibiotic Susceptibility Testing of *H. pylori*

2.1 Introduction

H. pylori is an extremely fastidious and slow growing organism *in vitro*. Initial culturing of this bacterium from clinical biopsies utilises selective media due to the intrinsic resistance these organisms show towards certain antibiotics; including vancomycin, cefsulodin, sulphonamides and trimethoprim (Blanchard and Nedrud 2012; Mégraud 1997). Primary culturing can take five to seven days before small, round and translucent colonies are observed on blood-based agars (Blanchard and Nedrud 2012). For more abundant growth of the bacterium, liquid media may be used, which are usually supplemented with a serum such as heat-inactivated horse or foetal bovine serum. Addition of serum has shown to increase the growth of *H. pylori*; however, this bacterium has also been observed to grow in liquid media without serum albeit less abundantly (Shahamat et al. 1991).

Antibiotic susceptibility testing is an important component of clinical microbiology used to identify and confirm susceptibility and resistance levels to chosen antimicrobial agents (Reller et al. 2009). For such analyses, the inoculum size required is important; however, for *H. pylori* antibiotic susceptibility testing, this parameter is not yet defined. A McFarland standard is a range of barium sulphate or latex particle suspension standards that are used as a visual reference test for the turbidity of bacterial suspensions (Courvalin and Soussy 1996). This visual reference is often used to determine the inoculum size required for accurate identification of antibiotic resistance, with each standard correlating to an approximate cell density (Table 2.1) (Dalynn Biologicals 2014). A general standard of 0.5 McFarland has been suggested to achieve semi-confluent growth for most bacteria; however, *H. pylori* inoculum sizes have ranged from McFarland 0.5 to 4 (Christofilogiannis 2001; Hartzen et al. 1997; Heep et al. 2002; Hombach et al. 2015).

Table 2.1: McFarland standards and their approximate cell density conversion (Dalynn Biologicals 2014).

McFarland standard	0.5	1	2	3	4	5
Approximate cell density ($x10^8$/mL)	1.5	3	6	9	12	15

The MIC is the lowest antibiotic concentration that will inhibit bacterial growth and can help in determining the breakpoint of an antibiotic, or the antibiotic concentration that defines resistance and susceptibility (Carson et al. 1995). Two antibiotic breakpoint guidelines are

primarily used to establish the resistance in clinical bacterial isolates; the Clinical Laboratory Standards Institute (CLSI) and the European Committee on Antimicrobial Susceptibility Testing (EUCAST) guidelines (Kassim et al. 2016). Phenotypic methods used for testing antibiotic susceptibility include agar dilution or breakpoint susceptibility testing, disc diffusion and E-test strips Figure 2.1. The agar dilution method involves the incorporation of varying concentrations of liquid antibiotic into molten agar before a known concentration of microbial inoculum is spread across the surface of the plates (Figure 2.1A). The plate where growth inhibition is observed identifies the MIC for that antibiotic for that amount of inoculum (Balouiri et al. 2016). Agar dilution is considered the gold standard method for establishing antibiotic resistance of *H. pylori*; however, it is not routinely used in clinical settings due to its high monetary and time costs (Clinical and Laboratory Standards Institute 2015 cited in: Miftahussurur et al. 2020; Valdivieso-García et al. 2009). E-test strips are also used to determine the MICs of antibiotics. These strips have a pre-defined gradient of antibiotic concentration from one end to the other, and are placed in the centre of bacterial inoculum spread on the surface of an agar medium (Figure 2.1B). The antibiotic diffuses from the strip into the agar, coming into contact with the bacteria on the plate. Just before the point of the strip where colonies are able to grow right up to the strip's edge is identified as the MIC (Miftahussurur et al. 2020; Miftahussurur et al. 2016). These two methods used for determining MICs of antibiotics are commonly used in antibiotic testing for *H. pylori* and have both achieved similar MICs for amoxicillin and clarithromycin (Glupczynski et al. 2002). In contrast, analysis of metronidazole has resulted in different MICs for each method and are considered unreliable (Glupczynski et al. 2002). Disc diffusion is a simple and economic method for routine testing of antibiotic susceptibility and is considered the gold standard method for confirmation of susceptibility for bacteria (Khan et al. 2019; Mégraud and Lehours 2007). Discs consisting of a single antibiotic concentration are placed onto inoculated agar plates (Figure 2.1 C). Resistance or sensitivity is determined by the zone of inhibition around the disc; these zones differ in size for each antibiotic and each concentration, with the MIC being the disc with the lowest concentration that causes growth inhibition (Ogata et al. 2015). This method has been employed with antibiotic susceptibility testing of *H. pylori* and is considered a cost-effective alternative to E-test methods (McNulty et al. 2002; Mishra et al. 2006).

Antibiotic resistance in NZ isolates of *H. pylori* has been understudied. While different strains of *H. pylori* have shown resistance to metronidazole and clarithromycin (Ahmed et al. 2004; Hsiang et al. 2013), the exact resistance(s) of the isolates in New Zealand is not well established. The aim of this part of the study was to establish antibiotic resistance phenotypes and MICs for different New Zealand isolates of *H. pylori* (Figure 2.2). Five isolates (A-E) were provided by Dr

Susan Taylor (Middlemore Hospital, New Zealand) for analysis. These isolates were initially retrieved from biopsies of patients undergoing gastroscopy and underwent routine antibiotic susceptibility testing to four commonly used antibiotics in New Zealand. In-house testing was conducted to confirm these findings using a variety of media and different testing methods. These data were then used to focus the genome analyses for molecular mechanisms of resistance (Chapters 3 and 4).

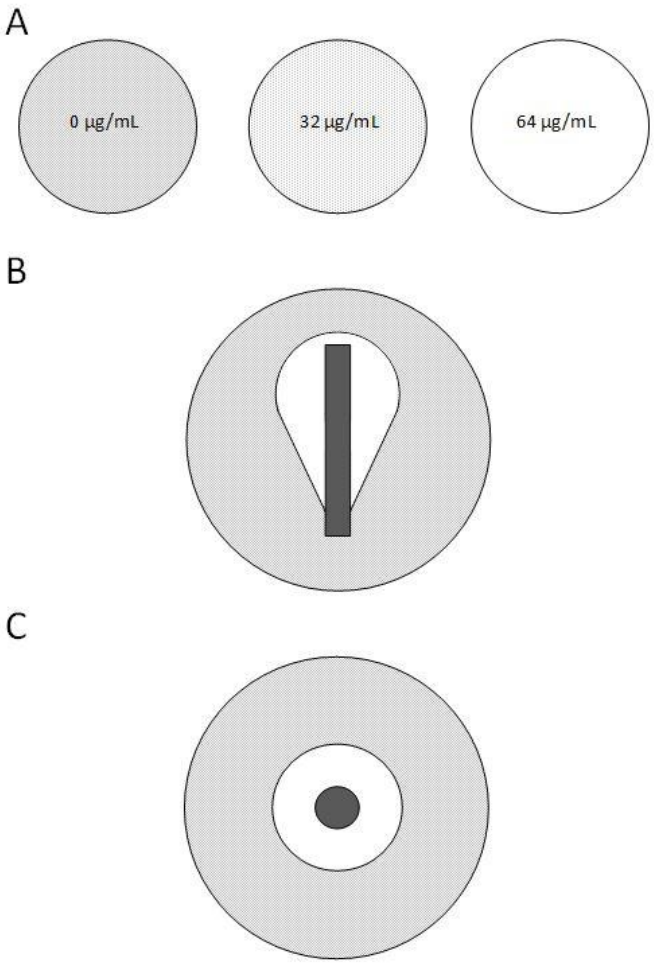

Figure 2.1: Different antibiotic susceptibility testing techniques. (A) Agar dilution, different concentrations of antibiotics added to each agar plate – examples shown; (B) E-test strips, a gradient of antibiotics along a strip that diffuse into the agar; and (C) disc diffusion, a single concentration of antibiotic that diffuses into the agar. Grey shading indicates bacterial growth, white areas indicate no growth.

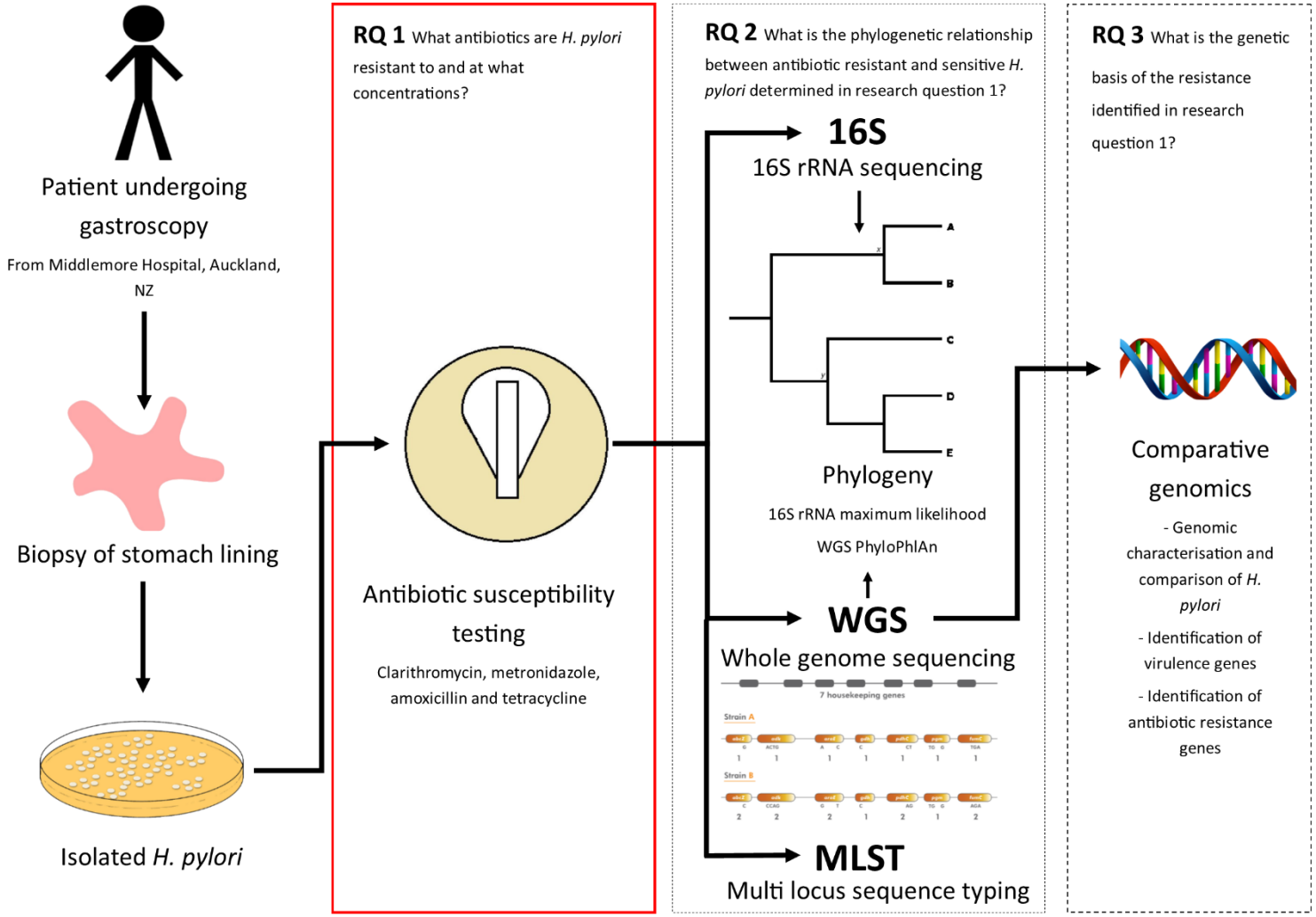

Figure 2.2: Schematic diagram of aims and research flow of this thesis. The RQ and aims of Chapter 2 are highlighted in red.

2.2 Methods and Materials

2.2.1 Ethics

No ethics approval was needed for this project. The proposal of this project was sent to the Health and Disability Ethics Committee (HDEC) board in NZ, who deemed this project not within their scope for ethics approval. It was then sent through Auckland University of Technology Ethics Committee (AUTEC) who reviewed this project and determined that it does not require an ethics review due to the research containing no patient identifiable information and no human tissue cells.

2.2.2 Acquisition, Growth and Storage of Isolates

H. pylori isolates (A-E) were isolated from stomach biopsies and provided by Dr Susan Taylor (Middlemore Hospital, Auckland, New Zealand). The isolates were stored in Laboratoire de Santé Publique du Québec (LSPQ) preservation medium until prepared for delivery. For transfer to Auckland University of Technology (AUT), isolates were plated from LSPQ medium onto supplemented Mueller Hinton (SMH) agar plates supplemented with 5 % sheep blood. Plates were incubated at 35-37 °C in microaerophilic conditions with campy GasPaks (Fort Richard, NZ) for three to seven days.

Isolates were received at AUT and sub-cultured to both liquid and solid media, as described by Blanchard and Nedrud (2012). The BHI broth (Fort Richard, NZ) was prepared following the manufacturer's instructions; 18.5 g of BHI powder was added to 500 mL of deionised water. The solution was thoroughly mixed before dispensing 10 mL into individual test tubes and autoclaving at 121 °C for 30 minutes. Columbia Sheep Blood agar (CSB) (Fort Richard, New Zealand) was used as the solid medium. Several colonies were transferred from the original plates to the BHI broth and CSB agar before being incubated in an anaerobic jar under microaerophilic conditions with campy GasPaks (Fort Richard, NZ) at 37 °C for seven to ten days. All isolates were incubated in these conditions from here on.

Glycerol stocks were prepared from the CSB agar plates; isolates were 'reverse spread plated' by adding 2 mL of 0.8 % saline to the plate and scraping off the lawn of growth before being transferred into a 1.5 mL Eppendorf tube. A 20 % volume per volume (v/v) glycerol stock was prepared using 0.6 mL of 40 % glycerol and 0.6 mL of re-suspended isolate in a 1.5 mL Eppendorf tube for each isolate. Glycerol stocks were stored at -80 °C until use.

A Gram stain was performed on each isolate for preliminary identification. From growth present on CSB agar, a colony was taken, and heat fixed onto a clean glass slide. A primary stain of crystal

violet was applied for 30 seconds before the addition of iodine for a further 30 seconds. The slide was rinsed with water, 95 % ethanol was added and then rinsed off with water. Safranin was added and left to sit for 30 seconds before being rinsed off. Slides were examined under 1000 x magnification.

2.2.3 Antibiotic Susceptibility Testing

2.2.3.1 Middlemore Hospital Testing

Antibiotic resistance was initially determined by Middlemore Hospital. Their method involved resuspending a 48-72-hour culture of *H. pylori*, grown on SMH agar, in saline to a 3 McFarland standard. The inoculum was then spread over the surface of SMH agar with 5 % sheep blood (four plates per organism). Antibiotics tested included amoxicillin, clarithromycin, metronidazole and tetracycline, all with a concentration gradient ranging from 0.016 – 256 µg/mL. One E-test strip was applied to each plate before incubating in microaerophilic conditions at 35 – 37 °C for 72 hours. Interpretations of MICs were determined against the EUCAST MIC breakpoints (http://www.eucast.org/).

2.2.3.2 In-house Testing

In-house antibiotic testing was conducted in an attempt to confirm the results provided from Middlemore Hospital. This testing was designed in two rounds with varying medias and antibiotic concentrations in an attempt to improve results (Table 2.2); using a similar method to that used by Middlemore hospital.

Isolates were prepared via spread plating with 200 µL glycerol stocks onto CSB and SMH EUCAST Agar F with 5 % horse blood and 20 mg/L nicotinamide adenine dinucleotide (NAD) (Fort Richard, NZ). Plates were incubated as described in Section 2.2.2 for seven days. From these plates, isolates were resuspended in 0.8 % saline to a McFarland standard of 3 and swabbed over the surface of new plates of both CSB and SMH EUCAST Agar F with 5% horse blood and 20 mg/L NAD (Fort Richard, NZ).

Table 2.2: In-house antibiotic susceptibility testing experimental design.

Antibiotic	Round 1		Round 2	
	CSB agar		MH EUCAST Agar F with horse blood and NAD	
	Disc or E-test strip	Concentration (µg)	Disc or E-test strip	Concentration
Amoxicillin	Disc	25	Disc	10 µg
Clarithromycin	Disc	15	E-test strip	0.016 – 256 µg/mL
Metronidazole	Disc	5	Disc	5 µg
Tetracycline	Disc	30	Disc	30 µg
Vancomycin	Disc	30	Disc	30 µg
Blank	Filter disc	-	Filter disc	-

Round one of testing involved CSB agar, a standard blood agar suitable for growing *H. pylori*. Antibiotic discs of amoxicillin (25 µg), clarithromycin (15 µg), metronidazole (5 µg), and tetracycline (30 µg) were used. Sterile filter paper dipped in water was used as a negative control and vancomycin (30 µg) was used as a positive control for resistance; due to the natural resistance exhibit towards this antibiotic by *H. pylori* (Table 2.2) (Mégraud 1997). Discs were placed on plates in two patterns: all discs on one plate and one disc per plate, a total of seven plates per isolate (Figure 2.3). All plates were incubated as described in Section 2.2.2. Only one technical replicate for each isolate was performed in this round of testing.

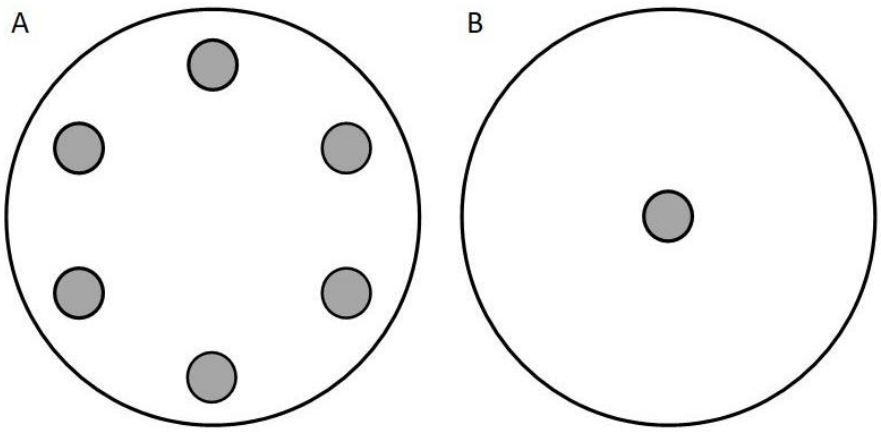

Figure 2.3: Example of disc placement for round one on antibiotic susceptibility testing. All discs present on one plate (A) and one disc per plate (B).

Although testing was not completed for round two, the following method was developed (Table 2.2). The media used was changed to MH EUCAST Agar F with 5 % horse blood and 20 mg/L NAD. Amoxicillin discs at a lower concentration were used (10 µg), as were an E-test strip for clarithromycin (0.016 – 256 µg/mL) as the desired antibiotic concentration was not available as a disc (Table 2.2). Three technical replications for each antibiotic per isolate were planned for round two. All plates were incubated as described in Section 2.2.2 for seven to ten days until growth was observed. Due to budget restraints, liquid antibiotics and/or E-test strips for all antibiotics were unable to be used to determine MICs.

When visible growth was observed, plates were removed, and the zone of inhibition for antibiotic discs was measured and recorded as the diameter of the zone around and across the centre of the disc. For *H. pylori* there are no published agreed standards for determining resistance using disc diffusion. Due to this, a literature search was conducted for previously described inhibition zone sizes conferring resistance, these zone sizes were used to determine resistance for this study (Table 2.3). If testing with E-test strips was complete, MICs were determined where the growth inhibition ellipse intersects with the MIC value indicated on the strip; EUCAST MIC guidelines would have been used to determine the breakpoints conferring resistance (http://www.eucast.org/).

Table 2.3: Published diameters of inhibition zones inferring resistance for chosen antibiotics.

Antibiotic	Concentration (µg/disc)	Inhibition zone diameters (mm)	Reference
Clarithromycin	2	No zone	(McNulty et al. 2002)
	15	≤ 30	(Mishra et al. 2006)
Metronidazole	5	< 10	(McNulty et al. 2002)
Amoxicillin	10	≤ 18	(Mishra et al. 2006)
	25	< 17	(Iovene et al. 1999; Lõivukene et al. 2002)
Tetracycline	30	≤ 30	(Mishra et al. 2006)

2.3 Results

2.3.1 Growth of *H. pylori*

Three different media were used in this study in an attempt to grow the fastidious, slow growing organism *H. pylori*; BHI broth, CSB agar and MH EUCAST Agar F with 5 % horse blood and 20 mg/L NAD. As measured by visual identification of small, translucent colonies. *H. pylori* did not grow using BHI broth but did grow on both types of agar (Figure 2.4).

Gram staining of isolates resulted in pink spirals, indicative of *H. pylori* for all isolates (data not shown). This provided preliminary confirmation of the isolates being *H. pylori*.

2.3.2 Antibiotic Susceptibility Testing

Antibiotic resistance was initially determined by Middlemore Hospital using E-test strips. In-house antibiotic testing was conducted in an attempt to confirm the provided results, primarily using the disc diffusion method.

2.3.2.1 Middlemore Hospital Antibiotic Testing

Using the E-test method, resistance profiles were established for each *H. pylori* isolate A-E (Table 2.4). These results, provided by Dr Susan Taylor and her team at Middlemore Hospital, established that isolate A is sensitive to all antibiotics tested. Isolates B and E were both resistant to both clarithromycin and metronidazole at a concentration above 256 µg/mL, while isolates C and D were resistant to only clarithromycin at concentrations of 24 µg/mL and 1 µg/mL respectively. These results are similar to published resistance data for *H. pylori* in NZ, with clarithromycin and metronidazole resistance being prominent (Hsiang et al. 2013).

Table 2.4: Antibiotic susceptibility testing results from Dr Susan Taylor, Middlemore Hospital for each isolate of *H. pylori*.

	Amoxicillin (µg/mL)	R/S	Clarithromycin (µg/mL)	R/S	Tetracycline (µg/mL)	R/S	Metronidazole (µg/mL)	R/S
A	<0.016	S	<0.016	S	0.047	S	0.5	S
B	0.047	S	>256	R	0.094	S	>256	R
C	0.032	S	24	R	0.032	S	0.25	S
D	0.032	S	1	R	0.032	S	0.25	S
E	0.023	S	>256	R	0.047	S	>256	R

R = resistant, S = sensitive.

2.3.2.2 In-house Antibiotic Testing

A McFarland standard of only two was achievable when isolates were grown on CSB agar. Although the desired McFarland standard of three was not achieved, confluent growth on the plates was observed (Figure 2.4).

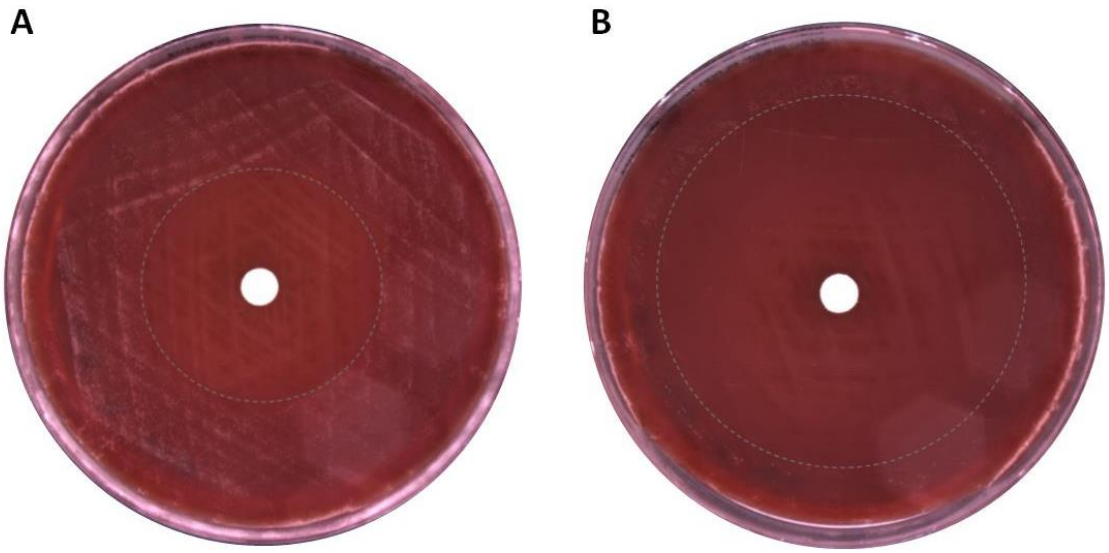

Figure 2.4: Example of antibiotic susceptibility testing on CSB agar. Dotted lines indicated where zone of inhibition starts. (A) Isolate C with clarithromycin disc (15 µg), (B) isolate C with tetracycline disc (30 µg).

Placement of antibiotic discs produced different results for in-house testing. The first round of in-house testing was conducted with all antibiotic discs on one plate, this produced no growth (data not shown) and was likely due to the overpowering effect of all the antibiotics being present at once. Using a single disc per plate yielded growth for some of the isolates (Table 2.5).

In-house testing for confirmation of Middlemore Hospital's results showed some discordant results. Middlemore Hospital results show that one isolate is sensitive to all antibiotics tested (isolate A), whereas two isolates were resistant to just clarithromycin and metronidazole (isolates B and E) and two isolates resistant to only clarithromycin (isolates C and D). Unfortunately, attempts to revive isolate D from glycerol stocks for antibiotic susceptibility testing were unsuccessful and the isolate was removed from progressing further through the study. Isolates A and B did not grow on any of the plates making confirmation with Middlemore Hospital's results unachievable. The lack of growth observed on the two control plates for these isolates indicates that the cells were not viable and not a direct result of the antibiotic concentrations being too high. In contrast, isolates C and E both produced growth on the two control plates as well as some of the antibiotic plates. For isolate E, growth was observed on both control plates (blank and vancomycin) as well as on plates containing clarithromycin and metronidazole; however, results indicate that this isolate is sensitive to both clarithromycin and

metronidazole due to the large zone of inhibition. This is contradictory with Middlemore Hospital's results that show this isolate is resistant to these two antibiotics. Growth was not present on plates with amoxicillin and tetracycline present for isolate E, suggesting that the concentration of these antibiotics may be too strong for this particular strain. Results from Middlemore Hospital show that isolate E is sensitive to amoxicillin and tetracycline as it is only able to grow in concentrations of 0.023 μg/mL and 0.047 μg/mL, respectively. Isolate C was the only isolate to exhibit growth on all plates in round one of testing. Zones of inhibition around the discs indicate that isolate is sensitive to all antibiotics tested. This result is in opposition with Middlemore Hospital in relation to clarithromycin. However, it is important to note that smaller zones of inhibition are noted for both clarithromycin and metronidazole than in isolate E.

Making a direct comparison between Middlemore Hospital results and in-house testing results is not possible due to difference in testing methods during this round of testing. However, using the disc diffusion method shows that isolates C and E are sensitive to clarithromycin and metronidazole at concentrations of 15 μg and 5 μg respectively. Sensitivity to amoxicillin and tetracycline is common amongst isolates from NZ and worldwide resistance is still low when compared to clarithromycin and metronidazole (Hsiang et al. 2013; Savoldi et al. 2018). The second round of antibiotic susceptibility testing, with different media and reduced antibiotic concentrations was unable to be completed as none of the isolates could be revived from glycerol stocks.

Table 2.5: Antibiotic susceptibility of each *H. pylori* isolate from in-house testing round 1.

Antibiotic disc concentration (μg)	Growth or zone of inhibition for each isolate			
	A	B	C	E
Clarithromycin (15)	NG	NG	40 mm z.o.i	60 mm z.o.i
Metronidazole (5)	NG	NG	20 mm z.o.i	70 mm z.o.i
Amoxicillin (25)	NG	NG	75 mm z.o.i	NG
Tetracycline (30)	NG	NG	55 mm z.o.i	NG
Vancomycin (30)	NG	NG	l.o.g	l.o.g
Blank	NG	NG	l.o.g	l.o.g

NG = no growth observed, z.o.i = zone of inhibition, l.o.g = lawn of growth.

2.4 Discussion

Due to the understudied nature of *H. pylori* antibiotic resistance in NZ, this part of the study aimed to establish the resistance levels of five *H. pylori* isolates isolated form Auckland, NZ. Antibiotic susceptibility testing is important for effectively treating an *H. pylori* infection as well as monitoring the fast-developing resistance of this species. Initial results from Dr Susan Taylor and her team at Middlemore Hospital showed that, of the five *H. pylori* isolates provided, isolate A was sensitive to all antibiotics tested, isolates B and E were resistant to >256 µg/mL of both clarithromycin and metronidazole and isolates C and D were resistant to 24 µg/mL and 1 µg/mL of clarithromycin respectively. In-house testing attempted to confirm these results. Results for in-house testing were incomplete but indicated that isolates C and E were susceptible to all clinical antibiotics tested, possibly due to antibiotic concentrations being too high. Revision of the literature led to alterations in the concentrations tested for some of the antibiotics and a second round of antibiotic susceptibility testing was planned; namely amoxicillin and clarithromycin (Table 2.2). Unfortunately, isolate D was lost before initial testing was done and subsequently the other isolates could not be revived from glycerol stocks and results were not obtained for the second round of testing. Discrepancies between the Middlemore Hospital results, and in-house testing may have been the result of the different testing methods used and lack of standardised breakpoints for disc diffusion methods, discussed below. However, it is important to note that although replicates of testing are desired to increase confidence in results, in clinical settings, antibiotic susceptibility testing is only performed once (Veses-Garcia et al. 2018). Thus, the results from Middlemore Hospital are considered reliable.

In addition to different testing methods, disparity observed between in-house testing and Middlemore Hospital may also be due to a loss or drop of resistance after storage at low temperatures. *H. pylori* has shown a loss in resistance to amoxicillin and dual resistance after long-term storage at -80 °C (Dore et al. 1999; Yakoob et al. 2001). However, the loss of amoxicillin resistance was shown to be reversible as resistance returned following plating of the organism onto agar containing an increasing gradient of the antibiotic (Dore et al. 1999). Loss of antibiotic resistance has also been observed in Staphylococcus aureus after storage at -70 °C for one year; resistance was either lost or the bacterium's tolerance was lowered towards oxacillin and cephalothin (Mayhall and Apollo 1980). The initial concentrations of antibiotics in this study may have been too high for the isolates as an initial exposure after storage at -80 °C. Utilising fresh cultures to establish resistance would be desired for repetition of this study to prevent any loss of resistance.

The rate of antibiotic resistance from Middlemore Hospital was generally similar to published studies, however the sample size of this study is small. Studies in NZ on the amount of *H. pylori* antibiotic resistance are limited, with only three studies publicly available. These three studies are dated, only sample the Auckland population and do not describe the level of resistance in MICs (Ahmed et al. 2004; Fraser et al. 1999; Hsiang et al. 2013). Therefore, a comparison of exact levels of resistance cannot be made between results obtained by Middlemore Hospital and these larger studies within NZ. Nevertheless, these studies show an increase in the rate of antibiotic resistance within NZ between 1999 and 2013 (Fraser et al. 1999; Hsiang et al. 2013).

Resistance to clarithromycin and metronidazole was expressed in results from Middlemore Hospital. Hsiang et al. (2013) have demonstrated that primary resistance of *H. pylori* to clarithromycin and metronidazole has risen in NZ since 1999, with almost half of their sample size exhibiting resistance to metronidazole (49.3 %) and 16.4 % resistant to clarithromycin, with previous rates being 32 % and 7 %, respectively (Fraser et al. 1999). A low prevalence of amoxicillin resistance (5.5 %) was noted and no tetracycline resistance was found (Hsiang et al. 2013). Dual resistance to antibiotics has been found in 12.7 % of isolates, of which clarithromycin and metronidazole dual resistance were the most common (5.5 %) (Iovene et al. 1999). These results are similar to Middlemore Hospital's results with dual resistance exhibited to clarithromycin and metronidazole (isolates B and E); however, clarithromycin resistance overall was more common (isolates B, C and E) than metronidazole resistance (isolates B and E) and no resistance was observed for amoxicillin.

Worldwide results are consistent with both Middlemore Hospital results and previous NZ based studies, in that clarithromycin and metronidazole resistance are the most common (Savoldi et al. 2018). Resistance to clarithromycin and metronidazole at an MIC of > 256 µg/mL has been noted in China (Wang et al. 2019), and resistance to clarithromycin at an MIC of 24 µg/mL has been observed in Vietnam (Quek et al. 2016). However, it is important to note that levels of resistance vary greatly worldwide and local surveillance is required for accurate comparisons of resistance MIC levels, an area that NZ is severely lacking in (Savoldi et al. 2018). Resistance to amoxicillin is still comparatively low, both in NZ and worldwide, with resistance to other first line treatment antibiotics developing at greater rates (Hsiang et al. 2013; Savoldi et al. 2018). However, it is important to monitor this resistance as amoxicillin is a crucial component to first line treatment regimens.

Although the results obtained by Middlemore Hospital are concurrent with the published literature, the small sample size of only five isolates used in this study makes it unreliable in

drawing a conclusion on the face of antibiotic resistant *H. pylori* in the whole of NZ. Increasing the sample size will aid in the representation of this issue in NZ as a whole.

The slow growth and fastidious nature of *H. pylori* made this a challenging organism to grow *in vitro*, as has been found by others (Abdalsadeg et al. 2012; Marshall and Warren 1984; Misiewicz et al. 1997). In this study both liquid broth and solid agar were tried to grow this meticulous organism. Both have been used previously to culture *H. pylori* (Blanchard and Nedrud 2012). However, liquid media is usually supplemented with a serum, such as heat-inactivated foetal bovine serum or horse serum, to increase growth of *H. pylori* (Blanchard and Nedrud 2012; Shahamat et al. 1991). The lack of a serum in the BHI broth used in this study may explain the lack of growth observed when attempting to grow *H. pylori*. While growth of *H. pylori* has been observed in serum-free liquid media, this has been at lower abundance than in broths containing serum (Shahamat et al. 1991). There is a lack in research around how sera encourage the growth of *H. pylori*, however it is suggested that they contain growth-stimulating factors due to their combined nutritional and toxin reducing effects for supporting growth (Dierikx et al. 2007; Hazell and Graham 1990). The high cost of sera was a limiting factor in this study due to monetary constraints and was therefore unable to be used during this study. There are several options for serum replacements for the growth of *H. pylori*, namely (2,6-di-methyl)-β- cyclodextrin, starch or charcoal; although growth with these alternatives is noticeably slower (Albertson et al. 1998; Buck and Smith 1987; Murray et al. 2020; Olivieri et al. 1993). These alternatives may be of further interest in future cultivation of this bacterium if large amounts of cells are required and there is ample time for slower growth.

Another associated issue with the growth of this bacterium is the loss of viability after long term storage. In this study isolate D was lost due to insufficient growth of the isolate from prepared glycerol stocks early in the study; as well as isolates A, B, C and E that were unable to be revived for the second round of antibiotic susceptibility testing. Reviving isolates from long term storage was also noted by Dr Susan Taylor who originally attempted to revive ten isolates for this study but was only successful with five (S. Taylor, e-mail message, March 19, 2019). Many studies on the cryopreservation of these isolates have demonstrated a 60-90 % recovery rate after long-term storage (Ansorg et al. 1991; Shahamat et al. 1992). Storage of isolates in refrigeration or freezing temperatures have shown a conversion of the spiral form of the bacterium into the coccoid form (Ohkusa et al. 2004). Conversion to the VBNC form (described in Section 1.3.1.1) of the bacteria may explain the difficulty in reviving the isolates from low-temperature storage conditions. Performing a Gram stain would confirm the presence of coccoid forms of the bacteria and may be a crucial step after low temperature storage of this species. Issues have also been noted with strain differentiation leading to different interactions and responses with

the long-term storage in low-temperatures (Shahamat et al. 1992). The difficulty associated with culturing this bacterium is an on-going issue in research settings. Further understanding of the advantage's sera and sera-alternatives provide in promoting the growth of this species may help elucidate a more cost and time effective option for cultivation of this bacterium. As well as identifying a specific long-term storage medium for the increased revival rate of this bacterium.

There are many different medias used for antibiotic susceptibility testing. For *H. pylori*, antibiotic susceptibility testing is generally performed on solid medium, which is employed routinely in clinical laboratories rather than liquid medium (Coudron and Stratton 1995; Mégraud and Lehours 2007). In this study, two types of solid media were used to grow *H. pylori*: CSB agar and MH EUCAST Agar F with 5 % horse blood and 20 mg/L NAD. For initial testing CSB agar was used; an adequate blood-based agar suitable for growing *H. pylori* (Blanchard and Nedrud 2012). Columbia agar is commonly used with supplementation of 5 – 10 % defibrinated sheep or horse blood to grow *H. pylori* and test antibiotic susceptibility (Blanchard and Nedrud 2012; Glupczynski et al. 1991; Smith et al. 2014). During the second round of testing that was not completed, MH EUCAST Agar F with 5 % horse blood and 20 mg/L NAD was planned to be used. This change was implemented to improve results and is similar to the media used by Middlemore Hospital. MH agar has been commonly used as the medium of choice for all antibiotic susceptibility testing (The European Committee on Antimicrobial Susceptibility Testing 2020). MH medium contains beef extract and casein hydrolysate, providing nutrients for the bacteria to grow, and the addition of starch absorbs toxic compounds (Åhman et al. 2020). The addition of 5 % horse blood agar and 20 mg/L NAD allows this medium to be used with many fastidious organisms, including *H. pylori* (Matuschek et al. 2014). MH agar is considered the go-to medium as it is a specialised medium for antibiotic susceptibility testing and most guidelines use this method; however, CSB agar is considered an acceptable alternative for testing *H. pylori* (McNulty et al. 2002). Although a direct comparison between different agars used for antibiotic susceptibility testing could not be made in this study, switching to the more widely used and specialised MH based media may have improved growth and results for identifying resistance.

The disparity of inoculum size required for antibiotic susceptibility testing of *H. pylori*, briefly discussed in Section 2.1, contributes to the lack of defined standards and procedures around antibiotic susceptibility testing for this defiant bacterium. The McFarland standard method relies on the user making an accurate visual comparison of their bacterial suspension against a recently vortexed set of standards and is, thus, subjective (Hombach et al. 2015). The inability to reach the desired McFarland standard in this study could be the result of the variability observed when making a visual comparison. In addition, the confluent growth observed on the plates, despite a lower McFarland standard, may have been due to the discrepancies observed between

McFarland standards and colony forming units (CFU) per millilitre. Hartzen et al. (1997) evaluated a 3 McFarland standard for *H. pylori*; 5 x 10^6 CFU/mL as well as 1 x 10^9 CFU/mL were both counted, showing a variation in potential inoculum size. Issues with inoculum size, for *H. pylori*, and resulting MICs, usually relate to an inoculum that is too small, 10^4 CFU/spot, and is unable to grow (Hartzen et al. 1997). When Iovene et al. (1999) increased the inoculum size of *H. pylori* from 10^4 to 10^9 CFU/mL, no changes in MIC were observed using both E-test and disc diffusion methods, however, full results are not given. In contrast, the increase of inoculum size from 10^3 to 10^7 CFU/spot resulted in an increase in MICs for *H. pylori* isolates utilising agar dilution methods (Berger et al. 1993; Hartzen et al. 1997). For other Gram-negative bacteria such as *E. coli* and *P. aeruginosa*, an increase in inoculum size has been shown to result in an increase in final MIC (Bulitta et al. 2010; Ferran et al. 2007; Tilton et al. 1973). In spite of the use of McFarland standards being subjective, they are still routinely used in antibiotic susceptibility testing practices (The European Committee on Antimicrobial Susceptibility Testing 2020). The inoculum effect is an important factor to consider when trying to accurately determine MICs. This effect does not appear to be well studied in *H. pylori* and is an area of further interest to improve the standards around antibiotic susceptibility testing of *H. pylori* to reduce the chance of inaccurate MICs.

There is currently no global standard method for the delivery of antibiotics in antibiotic susceptibility testing of *H. pylori* that is easily applicable to routine testing. All three methods, agar dilution, E-test, and disc diffusion, have been used to test antibiotic susceptibility in *H. pylori* isolates (Ogata et al. 2015). CLSI have defined agar dilution as the gold standard method for *H. pylori*, which is used as the reference to check the validity of other methods (Clinical and Laboratory Standards Institute 2015 cited in: Miftahussurur et al. 2020). However, this method is not routinely used in clinical laboratory settings due to its high labour, time and monetary costs (Valdivieso-García et al. 2009). As a result of the high costs associated with this reference standard, other methods have emerged. Researchers have agreed that the E-test method produces similar results to agar dilution with few discrepancies and is currently the superior method for initial recognition of resistance for levofloxacin, tetracycline, and clarithromycin (Miftahussurur et al. 2020; Ogata et al. 2015; Shakir et al. 2018). Metronidazole resistance is often overestimated when using the E-test method and often has to be re-tested with other methods for confirmation (Glupczynski et al. 2002; Hachem et al. 1996; Miftahussurur et al. 2020; Osato et al. 2001). Due to the low prevalence of amoxicillin resistance identified in *H. pylori*, establishing resistance using E-test strips has low sensitivity (Miftahussurur et al. 2020). The other available diffusion method is disc diffusion, the simplest method used for routine antibiotic susceptibility testing (Mégraud and Lehours 2007). The use of this method with *H.*

pylori is somewhat controversial due to its difficulties with fastidious and slow growing bacteria, its qualitative interpretation and its inaccuracies (King 2001; Mégraud and Lehours 2007; Ogata et al. 2015). Comparability of the disc diffusion method with other methods is controversial, with some studies presenting similar results between methods and others showing wide discrepancies and errors (Mishra et al. 2006; Ogata et al. 2015; Tang et al. 2020). Nonetheless, this method is economically beneficial, easy to perform for routine analysis and is currently used in *H. pylori* research (Akhtereeva et al. 2018; Alarcón-Millán et al. 2016; Ogata et al. 2015). While initial tests were done with E-test strips, this method could not be employed with in-house confirmatory testing. Due to monetary restraints the more economical option of disc diffusion was used.

Issues with the delivery of antibiotics are intrinsically linked to the breakpoints associated with each method. A lack in defined or standardised breakpoints to determine resistance are another limitation of antibiotic susceptibility testing for *H. pylori*. MIC breakpoints used to distinguish resistance and susceptibility in *H. pylori* are only defined for clarithromycin; with them being susceptible < 0.25 µg/mL, intermediate 0.25 – 1 µg/mL, and resistant > 1 µg/mL (Clinical and Laboratory Standards Institute, 2015 cited in: Arslan et al. 2017). For other antibiotics there are no standardised breakpoints defined and MICs are chosen from published guidelines or from other literature (Ogata et al. 2015). The use of previously defined breakpoints from literature and EUCAST guidelines, have provided similar resistance rates for clarithromycin, metronidazole and tetracycline; but different rates were observed for amoxicillin (Alarcón et al. 2017). As well as producing similar results to agar dilution methods, E-test strips are a quantitative method allowing for direct expression of MICs and have been well adapted to slow growing bacteria due to their slow release of antibiotics withstanding the incubation times required for slow growing organisms, such as *H. pylori* (Cederbrant et al. 1993). This method has also shown good inter- and intra-laboratory correlations of results for clarithromycin and amoxicillin (Glupczynski et al. 2002). A disadvantage of these strips is the high cost when compared to discs and the issues of resolving resistance for amoxicillin and metronidazole (Mishra et al. 2006; Reller et al. 2009).

Disc diffusion is a qualitative method for determining resistance (Ogata et al. 2015). This method also has no international standardisation of breakpoints and has not been validated for most clinical antibiotics, except for macrolides in France (Grignon et al. 2002; Mégraud and Lehours 2007). The lack of standards around breakpoints for determining resistance makes it difficult for comparisons to be made between methods and for the accurate determination of resistance and therefore, standardised and clinically validated breakpoints are required for antibiotic susceptibility testing of *H. pylori*.

The lack of replication and confirmation of antibiotic resistance poses a limitation to this study. Antibiotic susceptibility testing is often performed after treatment failure to allow the clinician to establish a more efficient course of action (Miwa et al. 2003). This testing can be expensive and time consuming depending on the method employed (Mascellino et al. 2017). Therefore, typical antibiotic susceptibility testing performed in clinical settings is only performed once; however to reduce the chance of false diagnosis technical replicates should be increased to at least three (Veses-Garcia et al. 2018). This was unable to be completed in this study however, as previously stated, the results retrieved from Middlemore Hospital are considered accurate based on current clinical practices.

This chapter highlights the lack of approved and validated standards for antibiotic susceptibility testing of *H. pylori*. Research studies and clinical guidelines do not have agreed standards for antibiotic susceptibility testing for *H. pylori*; disparity in media used, inoculum size and delivery of antibiotics makes it difficult for direct comparisons between studies. More defined standards are needed to make this process streamlined and consistent across various studies and clinical practices.

The four isolates described in this chapter (A, B, C, and E) were selected for further analysis due to their resistance profiles to key antibiotics used in first-line treatment regimens for treating *H. pylori* infections; one of these (isolate A) was sensitive to all antibiotics tested while isolates B, C and E showed a range of resistances. Isolate D was not studied any further as it did not survive the experimental process. While the in-house testing proved inconsistent with the results from Middlemore Hospital, this was most likely due to varying antibiotic testing methodologies and reduced viability from continual sub-culturing of *H. pylori*. Although technical replicates are desired to increase confidence in results, especially for metronidazole resistance, current clinical testing is only performed once and thus the results provided by Middlemore Hospital were considered reliable. These isolates were analysed further for potential resistance mechanisms to these antibiotics. Initial genome analyses of these isolates are described in Chapter 3.

Chapter 3 Phylogenetic Relationship of *H. pylori* isolates

3.1 Introduction

There is a strong association between antibiotic resistance and treatment failure of *H. pylori* infections (Savoldi et al. 2018). To begin to understand the underlying mechanisms of resistance and how they arose, a phylogenetic method can be employed. Performing a phylogenetic analysis can help discern the patterns and processes of evolution, as well as help identify closest relatives that can be used for comparative analyses (Hollister et al. 2015; Soltis and Soltis 2003). In Chapter 2 the antibiotic resistance phenotypes of isolates A, B, C, D and E were established. Isolates B and E were resistant to > 256 µg/mL of clarithromycin and metronidazole, isolate C was resistant to 24 µg/mL of clarithromycin, while isolate A isolate was susceptible to all antibiotics tested. Since isolate D did not survive the experimental process it was not studied further. Understanding the phylogenetic relationships of these isolates is an important first step for evaluating the underlying mechanisms of resistance for isolates A, B, C and E.

In clinical settings, it is important to determine the pathogen responsible for a particular disease or infection to help generate an accurate treatment plan (Bertelli and Greub 2013). Identification of clinically relevant bacterial pathogens is traditionally done using standard culture-based techniques such as Gram staining, biochemical tests, growth time and morphology (Woo et al. 2008). However, these techniques are unusable for bacteria that have undefined or complex biochemical characteristics or are fastidious and/or slow growing (Bosshard et al. 2003; Srinivasan et al. 2015; Woo et al. 2003). For the clinical identification of *H. pylori* current guidelines recommend invasive (gastric biopsy, histology and bacterial culture) and non-invasive tests (urea breath test, stool antigen assay and serology); with each method having different sensitivities (Chey and Wong 2007; Malfertheiner et al. 2017). Although these methods provide the identification of *H. pylori*, they do not provide any information on the antibiotic resistance phenotypes of the strains. To overcome the challenges of traditional culture-based techniques and the lack of information about antibiotic resistance, genomic based techniques may be utilised. Molecular based techniques have been utilised in clinical settings for identifying and typing bacteria into closely related groups; this may be achieved using 16S rRNA sequencing, MLST and WGS (Enright et al. 2000; Patel 2001; Salipante et al. 2015). WGS is more commonly used in research settings due to its high costs, making it too costly for routine clinical analysis. However, utilising this approach can greatly increase the amount of information obtained for a single isolate, which may be used for surveillance of antibiotic resistance, identification of novel resistance mechanisms and the creation of novel antibiotics (Binh et al. 2015; Chewapreecha et al. 2014; Doostparast Torshizi and Wang 2018).

The ubiquity of the 16S rRNA gene amongst bacteria, as well as areas of high and low conservation within this gene, have characterised it as a useful tool for taxonomic and phylogenetic classification of bacteria (Woese 1987; Woese and Fox 1977). As previously discussed (Section 1.3.6.1), the 16S rRNA gene is a valuable tool in clinical settings for fast and accurate identification of pathogens at the genus-, and potentially, species- and strain-level (Johnson et al. 2019; Patel 2001). In clinical settings, this method of identification has been employed with *H. pylori* for identification from stomach biopsies and stool samples (Gramley et al. 1999; Szymczak et al. 2020).

As well as identifying the bacterial pathogen, typing the bacterium is also an important step. Typing identifies different strains or sequence types (ST) within a species (Sabat et al. 2013). Identifying specific strains or STs is an important tool used in epidemiological studies to identify epidemically important strains and to help track and control outbreaks (Hammerum et al. 2015; Ye et al. 2011). There are many different typing methods available, with original methods focusing on phenotypic differences and more recent methods focusing on molecular differences (Destro et al. 1996; Kapperud et al. 1990; Maiden et al. 1998; Sørensen and Larsen 1986).

MLST is a molecular typing method used to systematically identify and classify strains of bacteria for molecular epidemiology. It is commonly used in clinical settings to identify and track infectious strains (Bougnoux et al. 2002; Maiden et al. 1998; Nukui et al. 2020). For MLST analysis, seven or eight housekeeping gene loci are used and the allelic profiles of internal gene fragments are determined based on their nucleotide sequences (Maiden et al. 1998). Unique allele sequences are assigned a number and combinations of these allele profiles and associated numbers (from each gene) create an overall ST for an isolate (Maiden et al. 1998). New numbers are assigned to each new allele that is discovered. The clear nature of the standardised nomenclature, as well as the reproducibility and ability for worldwide comparisons using internet-based databases makes this a desirable method for molecular based typing (Maiden et al. 1998). As described in the Section 1.3.2.1, MLST analysis of *H. pylori* uses seven housekeeping genes: *atpA*, *efp*, *trpC*, *ppa*, *mutY*, *yphC* and *urel*, as well as sometimes using the virulence gene *vacA* (Achtman et al. 1999). MLST analysis has primarily been used in *H. pylori* as a tool to track human migrations and to help understand the phylogeography of this expansive organism (Linz et al. 2007; Montano et al. 2015).

Understanding the phylogeny of bacteria and their genes is important to help observe their evolution. This chapter focusses on the second research question of this study (Figure 3.1), intending to understand the phylogeny of the isolates from this study, compared to other

published *H. pylori* isolates, as well as identifying the closest relatives of isolates A, B, C and E for further comparative analyses for antibiotic resistance mechanisms.

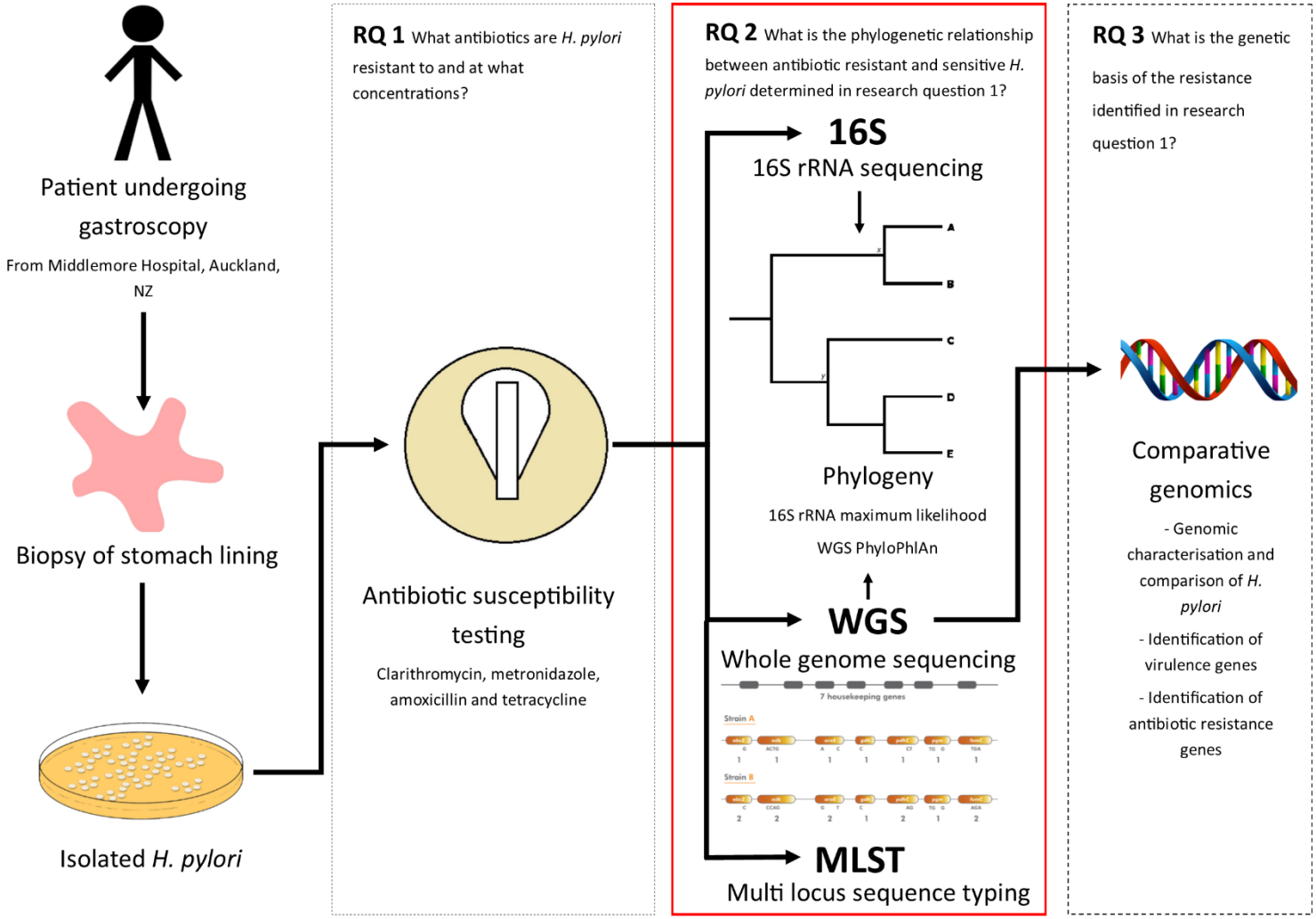

Figure 3.1: Schematic diagram of aims and research flow of this thesis. The RQ and aims of Chapter 3 are highlighted in red.

3.2 Methods and Materials

3.2.1 16S rRNA gene sequencing

3.2.1.1 DNA Extraction

Genomic DNA (gDNA) was prepared by reverse spread plating one plate of each isolate of *H. pylori* from CSB agar (Fort Richard, New Zealand) with 2 mL sterile water, followed by boiling at 100 °C for 10 minutes using an Allsheng Incubator MiniT-100 heating block (Zhejiang, China). *E. coli* was grown in nutrient broth for 24 hours before the gDNA was extracted at the same time to be used as the positive control in the PCR.

3.2.1.2 PCR Reaction

Amplification of the 16S rRNA gene region of each *H. pylori* isolate was performed using the universal bacterial forward primer 27F (5'AGAGTTTGATCMTGGCTCAG3') and the universal reverse primer 1492R (5'CGGTTACCTTGTTACGACTT3'). PCR components for each 25 µL reaction were 12.5 µL GoTaq Green Master Mix (Promega, Wisconsin, USA), 1 µL of 10 µM 27F, 1 µL of 10 µM 1492R, 2.5 µL gDNA template and 8 µL nuclease free water. *E. coli* gDNA was used as the positive control. Nuclease-free water (2.5 µL) was used instead of a DNA template for the no template control (NTC).

PCRs were carried out on a Techne TC-512 Gradient Thermal Cycler (Bibby Scientific™ Techne™, England) using the following conditions; initial denaturation of 5 minutes at 95 °C, 30 cycles consisting of 30 seconds at 95 °C (denaturation), 30 seconds at 51 °C (annealing) and 2 minutes at 72 °C (extension); followed by 10 minutes at 72 °C for final extension.

3.2.1.3 Agarose Gel Electrophoresis

To determine the size of the PCR product, a 1 % agarose/1 X Tris-Borate EDTA (TBE) gel, stained with 0.5 µL of 10 mg/mL ethidium bromide, was electrophoresed at 70 V for 50 minutes. Each lane contained 5 µL of PCR product or 0.1 µg/µL 100 bp DNA ladder (Solis BioDyne, Estonia). Gels were viewed and photographed under ultraviolet (UV) light using an AlphaImager® HP (Alpha Innotech, California, USA).

3.2.1.4 Sanger Sequencing

Sanger sequencing of the PCR products was performed by Macrogen Inc. (Seoul, South Korea) using the universal 27F and 1492R primer pair. Forward and reverse sequences were trimmed and aligned to create a consensus sequence. Consensus sequences were used to search GenBank on NCBI using BLASTn (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Preliminary identification of isolates was based on the top hit with the lowest E-value.

3.2.2 Whole Genome Sequencing

3.2.2.1 Genomic DNA Extraction

CTAB Extraction

Genomic DNA was extracted from isolates A, B, C and E using a method previously described by Minas et al. (2011). Isolates were grown for seven days on six plates of CSB agar (Fort Richard, New Zealand) per isolate. Cells were collected by reverse spread plate method (Section 2.2.2) with 2 mL of 0.85 % saline. Cells were washed twice with 20 mL of pH 7.2 phosphate-buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.2) and centrifuged at room temperature at 7,745 x gravity (*g*) (Eppendorf 5430R Centrifuge, New South Wales, Australia) for 10 minutes. The supernatant was discarded, and the cell pellet resuspended in 270 µL of PBS.

After washing the cells, the samples were prepared for DNA extraction. Two methods for cell lysis were compared, bead beating and boiling. For bead beating, the resuspended cell pellet was transferred to a tube containing 1 g of 0.1 mm silica-zirconia beads and homogenised at 4 meters per second (m/s) for 10 seconds using the FastPrep®-24 (MP Biomedicals, Ohio, USA). For the boiling method, the resuspended cell pellet was heated to 100 °C for 10 minutes using an Allsheng Incubator MiniT-100 heating block (Zhejiang, China). After cell lysis, 270 µL of sodium dodecyl sulphate (SDS) buffer (100 mM NaCl, 500 mM Tris pH 8.0, 10 % SDS) was added to each tube and samples were centrifuged at 19,090 x *g* for 3 minutes at room temperature.

RNase A (100 mg/mL) and proteinase K (20 mg/mL) were initially added at the recommended volume of 1 µL per enzyme per tube (EX 1, Table 3.1). However, to increase the quality of the extraction and to assess the effect of RNase A and proteinase K on the samples, the mixture and volume of these enzymes was adjusted in a presence/absence experiment (EX2 – EX7, Table 3.1). Tubes were then mixed by inversion before being left to sit for 10 minutes at room temperature.

Table 3.1: Addition of RNase A (100 mg/mL) and proteinase K (20 mg/mL) in CTAB extraction protocol for improved quality of gDNA.

	EX 1	EX 2	EX 3	EX 4	EX 5	EX 6	EX 7
RNase A	1 µl	0 µl	1 µl	0 µl	0 µl	2 µl	2 µl
Proteinase K	1 µl	0 µl	0 µl	1 µl	2 µl	0 µl	2 µl

EX = experiment.

A total of 180 µL of cetyltrimethylammonium bromide-polyvinylpyrrolidone (CTAB) extraction buffer (100 mM Tis-HCl, 1.4 M NaCl, 20 mM EDTA, 2% CTAB, 1% polyvinylpyrrolidone and 0.4% β-mecaptoethanol) was added to each sample and incubated at 60 °C with 300 revolutions per minute (rpm) in a SI-300R Shaking Incubator (GMI, Minnesota, USA) for 30 minutes. Samples were then centrifuged at 19,090 x *g* for 1 minute before 350 µL of chloroform:isoamyl alcohol (24:1) was added. Samples were then mixed by inversion before being centrifuged again at 19,090 x *g* for 5 minutes at room temperature. The upper aqueous layer was transferred to a new microcentrifuge tube and another 500 µL of chloroform:isoamyl alcohol (24:1) was added, mixed by inversion and centrifuged at 19,090 x *g* for 5 minutes. The upper aqueous layer was transferred to a new microcentrifuge tube and 10 M ammonium acetate was added for a final concentration of 2.5 M. Samples were mixed by inversion and centrifuged at 19,090 x *g*. The upper aqueous layer transferred to a new tube and 0.54 x sample volume of isopropyl alcohol was added. The samples were mixed by inversion before being stored at -20 °C for 48 hours. Samples were then centrifuged at 19,090 x *g* for 10 minutes at 4 °C. The supernatant was removed, and the DNA pellet washed with 1 mL of 70 % ethanol and centrifuged at 19,090 x *g* for 10 minutes at 4 °C. The ethanol was removed using a Concentrator plus (Eppendorf, Hamburg, Germany) at 60 °C for 4 minutes at 18,506 x *g*. The DNA pellet was resuspended in 30 µL sterile DNase free water (ThermoFisher Scientific, Massachusetts, USA) and stored at -20 °C until required.

Commercial Kit Extraction

The commercial DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) was used to extract gDNA following the manufacturer's instructions. Isolates were grown on CSB agar for one week before collecting the cells using the reverse spread plating method (Section 2.2.2) with 2 mL of 0.85 % saline. Suspended cells were then transferred to a microcentrifuge tube for centrifuging at 5000 x *g* for 10 minutes at room temperature. The supernatant was discarded, and the pellet resuspended in 180 µL Buffer ATL and 20 µL of proteinase K was added. The samples were vortexed thoroughly and incubated at 56 °C using an Allsheng Incubator MiniT-100 heating block (Zhejiang, China) for 1 hour with periodic mixing. Samples then had 4 µL of RNase A (100 mg/mL) added and were incubated at room temperature for 2 minutes. The samples were vortexed for 15 seconds before addition of 200 µL of Buffer AL and 200 µL of 96 - 100 % ethanol and thoroughly vortexed again. The solution was transferred to the DNeasy Mini spin column and collection tube and centrifuged at 6000 x *g* for 1 minute at room temperature. The column was transferred to a new collection tube and 500 µL of Buffer AW1 was added and centrifuged at 6000 x *g* for 1 minute at room temperature. A new collection tube was used and 500 µL of Buffer AW2 was added and centrifuged at 20,000 x *g* for 3 minutes at room temperature. For collection

of the extracted DNA, user provided microcentrifuge tubes were used; these were washed in 100 % bleach, 100 % ethanol and milli-Q water to remove any DNases present and were dried before use. Columns were placed in the cleaned microcentrifuge tubes and eluted with 50 µL of Buffer AE. The samples were incubated at room temperature for three minutes before being centrifuged at 6,000 x *g* for 1 minute at room temperature. A second elution was carried out using another 50 µL of Buffer AE, incubation at room temperature for three minutes and a final centrifugation at 6,000 x *g* for 1 minute at room temperature.

The gDNA was ethanol precipitated by adding. 0.1 x sample volume of 3 M sodium acetate (3 M NaOAc pH 5.2) and 2 x sample volume of ice cold 100 % ethanol. Samples were vortexed before placing at -20 °C for a minimum of 24 hours until required.

3.2.2.2 Quality and Quantity of gDNA Extractions

Quality of gDNA samples were assessed by measuring A_{260}/A_{280} and A_{260}/A_{230} ratios on a GE NanoVue spectrophotometer (GE Biosciences, New Jersey, USA) using default settings and 2 µL of Millipore water as a blank. Quality and contamination were also measured by electrophoresis with a 0.8 % agarose/1 X TBE gel, stained with 0.5 µL of 10 mg/mL ethidium bromide at 70 V for 60 minutes. All gels included 5 µL 0.21 – 21 kbp molecular weight marker (Sigma Aldrich, Missouri, USA) and 5 µL of gDNA plus 1 µL of loading dye. DNA concentration was also measured using the Qubit® 2.0 Fluorometer (ThermoFisher Scientific, Massachusetts, USA) by following the manufacturer's instructions.

3.2.2.3 Gel Purification

Following the CTAB extraction of gDNA, gel purification was performed using the GeneJET Gel Extraction Kit (ThermoFisher Scientific, Massachusetts, USA). gDNA was electrophoresed and visualised as previously described (Section 3.2.2.2). The gDNA band was excised using a sterile scalpel. The gel slice was placed in a clean tube and a 1:1 volume of binding buffer was added before incubation at 60 °C for 10 minutes with periodic inversions until the gel was dissolved and a yellow colour was seen. The solution was vortexed, and water was added at a 1:1 ratio with the solution before being thoroughly vortexed again. The solution was then transferred to a GeneJET purification column and centrifuged at 12,000 x *g* (Eppendorf 5430R Centrifuge, New South Wales, Australia) for 1 minute at room temperature. The flow-through was discarded, 100 µL of binding buffer was added to the column, and the solution was centrifuged at 12,000 x *g* for 1 minute at room temperature. The flow-through was discarded again, 700 µL of wash buffer was added to the column and then centrifuged at 12,000 x *g* for 1 minute at room temperature. The flow-through was removed and the column was spun again at 12,000 x *g* for 1 minute at room temperature to remove excess wash buffer. The column was transferred to a fresh

collection tube and 25 µL of elution buffer was added and centrifuged at 12,000 x *g* for 1 minute at room temperature. The purified gDNA was stored at -20 °C until used.

The purified gDNA was checked by running a 0.8 % agarose gel/1 X TBE as described previously (Section 3.2.2.2). Each lane had either 5 µL of purified gDNA and 1 µL loading dye or 5 µL of 0.21 – 21 kbp molecular weight marker (Sigma Aldrich, Missouri, USA). The gel was electrophoresed at 70 V for 60 minutes.

3.2.2.4 Genome Sequencing

gDNA was sequenced by GeneWiz (Shuzhou, China) using Illumina sequencing on an Illumina HiSeq™ 2500 sequencer. The GeneWiz workflow is as follows; library preparation was constructed using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (New England Biolabs, MA, USA), following the manufacturer's instructions. Each sample had 1 µg of genomic DNA randomly fragmented into < 500 bp by sonication using the S220 (Covaris, MA, USA). Fragments were treated with End Prep Enzyme Mix for end repairing, 5' phosphorylation and dA-tailing in one reaction; followed by a T-A ligation to add adaptors to both ends. Adaptor-ligated DNA then underwent size selection using AxyPrep Mag PCR Clean-up (Axygen, NY, USA), and fragments of approximately 410 bp (with an insert size of approximately 350 bp) were recovered. Each sample was then amplified using PCR for eight cycles using P5 and P7 primers, with both primers consisting of sequences that can anneal with flow cells to perform bridge PCR. The primer P7 also carries a six-base index to allow for multiplexing. The PCR products were then cleaned up using AxyPrep Mag PCR Clean-up (Axygen, NY, USA), validated using an Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA), and quantified using a Qubit® 2.0 Fluorometer (Invitrogen, CA, USA). Libraries with variable indexes were multiplexed before being loaded onto an Illumina HiSeq™ 2500 instrument following the manufacturer's instructions (Illumina, CA, USA). Sequencing was performed using a 2 x 150 PE configuration, and image analysis and base calling were completed using the HiSeq Control Software (HCS) + Real Time Analysis (RTA) 2.7 (Illumina, CA, USA) on the HiSeq instrument.

Quality control of raw reads was achieved by demultiplexing using bcl2fastq 2.17 conversion software (Illumina, CA, USA) and discarding reads if they met the following criteria; reads contaminated with adapter sequences, reads with an N base content of more than 10 %, and reads where the ratio of low quality bases (Q < 20) was more than 0.5. Additional trimming of raw reads was completed in-house using AdapterRemoval v2.3.0 (Schubert et al. 2016) to remove known adaptor sequences from PE reads.

3.2.3 Genome Assembly

Genome assembly was completed in-house by assembling the filtered reads into contigs using SPAdes v3.13.2 (Bankevich et al. 2012) and Unicycler v0.4.8 (Wick et al. 2017) using default settings. Assembled genomes were visualised using Bandage (Wick et al. 2015). Single, circular contigs, identified after assembly and presumed to be plasmids, were searched using the NCBI BLASTn suite (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to identify any similarity to other known *H. pylori* plasmids.

Nullarbor v2.0.20191013 was used as an all-encompassing software on the raw reads to trim reads, identify the species, perform a *de novo* assembly, annotate and construct an MLST, resistome, virulome and variance classification for each isolate (Seemann et al. n.d.). Specific programmes used for each process are listed in Table 1.5. This software was utilised to create a full report of the items discussed above for each isolate and compared to traditional analysis. Within this chapter, only the assembly, annotation and MLST analysis are discussed; the resistome and virulome are discussed in Chapter 4.

3.2.4 Phylogenetic Analysis

3.2.4.1 DNA Sequence Datasets

16S rRNA Dataset

A phylogenetic tree based on the 16S rRNA gene sequence was used for preliminary genotypic identification of the four *H. pylori* isolates. For isolates A, B, and C the consensus sequences of the 16S rRNA genes from Sanger sequencing were used in the dataset; the lengths of each is given in Table 3.2. For isolate E only the last 1,079 bp were used in the dataset due to poor sequencing from the forward primer. A search within the Ribosomal Database Project (RDP) v11.5 (http://rdp.cme.msu.edu/) for *H. pylori* 16S rRNA sequences revealed 18 distinct sequences used in this study (Cole et al. 2014). *Helicobacter acinonychis* and *Helicobacter cetorum* were used as the two outgroups for this dataset. The dataset can be viewed in Table 3.3.

Table 3.2: Length of 16S rRNA gene for each isolate from Sanger sequencing used in phylogenetic tree construction.

Isolate	Length of 16S rRNA from Sanger sequencing (bp)
A	1,320
B	1,356
C	1,356
E	1,079

Table 3.3: Dataset for Sanger 16S rRNA phylogeny from the RDP.

Strain name	Accession number	Reference
Isolate A	SAMN17734761*	This study
Isolate B	SAMN17734762*	This study
Isolate C	SAMN17734763*	This study
Isolate E	SAMN17734764*	This study
H. pylori BHHPSKP 69	KC525432	Unpublished (Patel, S. K., Nath, G. and Jain, A. K.)
H. pylori RIGLD HPD22	JN162414	Unpublished (Yadegar, A., Alebouyeh, M., Zali, M. R., Nazem Alhosseini, E., Mirzaei, T. and Shokrzadeh, L.
H. pylori 0021	EF217410	Unpublished (Ye, S.)
H. pylori 85D08	U00679	(Drazek et al. 1994)
H. pylori ATCC 49396	AF348617	Unpublished (Dewhirst, F. E., Paster, B. J. and Fox, J. G.)
H. pylori DM68A	DQ202380	(Khan et al. 2008)
H. pylori DM5A	DQ202378	(Khan et al. 2008)
H. pylori DM20B10	DQ202382	(Khan et al. 2008)
H. pylori ATCC49396	AF363064	Unpublished (Melito, P. L., Munro, C. K., Chipman, P. R., Woodward, D. L., Booth, T. F. and Rodgers, F. G.)
H. pylori DH195	DQ202375	(Khan et al. 2008)
H. pylori SJM180	CP002073	Unpublished (Kersulyte, D., Velapatino, B., Gilman, R. H. and Berg, D. E.)
H. pylori 0035	EF217412	Unpublished (Ye, S.)
H. pylori Lit69	AF535196	(Dailidiene et al. 2002)
H. pylori G27	CP001173	(Baltrus et al. 2009)
H. pylori BM012S	CP006889	(Linz et al. 2013)
H. pylori BM012A	CP006888	(Linz et al. 2013)
H. cetorum MIT 99-5656	AF292378	(Harper et al. 2000)
H. acinonychis Eaton 90-119-3	M88148	Unpublished (Eaton, K. A., Radin, M. J., Albrandt, K. G., Paster, B. J., Dewhirst, F. E., Krakowa, S. and Morgan, D. R.)

* = NCBI accession number

Whole Genome Sequence Dataset

To identify the whole genome relationships between isolates A, B, C and E other *H.* pylori isolates, a dataset was created consisting of complete *H. pylori* genomes. A search of the NCBI genome database (https://www.ncbi.nlm.nih.gov/genome) for complete *H. pylori* genomes returned 176 entries. From this list, 21 duplicate sequences were identified by PhyloPhlAn2 and were subsequently removed, reducing the dataset to 155 complete *H. pylori* genomes. These genomes were used for comparison with isolates A, B, C and E and the complete genome of *H. acinonychis* strain Sheeba (NC_008229.1) was used as the outgroup for this dataset. The dataset may be viewed in Appendix A: Whole Genome Sequence Data Set.

3.2.5 16S rRNA Phylogenetic Tree Construction

Maximum likelihood analysis of 16S rRNA sequences were initially carried out using the 16S rRNA sequences generated by Sanger sequencing as described in Section 3.2.1. A second analysis was done using 16S rRNA sequence extracted from the whole genome Illumina sequencing of isolates A, B, C and E described in Section 3.2.2. A comparison of data from the Sanger sequencing and Illumina sequencing was carried out to ensure the sequences from each isolate was consistent in both methods. The full length 16S rRNA gene was identified and extracted from the Illumina sequencing data through annotation of the genomes (described in Section 4.2.2).

3.2.5.1 Maximum Likelihood Analysis of 16S rRNA Sequences

A multiple sequence alignment of the dataset sequences was performed using the inbuilt MUSCLE algorithm v3.8.425 (Edgar 2004) integrated within Geneious v2020.0.4 (https://www.geneious.com). Default parameters were selected.

A model test was performed using the inbuilt function within the MEGAX software v10.0.5 (Kumar et al. 2018) to establish the most appropriate model to apply to the phylogenetic analyses. For the Sanger sequence dataset, this model was the Hasegawa, Kishino and Yano 1985 (HKY85) model, while for the Illumina sequence dataset, it was the Kimura, 80 (K80) model. This was based on their having the lowest Bayesian information criterion (BIC) scores of 2362.11 and 10865.19, respectively. As most *H. pylori* isolates carry multiple copies of the 16S rRNA gene, which are usually identical, only one representative 16S rRNA gene for each isolate from the Illumina sequence dataset was included in this analysis.

From the alignments, maximum likelihood trees were calculated using the PhyML (Guindon et al. 2010) v2.2.4 plugin for Geneious v2020.0.4 with either the HKY85 or K80 model, as

appropriate. Statistical analysis of the trees was carried out using 1000 and 500 bootstrap replications for the Sanger and Illumina data, respectively.

3.2.6 Multilocus Sequence Typing

As described above (Section 3.2.3), part of the report Nullarbor produces is an MLST analysis. This analysis was completed in Nullarbor using MLST (https://github.com/tseemann/mlst) and the pubMLST database (Jolley et al. 2018). Isolates A, B, C, and E were also run manually through the *H. pylori* MLST database (https://pubmlst.org/helicobacter/), which is part of the pubMLST database (Jolley et al. 2018), by uploading the DNA FASTA files of each genome for comparison against the database.

A phylogenetic tree was created based on the MLST alleles (ATP synthase, F1 alpha (*atpA*), elongation factor EF-P (*efp*), A/G-specific adenine glycosylase (*mutY*), Inorganic pyrophosphatase (*ppa*), anthranilate isomerase (*trpC*), urease accessory protein (*urel*) and GTPase (*yphC*) and vacuolating toxin (*vacA*)) and visualised using the Interactive Tree of Life (iTOL) plugin integrated within the pubMLST database (Jolley et al. 2018; Letunic and Bork 2016). Concatenated and aligned nucleotide sequences of the MLST alleles from 834 published isolates and the four isolates from this study were included (Breurec et al. 2011; Falush et al. 2003; Liao et al. 2009; Linz et al. 2007; Panayotopoulou et al. 2010; Wirth et al. 2004). The final tree was viewed and edited on the iTOL website (https://itol.embl.de/) where the continent of isolation for each isolate was overlaid.

3.2.7 PhyloPhlAn2

The whole genome dataset described in Section 3.2.4.1 and Appendix A: Whole Genome Sequence Data Set, was analysed using PhyloPhlAn2 (Segata et al. 2013) to create phylogenetic trees based on the derived sequences of 400 conserved proteins and their genes. PhyloPhlAn2 v0.40 was used with the supermatrix (concatenation) approach with rapid bootstrapping for both nucleotide and amino acid trees. The GTRCAT model and 100 bootstrap replicates were used for the nucleotide tree while the PROTCATLG model with 500 bootstrap replicates was used for the protein sequence tree. Both trees were constructed using RAxML v8.2.12 and edited in FigTree v1.4.4 to produce the final figures.

3.3 Results

3.3.1 16S rRNA PCR Amplification

Preliminary genomic identification of the new *H. pylori* isolates was done using PCR and partial sequencing of the 16S rRNA gene. The expected product of the 16S rRNA, approximately 1,500 bp long, was amplified for each isolate; an example of an agarose gel showing the expected product is shown in Figure 3.2. The positive control, *E. coli*, produced three bands, two of which are faint at approximately 2,000 and 900 bp. These two faint bands are likely a result of non-specific binding due to the large amount of DNA within the sample. The middle and brightest of the three bands has amplified at the greatest efficiency and is the desired product size, suggesting this is the 16S rRNA gene. The NTC had no bands present indicating no contamination was present in the reaction. Isolate C, shown in Figure 3.2, has two bands present, one around 1500 bp and the other around 1000 bp, the latter likely also due nonspecific binding.

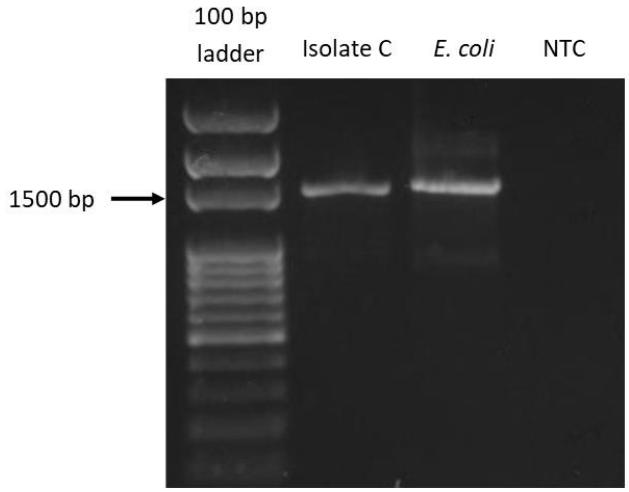

Figure 3.2: Example of 1 % agarose gel for 16S rRNA PCR. *E. coli* used as positive control, NTC = no template control used as negative control.

3.3.2 Genomic DNA Extraction

3.3.2.1 CTAB Extraction

In order to determine the best conditions for the gDNA extraction, two different preparation methods and five different concentrations of RNase and proteinase K were tested (Table 3.1, Figure 3.3). The comparison of cell lysis methods (boiling vs bead beating) showed that bead beating was the appropriate preparation method for the CTAB extraction as gDNA was observed on the gel (Figure 3.3 A). Boiling appeared to degrade the DNA completely. gDNA extraction from each isolate (A-C and E) using bead beating followed by the CTAB method with 1 µL of RNase A and 1 µL proteinase K (EX1, Table 3.1) showed a high molecular weight substance just under the top of the wells, that is likely protein contamination, followed by a band of gDNA, and

subsequently degraded DNA (smearing) leading to a large amount of degraded RNA and small DNA at the bottom of the gel (Figure 3.3 B). This indicated that the gDNA extraction conditions were not optimal for Illumina sequencing.

Using the six other enzyme conditions, outlined in Table 3.1, did not improve the quality of the gDNA extractions (Figure 3.3 C and D). Different combinations of 1 µL of RNase A and 1 µL proteinase K (Figure 3.3 C) were tested to determine if one or the other was contributing to the poor gDNA quality observed in Figure 3.3 B. The gDNA quality was not improved by any combination, indeed without either of these enzymes, no gDNA was observed (EX2). Increasing the quantity of RNase A and proteinase K to 2 µL still did not improve the quality of gDNA (Figure 3.3 D). In addition to the gDNA, some smearing of the DNA was observed, as were bands that were likely to be 23S/16S rRNA and 5S rRNA/tRNA with small degraded DNA and RNA. In a final attempt to improve the quality of the CTAB gDNA extraction, gel purification was tested. Figure 3.3 E shows the gDNA after CTAB extraction, prior to gel extraction. The band containing the gDNA is faintly apparent at the upper end of the gel, with large amounts of degraded DNA and RNA towards the lower end of the gel. After gel purification no gDNA was seen (data not shown) indicating that gel purification did not improve the quality of gDNA. This was likely due to an insufficient starting quantity of gDNA.

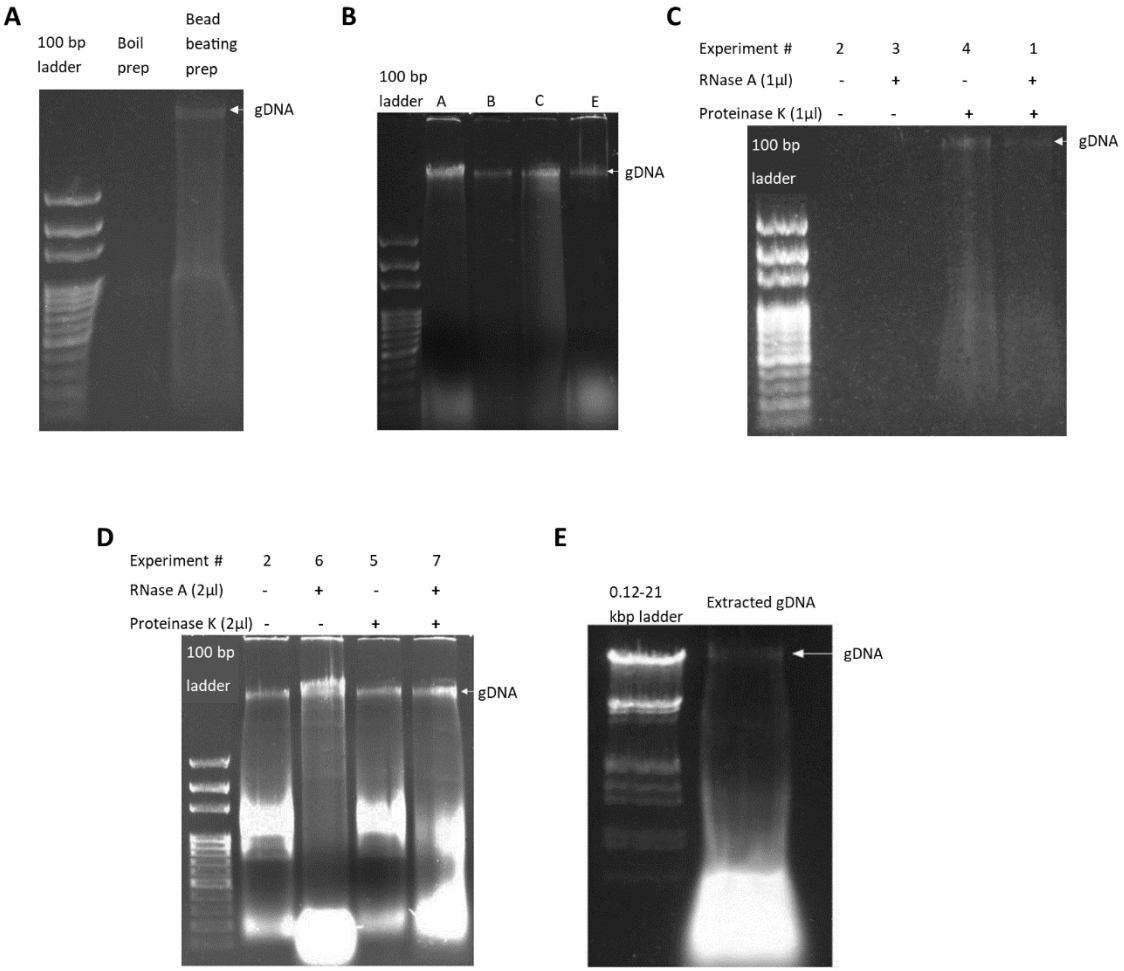

Figure 3.3: Example 0.8% agarose gels for various CTAB extraction methods. (A) Boiling vs bead beating for preparation of DNA for CTAB extraction. (B) CTAB protocol using bead beating DNA preparation and 1 µL of RNase A and 1 µL proteinase K. (C) Presence (+)/absence (-) gel using 1 µL of RNase A and proteinase K. (D) Presence (+)/absence (-) gel using 2 µL RNase A and proteinase K. (E) CTAB extraction gel for gel purification. Experiment # = RNase A and Proteinase K combinations from Table 3.1.

3.3.2.2 Commercial Kit Extraction

In order to improve the quality and quantity of the extracted gDNA, a silica column-based commercial kit was used as described in Section 3.2.2.1. The commercial kit extraction of gDNA for each isolate is shown in Figure 3.4. This gel shows a distinct gDNA band for each isolate above the 21 kbp molecular marker. Slight smearing, degradation of DNA, is also present but is not as severe as seen for the CTAB extracted gDNA (Figure 3.3 B). From visual inspection of the agarose gels, the kit extraction has produced higher quality gDNA for all isolates.

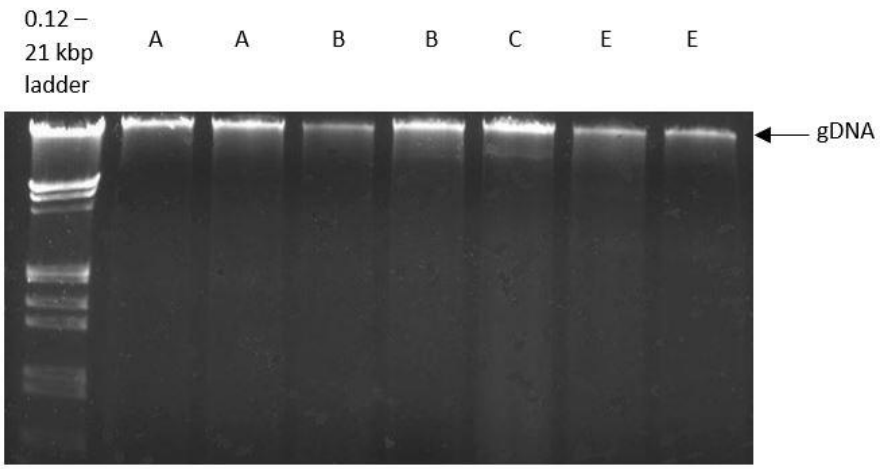

Figure 3.4: 0.8 % agarose gel of gDNA extraction of all four isolates using the Qiagen DNeasy Blood and Tissue Kit.

3.3.2.3 Quality and Quantity of gDNA

The quantity and quality of the gDNA extraction for both extraction methods, CTAB and the commercial kit, are shown in Table 3.4. Generally, the quality of gDNA, based on spectrometric measurements, was similar between the two extraction methods. Both methods gave comparable A_{260}/A_{280} ratio, of around 2 indicating the extractions had low protein contamination as they were close to the desired ratio of 1.8. The A_{260}/A_{230} ratio, measuring other contaminants such as salt and carbohydrates, was not measured for the CTAB method. For isolates B and E, this ratio was high at 4.490 and 4.582, respectively, indicating some potential contaminants while the ration for isolates A and C were 2.045 and 2.221, respectively, closer to the desired ratio range of 2.0 - 2.2. The yield of gDNA was higher using the CTAB method compared to the kit method. This was likely due to the larger number of cells used in the initial steps of the CTAB method compared to the kit, as the kit extraction has a limit to the amount of material that can be applied to the column.

The gDNA extracted using the commercial kit gave superior quality and sufficient yield for Illumina sequencing. Thus, this was the method chosen for gDNA preparation.

Table 3.4: Quantity and quality of gDNA extracted from each isolate using the CTAB and kit extraction methods as measured by spectrophotometry.

	A	B	C	E
CTAB extraction				
Quantity of DNA (ng)	2,640,000	1,548,000	6,300,000	1,704,000
Quality of DNA (A_{260}/A_{280})	2.063	2.027	1.733	2.219
Kit extraction				
Quantity of DNA (ng)	411.32	704.34	2,184	345.8
Quality of DNA (A_{260}/A_{280})	1.957	2.031	1.921	2.032
Quality of DNA (A_{260}/A_{230})	2.045	4.490	2.221	4.582

3.3.3 Genome Assembly

From Illumina high-throughput sequencing, approximately five million PE reads were obtained for each *H. pylori* isolate, with an average nucleotide coverage of around 1,000 times for each isolate (Table 3.5). Raw reads were of good quality and had a low error rate, as seen in the base content and distribution analysis and Phred quality score (Appendix B: Quality of Raw Reads)

Table 3.5: Coverage of *H. pylori* genomes sequenced by Illumina sequencing.

Isolate	Number of raw reads	Average nucleotide coverage
A	5,672,487	1,039x
B	5,104,579	906x
C	4,905,426	942x
E	5,599,851	1,020x

Assembly of the raw sequence reads into contigs produced a number for each isolate, ranging from 36 (isolate A) to 53 contigs (isolate E; Table 3.6). Multiple contigs suggest that regions of the genome are missing, making this a draft and incomplete genome. The assembly programme determined the optimal Kmer value of 127 for each isolate. The N50 and L50 values ranged from

65,734 (E) to 215,027 (B) and 3 (B) to 8 (E), respectively, indicating that the contigs produced were of reasonable quality (Table 3.6).

Table 3.6: Assembly statistics of isolates assembled using SPAdes. Plasmid statistics are included in the total assembly statistics for each isolate.

Assembly statistics	A	B	C	E
Kmer value used for assembly	127	127	127	127
Number of contigs	36	39	37	53
Number of contigs >500 bp	29	30	28	45
N50	136,419	215,027	145,451	65,734
L50	5	3	4	8
Longest contig (bp)	232,847	464,687	302,275	259,545

Visualisation of the *de novo* assemblies using Bandage showed an estimation of the connection of contigs for each isolate and the presence of a singular, circular contig in isolates A, C and E, suggesting the presence of a plasmid in each (Figure 3.5). No plasmid was identified in isolate B. A BLAST search of the plasmids indicated high similarity with other known *H. pylori* plasmids (Table 3.7).

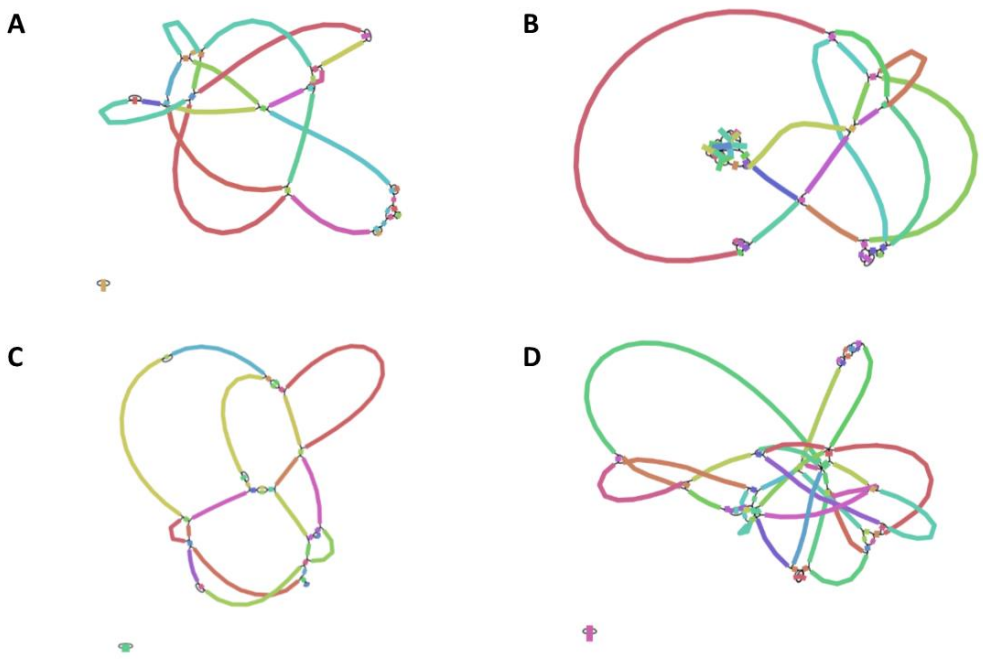

Figure 3.5: Assemblies of sequenced genomes ((A) A, (B) B, (C), C, (D) E) visualised with Bandage. Isolates A, C and E have a singular circular contig suggesting the presence of a plasmid.

Table 3.7: BLASTn results of plasmids from isolates A, C and E.

Plasmid	Size	BLASTn result				
		Description	Query cover	E-value	Percentage (%) identity	Accession
Isolate A plasmid	2,993	*H. pylori* HUP-B14 plasmid pHPB14	84 %	0.0	91.81 %	CP003487.1
Isolate C plasmid	9,121	*H. pylori* strain UM171 plasmid pUM171	100 %	0.0	97.86 %	KM583818.1
Isolate E plasmid	4,212	*H. pylori* strain UM291 plasmid pUM291	95 %	0.0	92.75 %	KM583821.1

Part of the Nullarbor pipeline is to assemble the raw reads before carrying out further analysis. A comparison of the two assembly methods, Unicycler and Nullarbor, can be seen in Table 3.8. Similar G+C contents were calculated by each assembly method, as well as having a similar coverage for each nucleotide. The number of contigs assembled for each isolate differs between the assembly methods, with Nullarbor producing a higher number for isolates A and C (50 and 59) but a lower number for isolates B and E (31 and 49), when compared with Unicycler. N50 values were lower for Nullarbor and L50 values were not produced in the report output. Although each programme produced slightly different assemblies, the Nullarbor assembly was

used to produce the final Nullarbor report and the Unicycler assembly was used for further, manual, genomic characterisations and comparisons for Chapters 3 and 4.

Table 3.8: Comparison of genome assemblies for each isolate using Nullarbor and Unicycler.

	A		B		C		E	
	Nullarbor	**Unicycler**	**Nullarbor**	**Unicycler**	**Nullarbor**	**Unicycler**	**Nullarbor**	**Unicycler**
Reads assembled	11,344,974	11,344,974	10,209,158	10,209,158	9,810,852	9,810,852	11,199,702	11,199,702
G+C content	38.9	38.9	38.4	38.7	38.6	38.8	38.9	38.9
Depth	1,003x	1,039x	903x	906x	867x	942x	990x	1,020x
Contigs	50	36	31	39	59	37	49	53
bp	1,621,082	1,622,179 (+2,993)	1,669,367	1,669,418	1,550,668	1,544,892 (+9,121)	1,632,175	1,629,055 (+4,212)
N50	59,296	136,419	136,576	215,027	52,145	145,451	55,257	65,734
Length of largest contig	198,671	232,847	413,152	464,687	142,816	302,275	169,862	259,545

Numbers in brackets = length of circular contig, presumed to be a plasmid.

3.3.4 16S rRNA Sanger Phylogeny

The 16S rRNA gene was sequenced in both the forward and reverse directions for isolates A, B and C using Sanger sequencing. For isolate E, only the reverse sequence (1,079 bp) was of an acceptable quality and therefore was only used for future analysis (Table 3.2).

Utilising Sanger sequencing data for the 16S rRNA gene phylogeny showed that isolates A, B, C, and E were *H. pylori*. Isolate A, B, C, and E were positioned within the same clade as other *H. pylori* 16S rRNA sequences with a bootstrap value of 95 %, providing sequenced-based identification that these isolates are *H. pylori* (Figure 3.6). This tree suggests that isolates A, C and E belong to the same clade, while isolate B has a different evolutionary history, forming its own clade. However, within this clade the bootstrap values were low, which indicates low resolution between 16S rRNA sequences of *H. pylori* isolates and highlights the need for further resolved phylogenetic tree analysis.

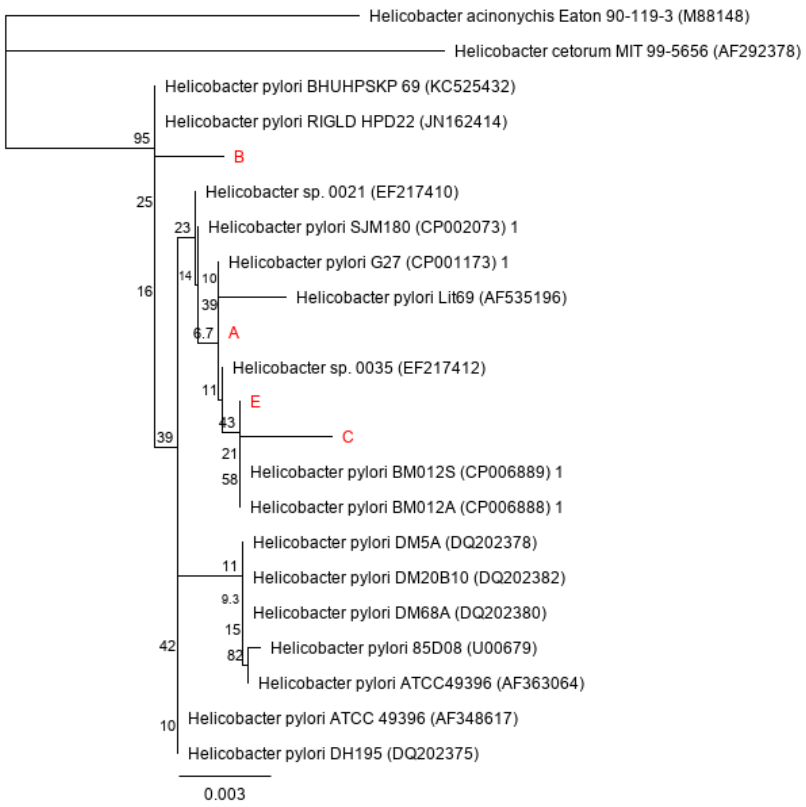

Figure 3.6: Maximum likelihood tree based on 16S rRNA Sanger sequencing of the new *H. pylori* isolates and 16S sequences from the RDP. Isolates (A, B, C and E) from this study are coloured red. Only the reverse sequence of isolate E was used in this phylogenetic analysis. Scale indicates number of substitutions per site.

3.3.5 Comparison of Sanger and Illumina 16S rRNA Gene Sequences

Due to delays in obtaining gDNA for Illumina sequencing, confirmation that the same isolates were sequenced using both Sanger and Illumina sequencing methods was performed. Comparison of 16S rRNA sequences showed that each isolate shared 100 % identity between both sequencing methods, suggesting that the same isolates were sequenced using both sequencing methods (Figure 3.7). BLASTn results show similar lengths of 16S rRNA genes from Illumina sequencing with other *H. pylori* isolates (data not shown).

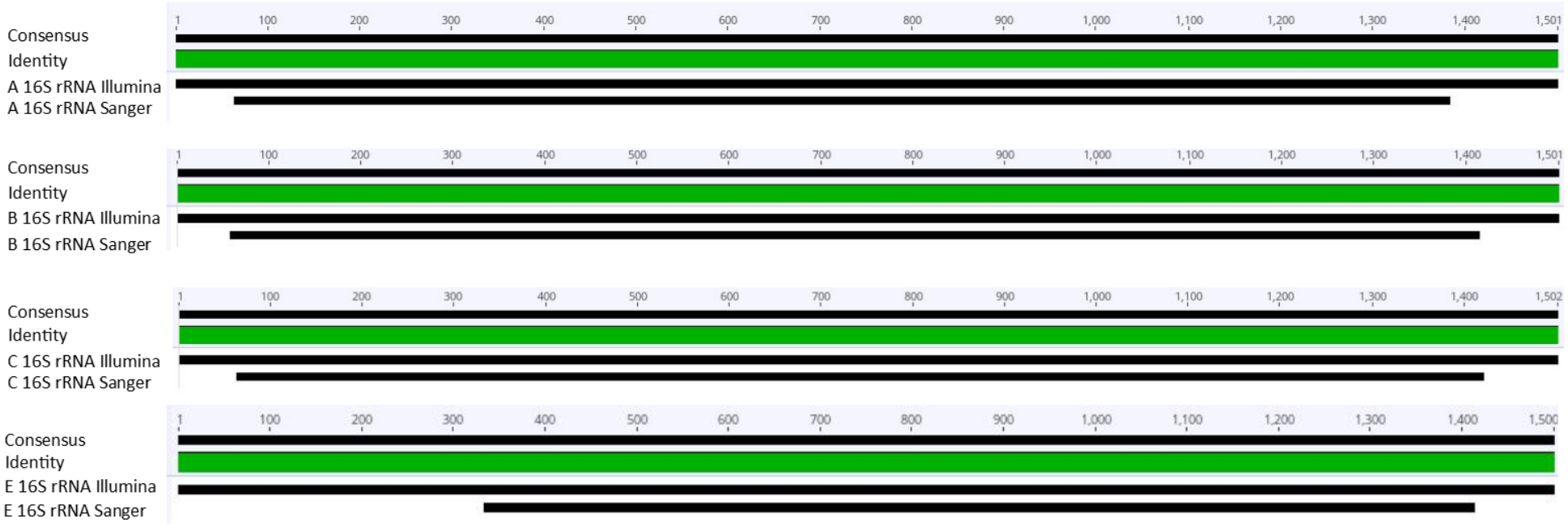

Figure 3.7: MUSCLE alignment of Illumina and Sanger sequences of the four isolates 16S rRNA genes. Isolate E 16S rRNA Sanger is only represented by the reverse sequence (1,079 bp).

3.3.6 Illumina 16S rRNA Tree

The phylogenetic tree presented in Figure 3.8 is based on the Illumina sequenced, full length 16S rRNA genes of the isolates and utilises the dataset created for whole genome analysis (Section 3.2.4.1, Appendix A: Whole Genome Sequence Data Set). The larger dataset was used to increase the resolution of the 16S rRNA tree, and to make direct comparisons with the PhyloPhlAn2 phylogenetic trees (Section 3.3.8). From this rooted tree, it can be seen that all known *H. pylori* and isolates A, B, C, and E group into one clade, confirming the identification of the isolates as belonging to this species. One 16S rRNA gene, NC_017063, appears to be basal to all other 16S rRNA genes included in this dataset; this has not been found in previous 16S rRNA phylogenetic studies on *H. pylori*. In contrast to the 16S rRNA tree shown in Figure 3.6, each isolate from this study is positioned within a different clade. The 16S rRNA sequence of isolate E appears to have the same sequence as the last common ancestor from all other *H. pylori* in the dataset, except for that of NC_017063. The 16S rRNA sequence from isolate A appears to be the same as that of NC_014560 but is poorly supported with a bootstrap value of 42 %. The 16S rRNA sequence from isolates C and B are both within low supported clades with bootstrap values of 32 % and 31 %, respectively. These isolates are also observed on their own branches, showing distinction from other isolates in their respective clades. The spread of these isolates throughout the tree indicated that these isolates (A-C and E) are from different sources.

Overall, this tree is not well supported or resolved since the bootstrap values are low and there is evidence of polytomy. While this might suggest rapid radiation of *H. pylori* strains, this is generally not desirable as it does not allow for accurate and defined relationships to be inferred. However, this tree does indicate that the isolates are distinct and from different sources. Using more than one gene to accurately and reliable understand the phylogeny of the *H. pylori* isolates may improve the quality of the trees.

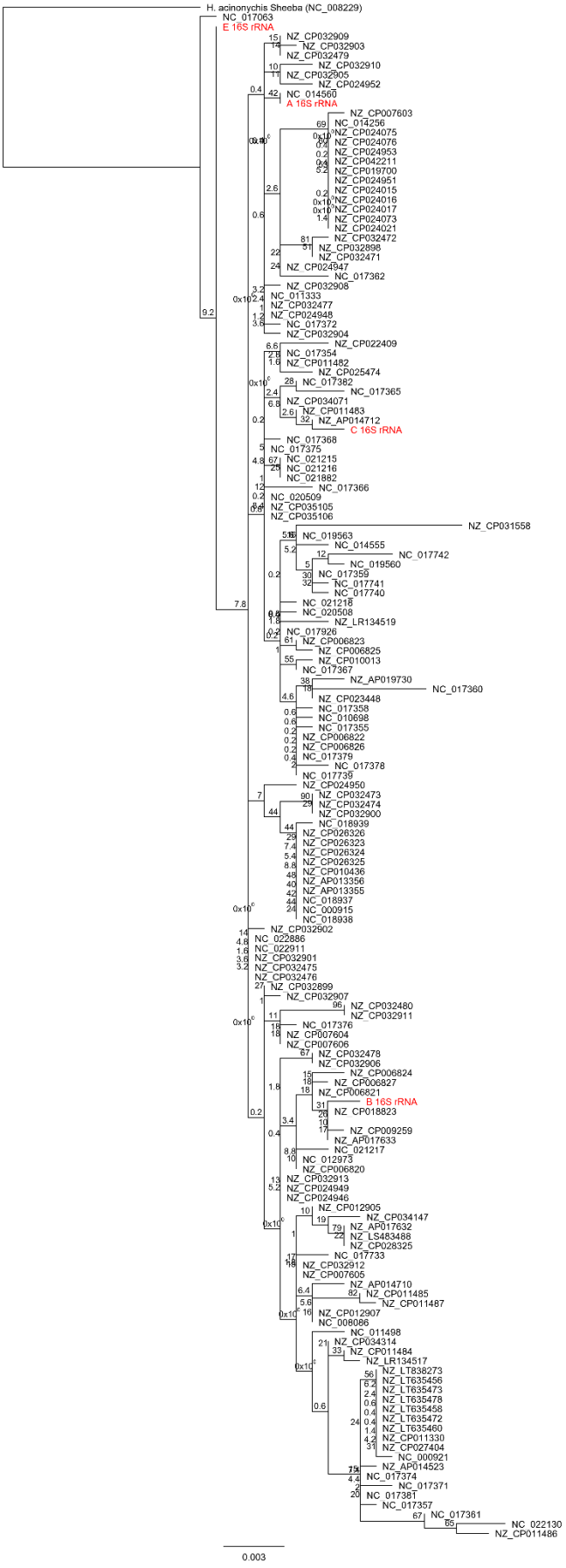

Figure 3.8: Maximum likelihood phylogenetic tree of full length 16S rRNA genes from *H. pylori*. Isolates from this study are highlighted in red. Scale indicates number of substitutions per site.

3.3.7 MLST Analysis

MLST analysis assesses relationships between the sequences shown in Table 3.9. MLST profiles for each isolate were established using both the pubMLST database and Nullarbor, with both methods showing identical results. MLST profiles for each isolate are presented in Table 3.9 and indicate that each isolate from this study is a new strain or ST of *H. pylori*. Isolates A, B and C contain alleles for some loci that are similar to those within the MLST database; however, for some loci they also all possess alleles that are not currently within the database and therefore not observed before. Contrary to this, isolate E possesses all new alleles for the genes currently within the MLST database.

Table 3.9: MLST profiles of each isolate (A, B, C and E) from Nullarbor and the *H. pylori* pubMLST database.

MLST genes	Isolate			
	A	B	C	E
atpA	New	1618	New	New
efp	1454	1454	New	New
mutY	New	New	New	New
ppa	New	New	New	New
trpC	1434	New	New	New
urel	2788	New	New	New
yphC	New	New	1376	New
vacA	New	New	New	New

Numbers (grey) relate to a specific entry within the pubMLST database and "New" (green) denotes a unique allele not within the database.

The MLST data was used to create a neighbour-joining tree to visualise the potential source continent of the isolates (Figure 3.9). This analysis suggests that isolate C came from Asia while isolates A, B and E are all within clades that share isolates from all continents. This further supports the earlier finding from the 16S tree (Section 3.3.6) that these isolates are from differing sources. This tree highlihgts the variation of *H. pylori* isolates based on their continent of isolation and demonstrates that *H. pylori* isolated in NZ likely originated from a variety of different continental origins.

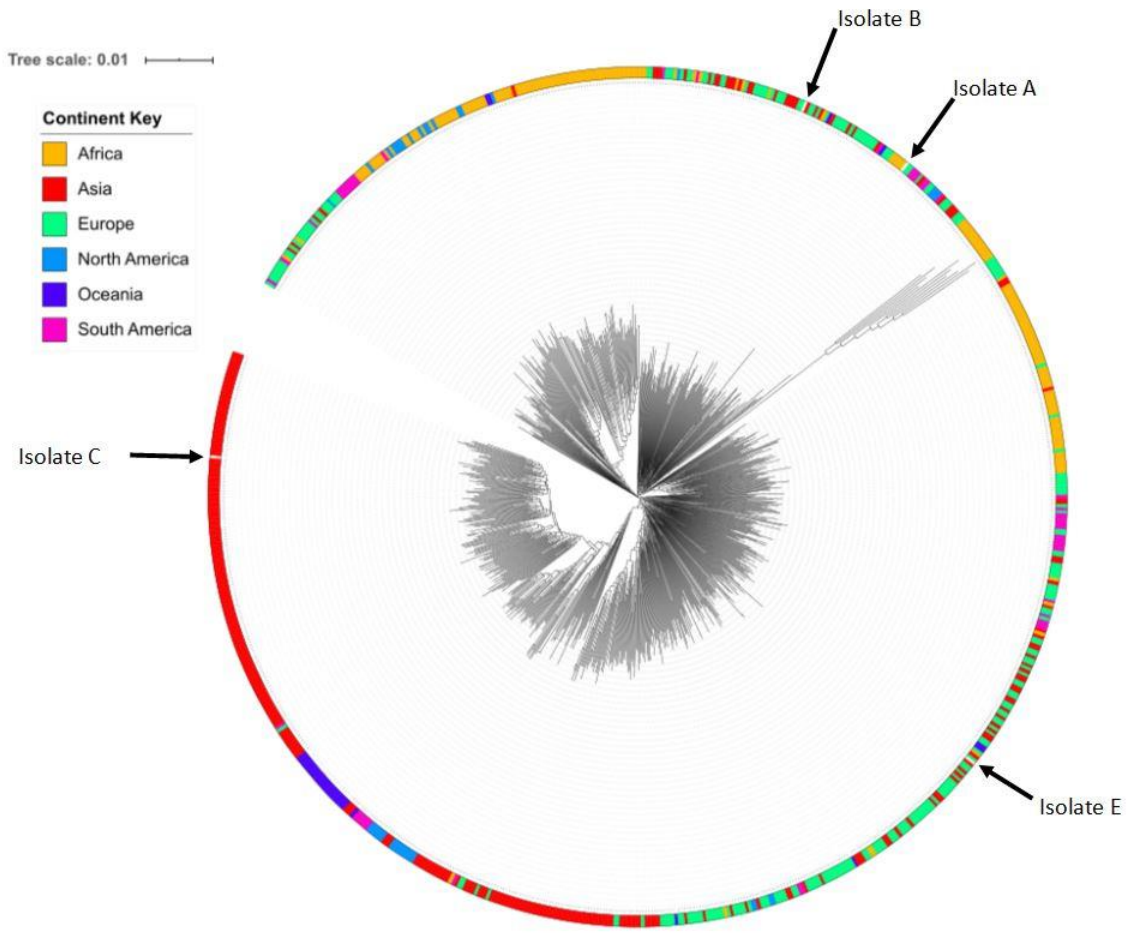

Figure 3.9: Neighbour-joining tree of concatenated MLST data of 834 *H. pylori* isolates and the four isolates from this study (A, B, C and E). Colours around the outside of the tree represent the continent of isolation metadata obtained from pubMLST and correspond to the branches within the Neighbour-joining tree. Isolates A, B, C, and E do not have continent of isolation metadata and are represented by a white space.

3.3.8 PhyloPhlAn2

3.3.8.1 Nucleotide Tree

The constructed nucleotide tree based on 400 conserved genes from whole genome data is presented in Figure 3.10. This figure has had minor clades collapsed to create an easier to read image, the full image may be viewed in Appendix C: PhyloPhlAn Trees. These have been collapsed as they did not contain any isolates from this study. This analysis shows that *H. pylori* is made up of two major clades, one with only three sequences (yellow box from NZ_CP011486, NC_017361 and NC_022130[1]) and the other consisting of all other *H. pylori* sequences in the dataset, including the isolates from this study.

The isolates from this study (A, B, C and E) all fit within one large clade with a supporting bootstrap value of 98 % (grey box; Figure 3.10); in support of the previous analyses, each isolate

[1] Since this analysis was completed NC_022130 has been removed from the RefSeq database.

is in a different clade. However, the clades they are in are different from those observed for the 16s rRNA analysis. Isolate A is grouped with four other isolates, NZ_CP032908, NZ_CP032907, NZ_CP032096 and NZ_CP032478, albeit with a low bootstrap value of 24 %. In the Illumina 16S rRNA tree (Figure 3.8), isolate A was grouped with NC_014560. The subclade highlighted in blue in Figure 3.10 contains isolate B and is well supported with a bootstrap value of 100 %. Within this clade are four other *H. pylori* isolates, NZ_CP006820, NZ_CP006824, NZ_CP006827 and NZ_CP006821. Similar to the 16S rRNA tree, isolate E forms its own reasonably well supported clade (bootstrap value of 74 %); however for this tree, it does not appear to be ancestral to almost other *H. pylori* strains (green box; Figure 3.10). The clade highlighted in orange (Figure 3.10) is the most recent subclade to emerge and contains isolate C; this clade has a high support value of 100 %

.

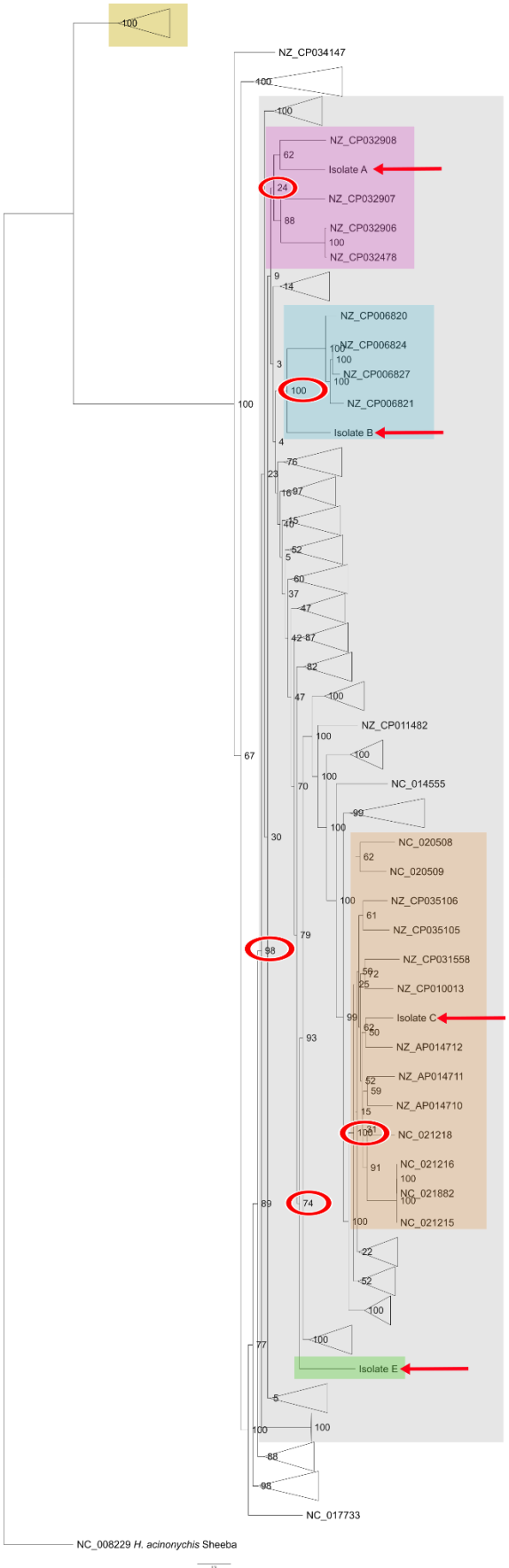

Figure 3.10: PhyloPhlAn2 tree of nucleotide sequences of 155 *H. pylori* complete genomes and four new *H. pylori* isolates. Red arrows indicate isolates from this study. Coloured blocks indicate major clades and red circles indicate major nodes. Scale indicates number of substitutions per site.

3.3.8.2 Protein Sequence Tree

To identify any new or different relationships of the *H. pylori* isolates, based on the more conserved protein sequences, a phylogenetic tree was constructed in PhyloPhlAn2 using the amino acid sequences of the 400 proteins encoded by the genes used for the nucleotide analysis describe above (Figure 3.11). As with the nucleotide tree above (Figure 3.10), minor clades and subclades, without isolates from this study, have been collapsed and a full image may be viewed in Appendix C: PhyloPhlAn Trees. Similar to the nucleotide tree is the first basal lineage of three isolates, highlighted in yellow. This clade consists of the same isolates in the nucleotide tree (NZ_CP011486, NC_017361 and NC_022130) and is in a similar position relative to all other isolates in the dataset.

All isolates from this study, once again, fit within a well-supported clade (bootstrap value of 98 %), highlighted in grey. Within this large clade, there are a number of bifurcations with low support values. This is similar to the nucleotide tree; however, in this tree, many have a support value below 10 % whereas in the nucleotide tree many were below 20 %. This suggests that the protein sequence is not well conserved and are undergoing rapid radiation.

Within the protein sequence tree, the isolates from this study are grouped with different *H. pylori* isolates and are positioned differently within the tree when compared to the nucleotide tree. Isolates B and E are grouped together indicating these protein sequences share a recent common ancestor, but this has a low support value of 48 % (blue box). These two isolates are grouped with several other *H. pylori* isolates (NZ_CP006820, NZ_CP006821, NZ_CP006824 and NZ_CP006827; blue box) into a larger clade with a support value of 66 %. The next isolate to diverge is isolate A (purple clade). This isolate is grouped only with NZ_CP032907 and has a low support value of 42.

Figure 3.11 suggests that isolate C (orange box) appears to have emerged most recently. This clade has good bootstrap support of 84 %; the only clade including isolates from this study to have a bootstrap value above the set threshold of 80 %. This observation is the same as seen in the nucleotide tree (Figure 3.10), with isolate C having the most recent emergence from all other isolates in this study. However, isolate C is grouped with NC_021215, NC_021216 and NC_021882 in the protein sequence tree whereas in the nucleotide tree it is grouped only with NZ_AP014712. The larger clade this subclade resides in (red box), has a low support value of 41 %, but has the same isolates present in the orange clade of the nucleotide tree (Figure 3.10). The low support of this red clade is similar to the other clades around it and the subclades isolates A, B and E fall into but is lower than the support values of the nucleotide tree (orange clade).

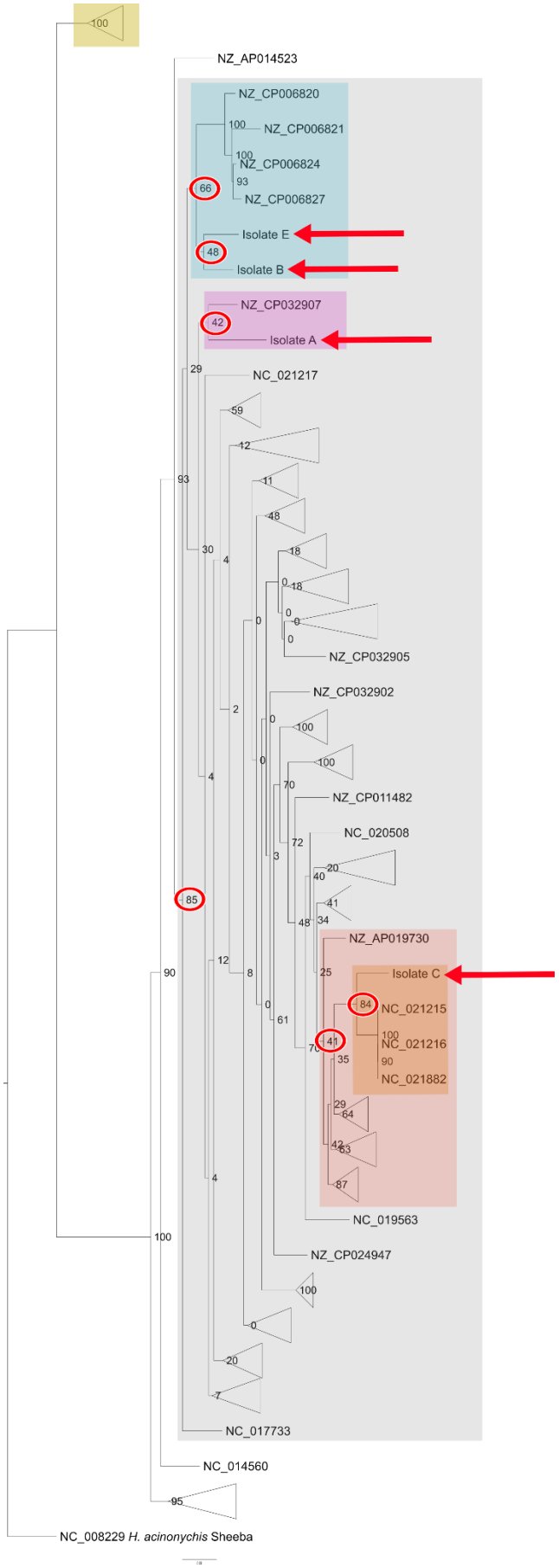

Figure 3.11: PhyloPhlAn2 tree of protein sequences of 155 *H. pylori* complete genomes and four new *H. pylori* isolates. The red arrows indicate isolates from this study. The coloured blocks indicate major clades and red circles indicate major nodes. Scale indicates number of substitutions per site.

3.4 Discussion

Phylogenetics is an important tool for the analysis and understanding of relationships and evolution of genes and species (Martiny et al. 2013). To understand the relationships between isolates A, B, C and E and between these isolates and other publicly available *H. pylori* isolates, phylogenetic analyses were employed. This chapter describes analyses carried out to determine and understand these relationships using a variety of phylogenetic approaches, as well as identifying the closest relatives for isolates A, B, C and E for further comparative analyses of antibiotic resistance.

3.4.1 Genome Extraction

Both single and multiple gene phylogenies were created to help understand the evolutionary relationships between the isolates A, B, C and E, and other *H. pylori* isolates. The single 16S rRNA gene was isolated using PCR amplification, whereas the multiple genes used in the MLST and PhyloPhlAn2 analyses were identified from whole genome data. gDNA was isolated using two methods: a non-commercial, CTAB-based method, and a commercial kit (Qiagen DNeasy Blood and Tissue Kit) to increase the quality of the extracted gDNA used for sequencing.

The CTAB extraction method has primarily been used for the extraction of DNA from plant materials; however, it has been adapted for bacterial DNA extractions (Minas et al. 2011). The main advantages of this method are the low cost of reagents when compared to commercial kits, and its high yield of DNA (Minas et al. 2011). Unfortunately, sufficient quality gDNA was not extracted from isolates A, B, C, and E using this method - an essential component for down-stream analysis. After changes to the CTAB method in an attempt to improve the gDNA quality were not successful, the commercial kit was used. Extraction of gDNA from *H. pylori* is primarily achieved through commercial kits (Dailidiene et al. 2002; Iwamoto et al. 2014; Yari et al. 2016). Limitations of the commercial kit used centre around the yield of DNA extracted due to a maximum cell density that is able to pass through the columns (Minas et al. 2011), although this limitation was not encountered in this study.

It is also important to note the differences in time taken to complete the gDNA extractions. *H. pylori* may take three to seven days for adequate growth using currently available methods (Blanchard and Nedrud 2012). The two extraction methods varied greatly in the time taken to complete; the CTAB method took three days, compared to the commercial kit that only took a few hours. The results from this study indicate that the commercial kit is the superior method for extracting gDNA from *H. pylori* isolates, in terms of quality of DNA as well as time taken to

complete the extraction. The CTAB method requires further adjustments to successfully work with these *H. pylori* isolates.

3.4.2 Genome Assembly

In this study, the four isolates A, B, C, and E were only partially sequenced, producing draft genomes. Illumina HiSeq sequencing technology produces short read lengths of up to 150 bp (Quail et al. 2012). As previously discussed (Section 1.3.6.3), the short read lengths create difficulties when resolving short sequence repeats and assembling *de novo* genomes (Morozova and Marra 2008; Quail et al. 2012; Wold and Myers 2008). Assembling the large amount of data produced through Illumina sequencing produced a number of contigs for each genome, ranging from 36 - 53 (Table 3.6), therefore, the presence of multiple contigs have left these genomes in draft form. Draft forms of genomes may be missing important data that is not included in assembly (Ricker et al. 2012). To accurately complete these genomes a hybrid approach of single-molecule sequencing (e.g. PacBio or Oxford Nanopore), producing long read lengths of 10 - 15 kbp, may be used along with Illumina sequencing to produce a highly accurate genome assembly (Mahmoud et al. 2019). The hybrid assembly uses the longer, single-molecule sequences as a scaffold for the Illumina sequences, as these shorter sequences have a lower error rate and can be used as correctors for the longer sequences (Mahmoud et al. 2019; Quail et al. 2012; Tyson et al. 2017).

The assembled draft genomes of isolates A, B, C, and E were similar to other available *H. pylori* genomes on the NCBI database (https://www.ncbi.nlm.nih.gov/). At the time of this study there were 1,665 publicly available *H. pylori* genomes on the NCBI website. The number of contigs assembled in this study were within the range of the number of contigs obtained for other draft genomes of *H. pylori*, ranging from 1 to 3,766 contigs (Giannakis et al. 2008; Tomb et al. 1997). The presence of a single circular contig was identified in three out of the four isolates, which are assumed to be plasmids. *H. pylori* are common carriers of plasmids ranging in size between 2 and 100 kbp, with approximately 50 % of isolates carrying them (Penfold et al. 1988). The size of the plasmids identified during assembly fit within the average size of plasmids identified in other *H. pylori* isolates and show high similarity to other known *H. pylori* plasmids. The assembled genomes ranged in size from 1,544,892 bp to 1,669,418 bp, fitting within the range of complete *H. pylori* genome sizes; 1,494,183 bp (Kersulyte et al. 2015) to 1,726,836 bp (Su et al. 2019a). G+C content was also comparable with the newly sequenced isolates averaging 38 %, while complete *H. pylori* genomes have a reported G+C content of 38 or 39 % (Avasthi et al. 2011; Duncan et al. 2013). Although the genomes from this study are in draft format, they are still similar to other sequenced genomes of *H. pylori*.

3.4.3 Phylogenetics and MLST

3.4.3.1 16S rRNA Phylogeny

Using small subunit rRNAs to understand and assign taxonomy and phylogeny to prokaryotic organisms has become a gold standard. The ubiquitous 16S rRNA gene has become widely used in both identification and taxonomic classification of bacteria, as well as phylogenetic studies (Watts et al. 2017; Yang et al. 2016). Since 1985, the 16S rRNA gene has been used as a phylogenetic marker due to its combination of highly conserved and hypervariable regions; the latter being useful for phylogenetic analysis and the former for universal primer attachment (Lane et al. 1985; Wang and Qian 2009). Taxonomic classifications using this gene have become a popular tool in clinical laboratories, especially for fastidious or biochemically complex bacteria (Sabat et al. 2017).

Preliminary sequenced-based identification of the four isolates from this study (A, B, C, and E) was achieved using an incomplete sequence of the 16S rRNA gene using the Sanger sequencing platform. The 16S rRNA gene was specifically targeted using universal primers (27F and 1492R); however, sequences were trimmed due to low quality sequences observed around primer attachment regions; a known limitation of this platform (Binladen et al. 2007). As well as an incomplete sequence, only using the last 1,079 bp of isolate E's 16S rRNA, due to low quality sequencing of the forward primer, highlighted another limitation of the Sanger sequence data from this study. However, alignment with extracted 16S rRNA sequences from WGS data from the Illumina sequencing platform provided confidence that the 16S rRNA sequences were true, as well as confirming the same isolates were sequenced in both extractions.

Phylogenetic analyses were completed with both the incomplete 16S rRNA sequences from Sanger sequencing as well as the full-length sequences from Illumina sequences. The presence of polytomy and low bootstrap values highlighted the poor resolution of these phylogenetic analyses performed on this single gene. Previous studies have also questioned the reliability of using 16S rRNA sequences for species-level identification of isolates within the *Helicobacter* genus due to their high sequence similarities (Vandamme et al. 2000).

The basis of utilising the 16S rRNA gene for identification and phylogenetic analysis is centred around the hypervariable regions being species-specific. The specificity of these regions is based on the "complexity hypothesis", that genes involved in transcription or translation are less likely to be transferred than housekeeping genes (Aris-Brosou 2005; Jain et al. 1999). Although original thoughts about sequence variation within these hypervariable regions was thought to be only due to evolutionary changes, it has since been shown that horizontal gene transfer can occur

within this gene, thus leading to potentially inaccurate assessments of evolution (Schouls et al. 2003; Tian et al. 2015; Trieber and Taylor 2002). High frequencies of horizontal gene transfer in *H. pylori* has led to polymorphisms within varying loci being in linkage equilibrium as well as defining this bacterium as panmictic (Salaün et al. 1998; Suerbaum et al. 1998). More specifically, 16S rRNA genes within *H. pylori* isolates have demonstrated natural transformation through the exchange of mutations conferring tetracycline resistance (Trieber and Taylor 2002). Through the increased rate of genetic transfer and recombination within the *H. pylori* population, as well as directly within the 16S rRNA gene, this may help explain the inability for phylogenetic resolution using this gene alone.

Dewhirst et al. (2005) demonstrated that using the 16S rRNA gene for phylogenetic analysis is not reliable for the *Helicobacter* genus due to horizontal gene transfer and mosaicism within the gene. They concluded that using the 16S rRNA gene itself is not the limiting factor of this method but the number of informative bases within the gene. Increasing the number of informative bases in a phylogenetic analysis may be achieved by choosing a larger gene such as the 23S rRNA or increasing the number of genes in the analysis (Dewhirst et al. 2005; Gontcharov et al. 2004). Increasing the number of genes within the phylogenetic analysis has shown to improve phylogenetic resolution and support of previously low supported nodes in single gene analysis (Gontcharov et al. 2004). Utilising a multi locus approach would improve understanding of the relationships between isolates A, B, C, and E and other *H. pylori* isolates.

3.4.3.2 MLST Analysis

MLST analysis is commonly used in clinical settings to identify and classify bacterial strains that may be of epidemiological significance (Bougnoux et al. 2002; Walker et al. 2012). As previously described, MLST data has been widely used with *H. pylori* data to track early human migration and to define seven major population types of this bacterium in modern day (Falush et al. 2003; Linz et al. 2007; Moodley et al. 2009).

Conventional MLST analysis involves PCR amplification of relevant housekeeping genes with specific primers followed by sequencing (Maiden et al. 1998). However, in this study, the genes were identified from the annotated whole genome data. Utilising WGS data as the source of the sequences analysed using MLST is becoming more common as WGS becomes more accessible for more laboratories (Deurenberg et al. 2017). Extracting standard MLST data from complete, incomplete or short sequence reads of bacterial genomes has become a standard part of MLST databases (Jolley et al. 2018; Larsen et al. 2012).

The isolates in this study being defined as new strains or STs based on the sequences available in the *H. pylori* pubMLST database was not an unexpected finding. *H. pylori* is known for having

a high amount of genome sequence diversity due to a high mutation rate and recombination frequency with unrelated strains; much higher than other pathogenic bacteria such as *E. coli* and *N. meningitidis* (Falush et al. 2001; Suerbaum et al. 1998). Due to this high amount of sequence diversity, some studies have indicated that each unrelated human may host its own unique strain or multiple strains of the bacterium (Akopyanz et al. 1992; Linz et al. 2007; Taylor et al. 1995). Strains showing closer relationships to each other have been identified, but only within families (mother and child) or people living closely together (Raymond et al. 2004). Unrelated *H. pylori* strains may arise from coinfection through transmission from another human host or through migration of other strains present in different areas of the stomach (Ailloud et al. 2019). Isolates A, B, C, and E appear unrelated based on both the 16S rRNA and MLST analyses as they are grouped in separate subclades and are defined as separate STs. Genetic diversity observed between different strains of *H. pylori* is attributed to a variety of factors. Migration events between *H. pylori* strains within different parts of the stomach and secondary infections increase recombination between strains (Ailloud et al. 2019; Bugaytsova et al. 2017; Mendoza-Elizalde et al. 2019). As well as recombination, the lack of proofreading activity of DNA polymerase I also contributes to genetic diversity by the accumulation of mutations (García-Ortíz et al. 2011). These events all contribute to the genetic diversity observed between strains of *H. pylori* and is no different from the strains from this study. However, it was not known whether isolates A, B, C, and E were part of a mixed infection.

The phylogenetic method employed in the MLST analysis indicated that isolates A, B, C, and E originated in different continental locations. There were two major clade groupings: Asia (red) and Africa (yellow); and isolates A, B, C, and E were each within different clades of the tree (Figure 3.9). This highlights the fact that each isolate from this study is from a different source and that the isolates are not unique to NZ. The diversity in strains observed within NZ, from this small sample size, may be related to the increasing amount of travel and migration to NZ from international destinations (Stats NZ Tatauranga Aotearoa 2020). Immigrant populations have shown an increased prevalence of *H. pylori* infections compared to the rest of the population in a variety of countries. It has been reported that migrants to countries such as the United Kingdom, Germany, the Netherlands and America, with a higher prevalence of *H. pylori* infections, are usually from third-world countries with lower socioeconomic conditions (Banatvala et al. 1994; Perez-Perez et al. 2005). Further, these *H. pylori* infections show an increase in metronidazole resistance, likely linked to an increased use of nitroimidazole antibiotics in their home country (Banatvala et al. 1994; De Vries et al. 2008; Perez-Perez et al. 2005; Porsch-Ozcürümez et al. 2003).

The birthplace of immigrants has shown to also be a contributing factor with *H. pylori* infection, as those born in lower socioeconomic areas are at a higher risk of becoming infected with *H. pylori* (Chow et al. 1995). Although the route of transmission for *H. pylori* is still not fully understood, as discussed in Section 1.3.2.3, living in lower socioeconomic areas with untreated water; an absence of garbage collection, sewerage system or indoor plumbing; and living in overcrowded areas, all increase the chances of acquiring *H. pylori* (Goodman and Correa 2000; Mhaskar et al. 2013; Parente et al. 2006; Yücel et al. 2009). Migrants from lower socioeconomic regions that have an increased risk of *H. pylori* infections may increase the *H. pylori* population within NZ. This may result in greater diversity within the NZ *H. pylori* population due to recombination with local strains as well as adaptation to new environments and host diets. Therefore, future studies into the implications of migrant *H. pylori* affecting the genetic diversity of the local NZ population are required.

From the Out of Africa migration, to migrations from the Old World to the New World, the expansion of humans across the globe has led to seven modern day populations and a number of subpopulations for *H. pylori* (Section 1.3.2.1) (Falush et al. 2003; Linz et al. 2007; Moodley et al. 2009; Thorell et al. 2017). Subpopulations have arisen from migrations to different places, exposing *H. pylori* to ethnically diverse hosts with varied diets and different stomach environments for *H. pylori* to adapt to (Thorell et al. 2017). With modern populations travelling and migrating, this will likely influence the genomic diversity of this species. A concern relating to the high rate of recombination of *H. pylori,* and increased migration of its hosts, is that the genome may begin to reflect the evolutionary history of its local gene pool, rather than its continent of origin, which would make understanding evolutionary relationships harder to track and observe (Thorell et al. 2017). Understanding the evolution of the *H. pylori* population in NZ and implications of travellers and migrants may help further understand how this bacterium is evolving and developing resistance to antibiotics.

Several phylogenetic studies of *H. pylori* are based on geographic locations and population structures of the isolates, as this species has become an important tool for tracking human migration patterns (Bie et al. 2019; Duncan et al. 2012; Falush et al. 2003; Lamichhane et al. 2019; Muñoz-Ramírez et al. 2017). Population structures are primarily identified using STRUCTURE software and MLST genes (Falush et al. 2003; Montano et al. 2015). These population structures have shown importance not only for tracking human migrations but also for the geographic spread of certain virulence genes (Sheh et al. 2013). Further analysis using software such as STRUCTURE to identify the modern-day population structures of these isolates may help to improve understanding of the origins and evolutionary history of these isolates based on geographic location.

In the current study, MLST analysis was performed based on MLST allele sequences and associated metadata submitted to the *H. pylori* pubMLST database. As of June 2020, there were over 2,600 isolates within the *H. pylori* pubMLST database. Analysis using the iTOL plugin within the pubMLST database is limited to 2,000 records. Due to this limitation, all records within the pubMLST database could not be included in the phylogenetic analysis. Only isolates associated with a publication were included in the analysis so sufficient metadata could be included in the analysis. The remaining isolates within the database, not associated with a publication, are direct submissions into the database. Only including MLST data associated with a publication limited the number of isolates included in the phylogenetic analysis that may help further understand the relationships between *H. pylori* isolates.

Utilising a multi locus approach identified isolates A, B, C, and E as unique strains or STs within the pubMLST database. Although a direct comparison cannot be made between the Illumina 16S rRNA tree (Figure 3.8) and the MLST tree (Figure 3.9), due to different datasets having been used and varying methods, both trees suggested that isolates A, B, C, and E are different sequences and arise from different sources. While MLST data are useful for identifying epidemiologically relevant strains of bacteria and discerning population structures for *H. pylori*, it is limited by sufficient representation of whole genome phylogeny due to a lack of informative bases retrieved from the internal fragments of housekeeping genes (Tsang et al. 2017). To further understand the relationships of isolates A, B, C, and E with available and complete *H. pylori* genomes, PhyloPhlAn2 was utilised.

3.4.3.3 PhyloPhlAn2

To improve the phylogenetic resolution in this study, PhyloPhlAn2 was used. This approach utilises 400 genes, greatly increasing the number of genes used in the phylogenetic analysis. To date, this is the first phylogenetic tree of *H. pylori* isolates constructed using 400 genes and 155 complete genomes. Overall, both the nucleotide and protein sequence PhyloPhlAn2 trees showed similar patterns. Further, these trees were similar to the other trees suggesting rapid radiation of *H. pylori* strains likely the result of high rates of recombination and mutation.

Increasing the number of genes included in the phylogenetic analysis was expected to increase the resolution of the trees and provide greater insight into the evolutionary histories and interspecies relationships of isolates A, B, C, and E with other *H. pylori* isolates. However, low bootstrap values were still observed in both the nucleotide and protein sequence trees (Figure 3.10 and Figure 3.11). The low bootstrap values indicate rapid radiation has occurred at both the nucleotide and protein level. This is likely due to the high rates of mutation and recombination present within this species.

High mutation rates within *H. pylori* have been associated with duration of infection. During the initial, acute stages of infection, *H. pylori* shows a high mutation rate of 6.1 – 8.4 x 10^{-4} changes per site per year, primarily observed in OMPs (Linz et al. 2014). In longer, chronic infections, the mutation frequency appears to decrease to 1.38 x 10^{-5} to 2.5 x 10^{-5} per site per year (Didelot et al. 2013; Kennemann et al. 2011). The changes observed in mutation rates suggest that during initial infection the bacterium is adapting to the new host environment and, once established within the host, may survive (Linz et al. 2014). Initial adaptation to a new host through mutations primarily observed in OMPs contributes to the genetic diversification of isolates within this species and may contribute to the rapid radiation observed through the PhyloPhlAn2 trees.

Recombination may also contribute to the genetic diversity of this species. As previously discussed in Section 3.4.3.2 coinfection with multiple unrelated strains of *H. pylori* can lead to recombination and increased genetic diversity (Didelot et al. 2013). While the genome of a single strain infection appears to be remain stable, the recombination frequency in multiple strain infections has been reported as 5.5 x 10^{-5} per initiation site per year (Falush et al. 2001; Kennemann et al. 2011). It is unknown whether the isolates from this study belonged to a mixed or single infection. Identifying and understating the rate of mixed *H. pylori* infections in NZ may help understand the evolution of strains.

Although mutations and recombination with exogenous DNA cannot be excluded, PhyloPhlAn2 does not identify where the variation is within the genes and proteins analysed. Further analysis would be required to identify the exact regions of diversity within the 400 genes and proteins included in the phylogenetic analysis.

Without more data such as birthplace of the patient, date of isolation and outcome of infection, there is a limit to the amount of data that can be extrapolated from the PhyloPhlAn2 trees. Although the ancestry of these isolates is unable to be accurately estimated using this method the high bootstrap values observed at the tips of the branches provides greater support for identifying the closest relatives of these isolates. Based on the nucleotide sequence, the closest relatives for isolates A, B, C, and E have been identified as *H. pylori* 23-A-EK1 (NZ_CP032908), *H. pylori* oki102 (NZ_CP006820), *H. pylori* ML3 (NZ_AP014712) and *H. pylori* J182 (NZ_CP024947), respectively.

The question still remains about how these isolates of *H. pylori* developed antibiotic resistance. Although this phylogenetic analysis did not primarily focus on genes related to antibiotic resistance, determining the evolutionary histories of these isolates may have provided clues about their acquisition of resistance. As described in Chapter 2, isolate A was sensitive to all antibiotics tested, isolates B and E were resistant to > 256 µg/mL of clarithromycin and

metronidazole, and isolate C was resistant to 24 µg/mL clarithromycin. While their positions in the trees did not appear to correlate with antibiotic resistance, the PhyloPhlAn2 trees and the positions of isolates A, B, C, and E within these trees, suggests that antibiotic resistance is not a deep-rooted trait and is instead likely to have developed independently and through different ways. Most antibiotic resistance mechanisms reported for *H. pylori* are the result of gene mutation (Section 1.3.5), likely as an adaptation to the changing environment. Rapid convergent evolution has been identified in *H. pylori* and is likely associated with micro-niches within hosts, geographic spread and differences between hosts, such as diet and genotype (Chattopadhyay et al. 2018). It is unknown whether these single isolates (A, B, C, and E) have developed resistance independently or if the clades they reside in have all developed similar resistance phenotypes. Using comparative genomics with the respective isolate's closest relatives, based on the nucleotide tree, may help elucidate and increase our understanding of what causes the resistance phenotypes and how this resistance develops. This analysis is described in Chapter 4.

To further increase the resolution of these trees and help resolve the ancestry of modern *H. pylori* populations, utilising a whole genome phylogeny approach or increasing the number of sequences may help. Employing a whole genome phylogeny, such as a SNP tree, will greatly increase the number of informative bases within the analysis and may provide greater understanding of the evolutionary relationships amongst this diverse species (Shakya et al. 2020).

As the genomes for the isolates A, B, C, and E were in draft format, a whole genome phylogenetic approach was unable to be used in the current study. To complete a whole genome phylogeny, firstly, the draft genomes for isolates A, B, C, and E would need to be completed as described in Section 3.4.2. Secondly, increasing the number of complete genomes may help further resolve the ancestry of modern-day isolates. In the last few years, the number of complete *H. pylori* genomes within the GenBank database has greatly increased. This is likely due to the reduced cost of sequencing and the clinical importance of this bacterium, especially in relation to antibiotic resistance (Savoldi et al. 2018). As WGS becomes more accessible and more genomes are completed, including complete genomes from *H. pylori* strains isolated within NZ and worldwide may help further our understanding of how the *H. pylori* population within NZ is changing. Further studies are required to fully understand the evolutionary history of *H. pylori* isolates, especially in NZ; completing the genomes from this study is the first step to create a more robust and resolved phylogenetic tree.

This chapter highlights the significance of genomic diversity seen within *H. pylori* genomes. Due to this extreme diversity, probably from a combination of recombination and mutation, using a

single gene such as 16S rRNA has failed to resolve interspecies relationships sufficiently enough to elucidate where isolates A, B, C, and E sit within the tree. Utilising MLST increased the number of genes used, identified isolates A, B, C, and E as new strains and highlighted that *H. pylori* in NZ did not originate form one source, but rather has had a range of continental origins. However, due to the limited informative sequences within the housekeeping genes selected it was not used as a representation for whole genome phylogeny. Further increasing the number of genes analysed to 400 in an attempt to increase the informative sequences, PhyloPhlAn2 was used to improve the resolution of the phylogenetic trees. This analysis, consisting of 155 complete *H. pylori* genomes and the four isolates from this study (A, B, C, and E) is, to date, the first phylogenetic analysis using PhyloPhlAn2 of this size for *H. pylori*. Although the ancestry of the isolates was not fully resolved using this method and inference could not be made directly about antibiotic resistance, the closest relative for each NZ isolate were identified. To further understand the mechanisms of resistance and how these mechanisms have developed, comparative genomic analysis was performed with the closest relative for each isolate based on the PhyloPhlAn2 nucleotide tree; this analysis is described in Chapter 4.

Chapter 4 Genomic Characterisation and Comparison of *H. pylori* Isolates

4.1 Introduction

Antibiotic resistance is a serious threat to public health, affecting individuals worldwide from all socioeconomic backgrounds (Laxminarayan et al. 2016; World Health Organization 2020). *H. pylori* has been defined as a high priority antibiotic resistant organism in need of new therapeutic treatments (Tacconelli et al. 2018). Resistance rates to common first-line treatment options for *H. pylori* vary worldwide but an increase towards clarithromycin and metronidazole has been observed, especially in NZ (Ahmed et al. 2004; Hsiang et al. 2013; Savoldi et al. 2018). Antibiotic resistance can have a negative effect on the treatment of an infection leading to treatment failure (Savoldi et al. 2018). The slow growing and fastidious nature of this organism, coupled with the lack of agreed and defined standards for antibiotic susceptibility testing (Section 2.4) calls for a fast and accurate diagnosis of resistance for an appropriate treatment regimen to be selected for eradication of the infection (Malfertheiner et al. 2017). Understanding the underlying, genomic features conferring antibiotic resistance will help in developing fast and accurate tests to understand the susceptibility profiles of bacteria (Su et al. 2019b). Determining the genomic features of resistance will also help track these mechanisms and aid in the understanding of the evolution of antibiotic resistance (Hendriksen et al. 2019).

From the previous two chapters isolates A, B, C, and E were characterised by their resistance profiles, their evolutionary relationships with other *H. pylori* isolates were analysed and their respective closest relatives were identified. All isolates were identified as new strains or STs and a high level of genomic diversity was observed throughout the phylogenetic trees produced. For the isolates themselves, isolate A was sensitive to all antibiotics tested and its closest relative was *H. pylori* strain 23-A-EK1. Isolate B and E were both resistant to > 256 µg/mL of both clarithromycin and metronidazole and their closest relatives were *H. pylori* strains oki102 and J182, respectively. Finally, isolate C was resistant to 24 µg/mL of clarithromycin and its closest relative was *H. pylori* strain ML3.

Comparative genomics is an important tool used to help further our understanding of biological processes through the comparison of genomic features (Xia 2013). *H. pylori* was the first bacteria to have two complete genomes sequenced and was subject to the first comparative genomic analysis (Alm et al. 1999). This tool has been applied to multidrug resistant bacteria such as *A. baumannii* and *K. pneumoniae* to elucidate resistance mechanisms and evolutionary patterns

(Fournier et al. 2006; Kumar et al. 2011). This comparative method has not been performed on *H. pylori* isolates from NZ.

A review of the known or suggested mechanisms of resistance towards first-line treatment antibiotics for *H. pylori* infections was described in Section 1.3.5. As stated above, the isolates from this study only showed resistance towards clarithromycin and metronidazole; belonging to the macrolide and nitroimidazole drug classes, respectively. Thus, these two antibiotics and the mechanisms associated with their resistance will be the focus of this chapter. The review in Sections 1.3.5.3 and 1.3.5.4 highlighted a number of mutations within key genes associated with clarithromycin and metronidazole resistance, respectively. Key genes and associated mutations include mutations within domain V of the 23S rRNA gene relating to clarithromycin resistance, and amino acid substitutions within RdxA and FrxA, preventing metronidazole from being reduced into its bactericidal components (Jenks et al. 1999; Saranathan et al. 2020; Tran et al. 2019; Versalovic et al. 1997). Although many potential mechanisms have been identified in resistant strains, some causes of resistance are still yet to be determined and defined (Chang et al. 1997; Smiley et al. 2013). Conducting further studies to advance the knowledge around resistance mechanisms within *H. pylori* is crucial for the accurate diagnosis and development of alternative treatment regimens as well as understanding what mechanisms of resistance are present within *H. pylori* isolates from NZ.

The aims of this chapter were to further our understanding of the antibiotic resistance mechanisms within *H. pylori* isolates from NZ using comparative genomic techniques (Figure 4.1). Isolates A, B, C, and E underwent genomic characterisation and comparative genomic analysis with the closest relatives identified from the nucleotide PhyloPhlAn2 tree from Chapter 3 Phylogenetic Relationship of *H. pylori* isolates.

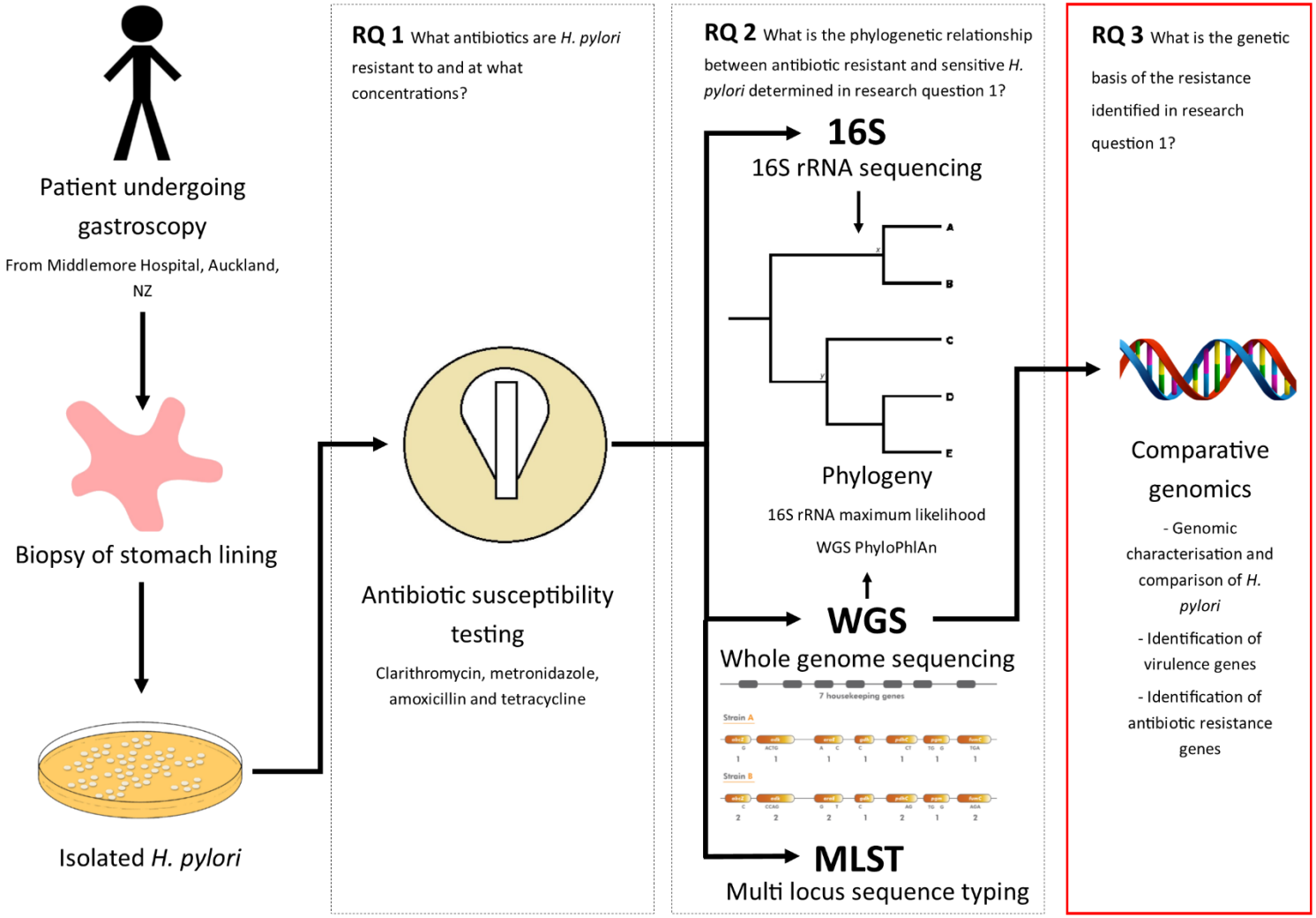

Figure 4.1: Schematic diagram of aims and research flow of this thesis. The RQ and aims of Chapter 4 are highlighted in red.

4.2 Methods and Materials

4.2.1 Genomic DNA Extraction and Assembly

H. pylori isolates were grown and gDNA was extracted following methods previously described in Sections 2.2.2 and 3.2.2.1 (commercial kit), respectively. Genomes were sequenced and assembled as described in Section 3.2.2.4 and 3.2.3, respectively.

4.2.2 Gene Predication and Annotation

Annotation of the genomes were conducted using three pipelines; RAST v2.0 (Aziz et al. 2008), Prokka v1.13.3 (Seemann 2014) and the previously mentioned Nullarbor v2.0.20191013 (Seemann et al. n.d.). An overview of the workflow and use of output files may be viewed in Figure 4.2. RAST annotation was performed by uploading each *H. pylori* genome onto the RAST server (http://rast.nmpdr.org/). Prokka annotation was completed in-house using standard settings and parameters. Nullarbor was conducted in-house using standard parameters and *H. pylori* 26695 (NC_000915) was used as a reference.

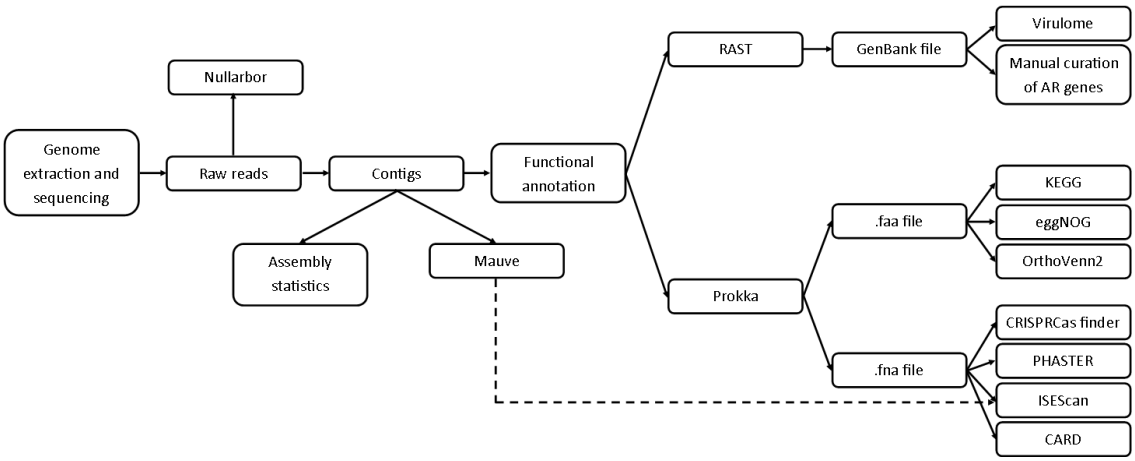

Figure 4.2: Workflow of genomic extraction, sequencing and functional annotation of isolates A, B, C, and E used in this thesis. Programmes used with different file types are presented at end of workflow. Dotted line indicates that the progressiveMauve alignment was used with ISEScan as well. AR = antibiotic resistance.

4.2.3 Genomic Features

CRISPR (clustered regularly interspaced short palindromic repeats) regions were identified using the online tool, CRISPRCasFinder (https://crisprcas.i2bc.paris-saclay.fr/). The *.fna file was uploaded to the webpage and analysed using default settings plus CAS gene detection (Couvin et al. 2018). Phage regions were identified using PHASTER (PHAge Search Tool Enhanced Release) (https://phaster.ca/). Once again, the *.fna file was uploaded and analysed using the default settings (Arndt et al. 2016; Zhou et al. 2011). Proteins identified by PHASTER were subsequently searched within InterPro v80.0 using InterProScan and through a BLASTp search

on the NCBI database to further identify conserved regions within hypothetical and phage-like proteins that may be involved in antibiotic resistance. Genes surrounding this region were also investigated for any points of potential interest relating to antibiotic resistance. ISs were identified through ISEScan, a pipeline used to identify both complete and partial ISs in prokaryotic genomes (Xie and Tang 2017). The *.fna files and the sorted and concatenated Mauve files were input into the pipeline and default parameters were selected. A BLASTp search was conducted through the NCBI database on the IS identified by ISEScan.

4.2.4 Mauve

The Mauve v1.1.1 plugin for Geneious was used to rearrange the contigs and align the newly sequenced *H. pylori* genomes against reference genomes identified from the nucleotide PhyloPhlAn2 tree (Section 3.3.8.1, Table 4.1). To reorder the contigs the MCM algorithm was used and the respective closest relative was chosen as the reference. The sorted sequences were concatenated and circularised. The progressiveMauve algorithm was then run to align the genomes. The match seed weight was adjusted to 11, to match the approximate size of the *H. pylori* genomes (1.6 Mbp), and the minimum weight of the LCB was set as the default.

Table 4.1: *H. pylori* isolates from this study and their closest relative(s) identified from the nucleic acid PhyloPhlan2 tree.

Isolate	Closest relative	NCBI accession number	Reference
A	*H. pylori* 23-A-EK1	NZ_CP032908.1	(Ailloud et al. 2019)
B	*H. pylori* oki102	NZ_CP006820.1	(Satou et al. 2014)
C	*H. pylori* ML3	NZ_AP014712.1	(Wang et al. 2015a)
E	*H. pylori* J182	NZ_CP024947.1	(Su et al. 2019a)

4.2.5 Whole Genome Analysis

4.2.5.1 Nullarbor

As previously described in Sections 1.3.6.4 and 3.2.3, Nullarbor produces a public health microbiology report using a range of different programmes. This report consists of previously described analyses (Section 3.2.3), as well as resistome, virulome and single nucleotide polymorphism (SNP) analyses. Features of this report were compared with other methods.

4.2.5.2 Virulome

The virulence factor database (VFDB) 2019 (http://www.mgc.ac.cn/VFs/main.htm) was used to predict the virulome for isolates A, B, C, and E (Liu et al. 2019). Utilising the VFanalyzer (www.mgc.ac.cn/cgi-bin/VFs/v5/main.cgi?func=VFanalyzer), the GenBank file for each assembled draft genome, produced from RAST, was uploaded to the web-based interface. *H. pylori* 26695 was selected as the representative genome.

4.2.5.3 Metabolic Potential

The metabolic potential of isolates A, B, C, and E was determined using the KEGG database. The *.faa files for each genome were uploaded to the BlastKOALA webpage (https://www.kegg.jp/blastkoala/). The bacteria taxonomic group was selected and the species_prokaryote KEGG GENES database was selected. The genomes were annotated and genes were assigned KO numbers and uploaded to the KEGG Mapper – Reconstruct Pathway webpage (https://www.kegg.jp/kegg/tool/map_pathway.html) to identify genes present within the pathways within the KEGG pathways database.

4.2.5.4 Clusters of Orthologous Groups Assignment

To assign OGs to each isolate genome (A, B, C, and E), the *.faa files were uploaded to eggNOG-Mapper (http://eggnog-mapper.embl.de/). The default settings were selected before running the analysis (Huerta-Cepas et al. 2017; Huerta-Cepas et al. 2018). The closest relative for each isolate (Table 4.1) was also analysed through eggNOG-Mapper for a comparison with each isolate (A, B, C, and E). The annotated eggNOG files were imported into Excel v1908 and normalised to create a bar graph to visualise the OGs assignments.

4.2.6 Identification of Potential Antibiotic Resistance Genes and Mutations

To identify genes or mutations that may be involved in antibiotic resistance within isolates A, B, C, and E, multiple methods were used. As part of the Nullarbor pipeline a resistome report is produced from a search within CARD. A manual curation of genes and mutations from the literature was also conducted. As well as a direct search through CARD's RGI and OrthoVenn2 to identify novel genes of interest shared between resistant isolates that may be associated with antibiotic resistance. Each of these methods are described below.

4.2.6.1 Manual Curation of Known Antibiotic Resistance Genes from *H. pylori*

From Section 1.3.5 a number of genes or proteins and mutations were discussed that are associated with antibiotic resistant *H. pylori*. Genes associated with clarithromycin resistance (Section 1.3.5.3), 23S rRNA, translation initiation factor IF-2 (*infB*) and ribosomal protein L22 (*rpl22*); proteins associated with metronidazole resistance (Section 1.3.5.4), RdxA (*rdxA*, oxygen-

insensitive NADPH nitroreductase), FrxA (*frxA*, NADPH-flavin oxidoreductase) Fur (*fur*, ferric uptake regulator) and RecA (*recA*, recombinase A) were extracted from the Genbank format files of isolates A, B, C, and E and from their respective closest relative genomes (Table 4.1). Eflux pumps, HefABC (*hefA*, outer membrane protein; *hefB*, membrane fusion protein; *hefC*, cytoplasmic pump protein), HP1184 (*hp1184*, multidrug efflux MATE transporter) and HP1181 (*hp1181*, multidrug efflux MFS transporter) were also extracted from the GenBank format files of their isolates and respective closest references. The *H. pylori* type strain, 26695, was used as a general reference and all genes mentioned above were extracted from its Genbank file. Each gene and/or predicted protein sequence was analysed by multiple sequence alignment using the MUSCLE algorithm v3.8.425 within Geneious v2020.0.4.

4.2.6.2 Novel Antibiotic Resistance Determinants

CARD

CARD was used to identify potential antibiotic genes through their online tool (RGI; https://card.mcmaster.ca/) and through Nullarbor. The RGI was used to identify potential resistomes in each of the isolates from this study. The DNA sequence for isolates A, B, C, E and their respective closest relatives (Table 4.1) were uploaded through the RGI interface. The selection criteria were initially chosen to only include perfect and strict hits, this returned few results, so the criteria were adjusted to include perfect, strict and loose hits to increase the pool of potential antibiotic resistance genes.

The RGI results were condensed to only include genes associated with the relevant classes of antibiotics the isolates (A, B, C, and E) were tested against; macrolides, nitroimidazoles, tetracyclines and penicillins. The RGI output of genes associated with macrolide and nitroimidazoles was visualised using a heatmap. The percentage similarity of genes associated with macrolides and nitroimidazoles were incorporated into a heatmap using the heatmap.2 function from gplots v3.0.4 (Warnes et al. 2015) in R v3.6.1 (R Core Team 2019) and visualised using Rstudio v1.3.1073 (RStudio Team 2020).

Protein predictions of the genes identified were analysed through InterPro (https://www.ebi.ac.uk/interpro/) and the NCBI BLASTp search tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to identify and analyse predicted protein families, conserved sites and domains and to observe similarity to other known protein sequences.

OrthoVenn2

OrthoVenn2 was used to compare whole-genome orthologous clusters between the isolates from this study. The *.faa files for each isolate were uploaded to the web-based interface

(https://orthovenn2.bioinfotoolkits.net/home). Analysis was carried out using the default settings of an E-value of $1x10^{-2}$ and an inflation value of 1.5.

To identify orthologous protein clusters (OPC) that may be involved with antibiotic resistance a number of comparisons were carried out. As isolate A was sensitive to all antibiotics tested it was used as a negative control for comparison to the other isolates. Isolate A was compared to isolates B, C and E individually and OPCs that were present only within all isolates B, C, and E were collated. OPCs that were only present in the resistant strains were also collated. As isolates B and E share the same resistance phenotypes, OPCs shared between only these two isolates were also collated.

The protein sequences for each identified cluster from the comparisons described above were searched through InterPro and NCBI BLASTp search, as described in Section 4.2.6.2 (CARD).

4.2.7 Presence/Absence of A2147G Mutation Associated with Antibiotic Resistance

To observe the evolution of the known antibiotic resistance mutation, A2147G from the 23S rRNA, identified in isolates B and C, these mutations were overlaid onto the nucleotide PhyloPhlAn2 tree from Chapter 3 (Section 3.3.8.1). The 23S rRNA gene was identified in the PhyloPhlAn2 dataset (Appendix A: Whole Genome Sequence Data Set) through the inbuilt BLAST search tool within Geneious v2020.0.4. The sequences were aligned using the MUSCLE algorithm within Geneious and the A2147G mutation was identified. The nucleotide PhyloPhlAn2 tree was edited in Inkscape v0.92.4 to show the presence or absence of the suggested mutations associated within clarithromycin and metronidazole resistance.

4.3 Results

4.3.1 Gene Prediction and Annotation

RAST, Prokka and Nullarbor were used to predict and annotate genes. A summary of the number of coding genes, rRNAs and tRNAs each programme identified can be viewed in Table 4.2. Each programme identified a varying number of genes, with RAST calling the highest number. Nullarbor was unable to identify any tRNA genes, therefore questioning its accuracy with gene prediction.

As well as a different number of genes and proteins identified, the description of identified proteins can differ between programmes. Nullarbor does not produce a searchable file for annotated genes, therefore, only Prokka and RAST gene calls were compared, as well as a BLASTX search in CDS = **coding sequence.**

Table 4.3. This table demonstrates that both RAST and Prokka were able to call predicted proteins, however, the description of these proteins differed. This highlights the importance of using both programmes to predict protein functions.

Table 4.2: Summary of genes and RNAs predicted for each isolate (A, B, C and E) utilising RAST, Prokka and Nullarbor.

	A			B			C			E		
	Nullarbor	**RAST**	**Prokka**	**Nullarbor**	**RAST**	**Prokka**	**Nullarbor**	**RAST**	**Prokka**	**Nullarbor**	**RAST**	**Prokka**
CDS	1,509	1,629	1,519	1,571	1,689	1,570	1,467	1,584	1,473	1,542	1,679	1,547
rRNA	2	3	2	3	3	2	2	3	2	2	3	2
tRNA	0	48	36	0	48	36	0	49	36	0	48	36

CDS = coding sequence.

Table 4.3: Comparison of gene annotations between RAST, Prokka and BLASTX for isolate A.

Prokka locus tag	**RAST description**	**Prokka description**	**BLASTX result**		
			Description	**E-value**	**Accession**
INKADMOK_01020	Type IIS restriction enzyme M protein (mod)	Restriction enzyme BgcI subunit alpha	N-6 DNA methylase	0.0	WP_108576009.1
INKADMOK_01028	ABC transporter, ATP-binding protein HP_1465	Methionine import ATP-binding protein MetN 2	ABC transporter ATP-binding protein	0.0	WP_023526679.1
INKADMOK_01117	(S)-2-hydroxy-acid oxidase	Putative FAD-linked oxidoreductase	glycolate oxidase subunit GlcD	0.0	WP_101005179.1
INKADMOK_01554	Outer membrane protein (omp27)	Hypothetical protein	Hop family adhesin HopQ	0.0	WP_128072307.1

4.3.2 Genomic Features

General genomic features of the *H. pylori* genomes are displayed in Table 4.4. Genome sizes ranged from 1.54 Mbp (isolate C) to 1.67 Mbp (isolate B). The average G+C content was approximately 38 % for each genome. RAST identified one copy of each rRNA gene (5S, 16S and 23S) in each isolate along with 34 (A and B) and 35 (C and E) tRNA genes in each respective genome. Isolates C and E were found to have one and four CRISPR repeat regions, respectively, and one phage region was identified in Isolate B. Both complete and partial ISs were identified in all isolate genomes.

Table 4.4: Genomic features of draft genomes of isolates A, B, C and E and three plasmids from RAST.

Features	A	A plasmid	B	C	C plasmid	E	E plasmid
Genome size (bp)	1,622,179	2,993	1,669,161	1,544,892	9,121	1,629,055	4,212
GC content (%)	38.9	35.4	38.7	38.8	33.7	38.9	37.8
Number of contigs	35	1	39	36	1	53	1
Total genes	1,682	3	1,760	1,631	11	1,737	6
Protein encoding genes (%)	1,626 (96.67%)	3 (100 %)	1,689 (95.97%)	1,573 (96.44%)	11 (100 %)	1,673 (96.32%)	6 (100 %)
Hypothetical proteins (% of CDS)	419 (25.77%)	1 (33.33%)	466 (27.59%)	379 (24.09%)	4 (36.36%)	461 (27.56%)	4 (66.67 %)
Genes assigned to COGs (% of CDS)	1,345 (82.72%)	1 (33.33%)	1,372 (81.23%)	1,333 (84.74%)	1 (9.09%)	1,382 (82.61%)	1 (16.67 %)
Total RNA genes	37	0	37	38	0	84	0
rRNA genes (5S, 16S, 23S rRNA)	3 (1, 1, 1)	0	3 (1, 1, 1)	3 (1, 1, 1)	0	3 (1, 1, 1)	0
tRNA genes	34	0	34	35	0	35	0
CRISPR repeats[1]	0	0	0	0*	0	0*	0
Phage regions[2]	0	0	1	0	0	0	0
Insertion sequences[3]	2♦	0	3♦	4♦	0	6♦	0

[1] = identified using CRISPRCas finder. [2] = identified using PHASTER. [3] = identified using ISEScan. * = incomplete CRISPR regions identified. ♦ = total number of IS identified (both complete and partial).

4.3.2.1 CRISPR Regions

One potential CRISPR region was identified in isolate C and four potential CRISPR regions were identified in isolate E (Table 4.5). All regions identified contained only two direct repeats (DR) and one spacer, with an assigned evidence level of one. This low evidence level, calculated by the low number of DRs and spacers, indicates that these short arrays are unlikely to be true CRISPR regions. A BLASTn search of these DR sequences indicates that these regions have been identified in other *H. pylori* isolates, including the cagA region, but not in any known CRISPR regions. No Cas regions were identified.

Table 4.5: Identified potential CRISPR repeats in isolates C and E.

Isolate	Length of DR (bp)	DR sequence	DRs in CRISPR locus	Number of spacers	Evidence level
C	23	ACTTGAGCGTAAATGGGCTCTTC	2	1	1
E	27	TTTTTTAAAAAAACCCTTGAGTGTTTT	2	1	1
E	23	GATGAGCCCATTTACGCTCAAGT	2	1	1
E	37	GTTTTTTTGAAATAATCTCTTCGGTAACGCTAGAGAG	2	1	1
E	26	CATGCGAGAAGTAGTGGTTCAAGAAC	2	1	1

DR = direct repeat.

4.3.2.2 Phage Analysis

One incomplete phage region was identified within contig 4 of isolate B. The region identified was 30.8 Kbp long, consisting of 22 proteins (10 phage hit proteins and 12 hypothetical proteins), a transposase region and had a G+C content of 34.54 % (Figure 4.3).

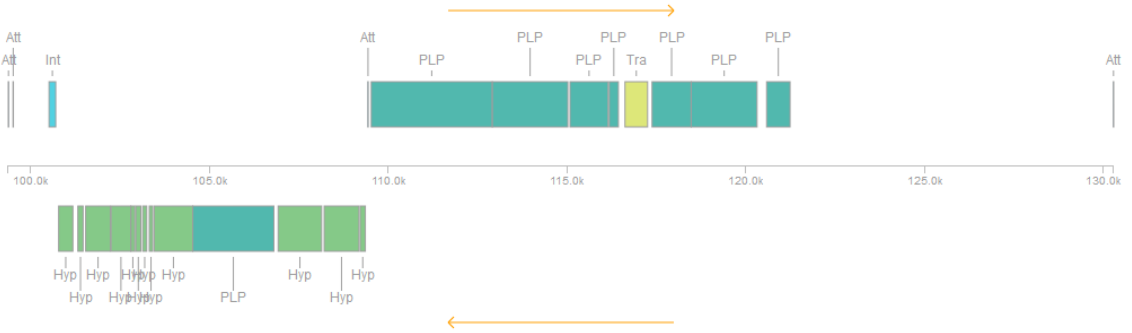

Figure 4.3: PHASTER genome viewer of phage region identified in contig 4 of isolate B. Att = attachment site, Int = integrase, PLP = phage-like protein, Tra = transposase, Hyp = hypothetical protein. Arrows indicate direction of gene.

Analysis of the hypothetical proteins and phage-like proteins identified by PHASTER did not return any domains or homologues to other proteins related to antibiotic resistance previously established (data not shown). The transposase region was identified as belonging to the transposase IS607 family, the phage-like protein preceding this region was also identified as a transposase belonging to the IS200/IS605 family, specifically OrfB. Genes surrounding this incomplete phage region were also analysed further for association with known antibiotic resistance mechanisms. No genes of interest were identified through this extended analysis.

4.3.2.3 Insertion Sequences

Both complete and partial ISs were identified in all isolates (Table 4.6), primarily belonging to the IS21 and IS200/605 families, with only one insertion sequence belonging to the IS595 family identified in E3. Complete sequences that were identified belonged to the IS200/605 family. Both the separate contigs and the rearranged and concatenated genomes (Section 4.3.4) were analysed using ISEScan to identify partial and complete ISs. Two additional ISs were identified in the rearranged and concatenated genomes of isolates B and E that were not identified when analysing the separate contigs of the respective isolates. In isolate E, the IS E6 was located over two contigs. This IS may be an artefact of the contig rearrangement in Mauve as when contigs 38 and 40 were placed next to each other different sequences were identified. Further investigation is required to elucidate whether IS E6really exists. Within isolate B, IS B2 was identified in the Mauve genome rearrangement but not within the search of the separate contigs. Unlike isolate E, this extra IS is located within the middle of a contig, not overlapping multiple contigs. Once again, further investigation is required to determine whether this is a true IS. The IS identified within Section 4.3.2.2 was also identified through this analysis (B3).

The BLASTp search (Table 4.6), conducted for both complete and partial ISs returned top hits related to the movement and replication of IS, except for E6. This analysis indicated that this protein is an OMP belonging to the Hop family, aiding in adhesion of the bacterium to gastric

tissue (Coppens et al. 2018). This suggests that this may not be a true IS and further investigation into the sequence is required; however, this is beyond the scope of this study.

The positions of the ISs in each draft genome and surrounding genes were identified to establish if they were near any antibiotic resistance determinants. Within the rearranged, draft genomes of isolates A, B, and C, no antibiotic resistant genes of interest were identified based on the RAST annotation. However, within isolate E, IS E3, E4, E5, and E6 were located near the 23S rRNA and 16S rRNA genes, both of which are involved in clarithromycin and tetracycline resistance, respectively. Although these ISs are located next to these antibiotic resistance genes, they do not appear to have caused any alterations to the genes.

Table 4.6: Insertion sequences identified from isolates A, B, C and E and BLASTp results.

Isolate IS	IS Family and Cluster	Description	Query cover	E-value	% identity	Accession number
A1	IS200/IS605_155	DNA methylase [*H. pylori*]	43%	0.00	96.79%	EJB82799.1
A2	IS21_259 ᴾ	Chromosomal replication initiator protein DnaA [*H. pylori*]	82%	0.00	99.31%	WP_128044524.1
B1	IS21_259 ᴾ	Chromosomal replication initiator protein DnaA [*H. pylori*]	96%	0.00	99.12%	WP_127945629.1
B2*	IS200/IS605_155	Eco57I restriction-modification methylase domain-containing protein [*H. pylori*]	99%	0.00	98.37%	WP_001170013.1
B3	IS200/IS605_155	Eco57I restriction-modification methylase domain-containing protein [*H. pylori*]	42%	0.00	96.11%	WP_001170013.1
C1	IS200/IS605_96 ᴾ	Transposase [*H. pylori*]	92%	6.00E-58	96.63%	WP_120901692.1
C2	IS21_259 ᴾ	Chromosomal replication initiator protein DnaA [*H. pylori*]	83%	0.00	99.56%	WP_060759823.1
C3	IS200/IS605_430 ᴾ	IS200/IS605 family transposase [*H. pylori*]	99%	3.00E-58	98.99%	WP_162966965.1
C4	IS200/IS605_96	IS200/IS605 family element transposase accessory protein TnpB [*H. pylori*]	99%	0.00	99.30%	WP_154569861.1
E1	IS21_259 ᴾ	Chromosomal replication initiator protein DnaA [*H. pylori*]	86%	0.00	99.34%	WP_128039963.1
E2	IS200/IS605_96 ᴾ	IS606 transposase [*H. pylori*]	91%	3.00E-56	93.62%	BAW74872.1
E3	IS1595_235 ᴾ	IS1595 family transposase [*H. pylori*]	64%	0.00	99.41%	WP_001969827.1
E4	IS200/IS605_449 ᴾ	IS200/IS605 family element transposase accessory protein TnpB [*H. pylori*]	74%	0.00	99.10%	WP_108262961.1
E5	IS200/IS605_96	IS200/IS605 family element transposase accessory protein TnpB [*H. pylori*]	99%	0.00	99.06%	WP_108529583.1
E6*	IS200/IS605_430	Hop family outer membrane protein HopJ/HopK [*H. pylori*]	70%	0.00	95.83%	WP_097568304.1

* = not identified in contig search. Number after underscore = IS cluster. ᴾ = Partial sequence.

4.3.3 COG Assignment

The categorisation of genes into COGs for each isolate and associated reference can be seen in Figure 4.4. Distribution of COG assignment is similar amongst all isolates and their respective references. This provides further confirmation that the isolates (A, B, C, and E) are *H. pylori*. The COGs were classified into 22 out of 26 functional categories, of which the largest groups across all isolates were function unknown (S), translation, ribosomal structure and biogenesis (J), and cell wall/membrane/envelope biogenesis (M). No COGs were identified belonging to the general function prediction only (R), extracellular structures (W), nuclear structure (Y), and cytoskeleton (Z) categories.

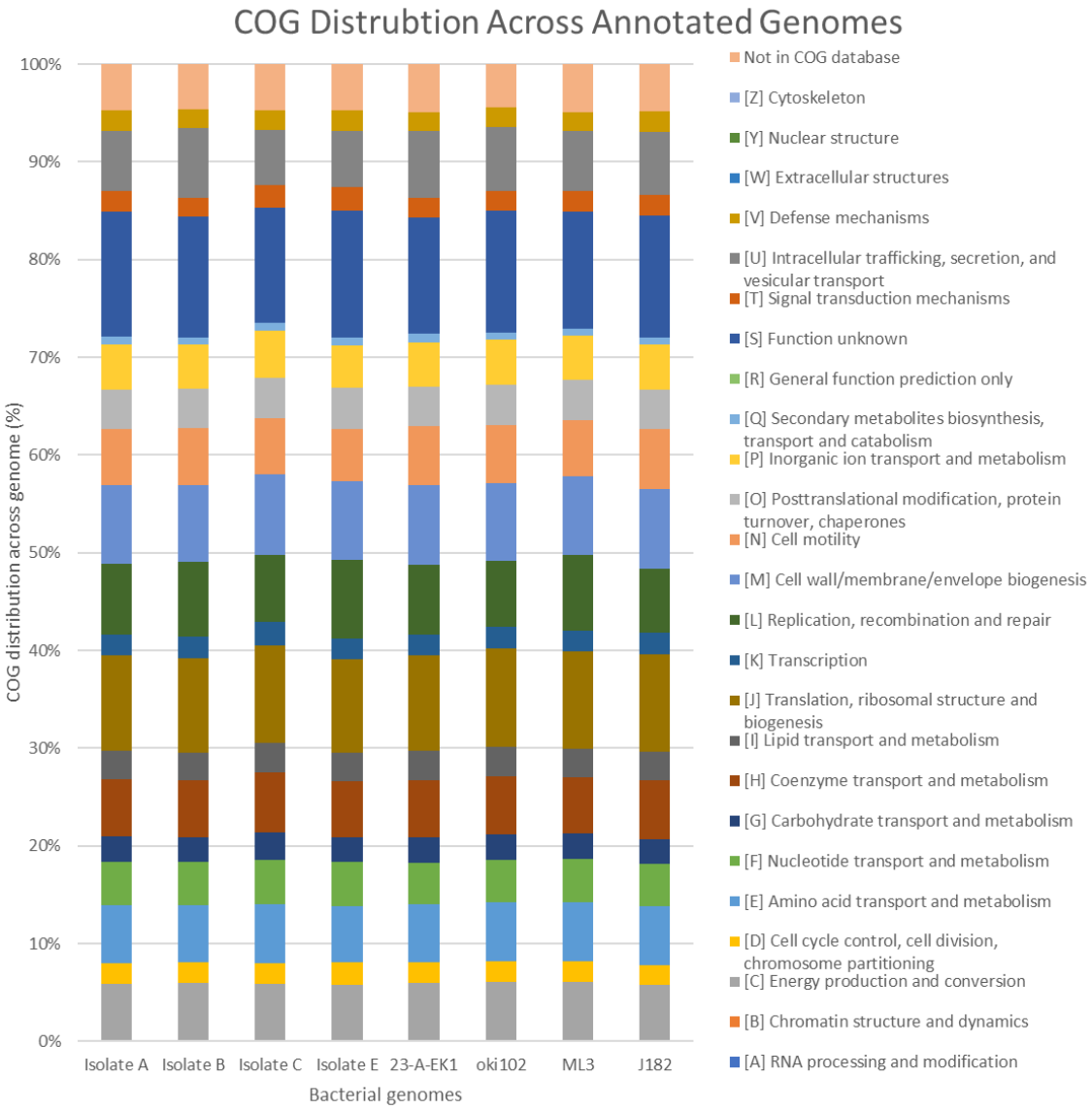

Figure 4.4: COG distribution of isolates (A, B, C and E) and reference genomes (23-A-EK1, oki102, ML3, J182).

4.3.4 Mauve Analysis

Mauve was used to reorder and align the draft genomes from this study against complete genomes of their respective closest relative identified from the nucleic acid PhyloPhlAn2 tree. The alignments for each isolate and their identified closest relative are presented in Figure 4.5. Isolate A had a minimum LCB weight of 571 nt and this parameter produced 29 LCBs (Figure 4.5 A). Nine inversions were identified in isolate A compared to its reference 23-A-EK1; however, these were small regions of the genome. Isolate B had a larger minimum LCB weight of 177,250 nt, producing two LCBs (Figure 4.5 B). The two LCBs indicate the presence of one large homologous region (green) and then one smaller homologous region (red). Within the middle of the green LCB in isolate B, is a white region, this indicates that isolate B is missing a number of genes that the reference genome (oki102) has within this genome section. These white regions are also observed in the reference genome (oki102) in both the green and red LCBs but are much smaller. This indicates that isolate B has additional genes that oki102 does not have based on nucleotide sequence homology. The alignment of isolate C against its closest relative produced one LCB and therefore no minimum LCB weight (Figure 4.5 C). There are two large white gaps observed in the reference genome (ML3) that indicate additional genes are present in isolate C that are not present within *H. pylori* ML3. Finally, isolate E had a minimum LCB weight of 209 nt, producing 19 LCBs (Figure 4.5 D). Six inversions were observed when compared to its reference (J182). The gaps within the isolates (A, B, C, and E) may be due to the incomplete nature of the genomes. However, these gaps were examined further, and no known antibiotic resistant determinants were identified. Gaps present in the reference genomes, which are complete, were also analysed further but no known antibiotic resistance genes were identified.

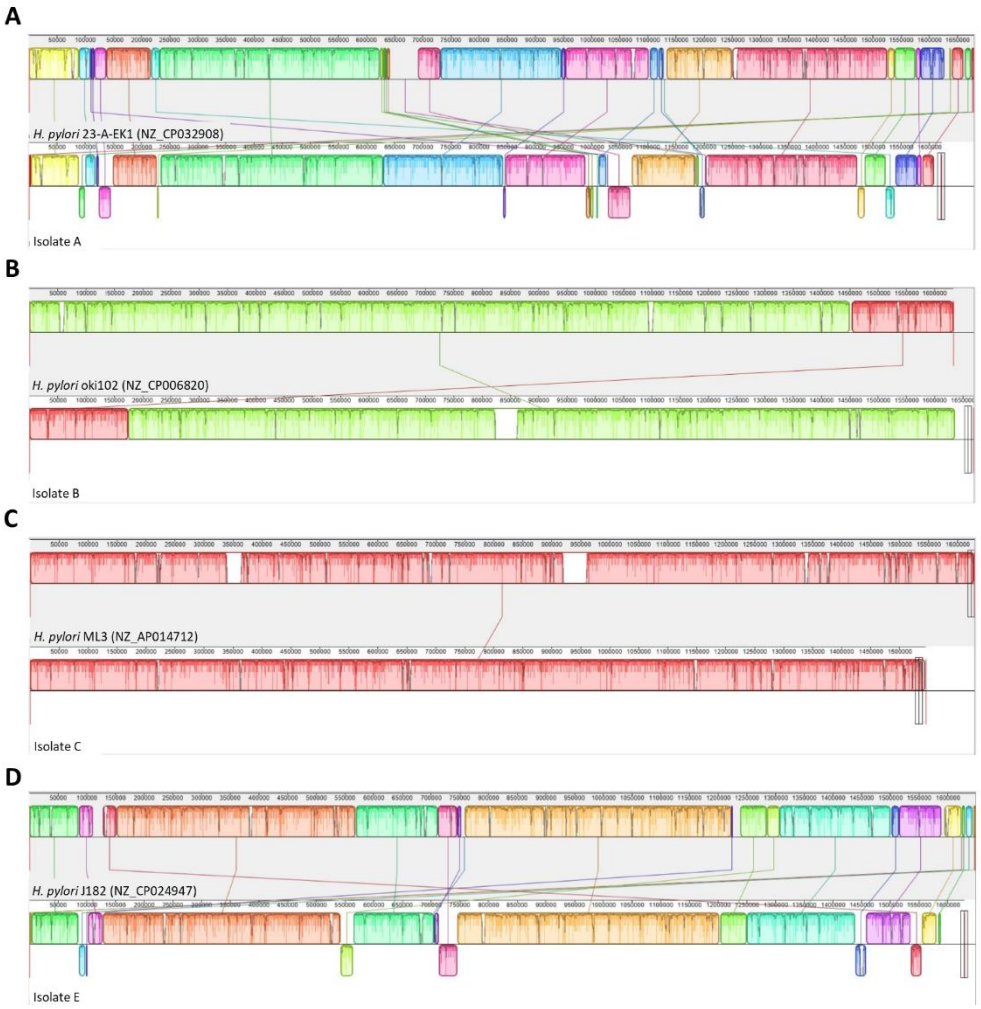

Figure 4.5: Mauve alignments for isolates A (A), B (B), C (C) and E (D) with their respective reference genomes. Coloured blocks represent LCB (homologous regions). Blocks below the line of the isolate genomes indicate an inverted sequence compared to their reference.

4.3.5 Virulome

The presence of potential virulence genes within the genomes each isolate (A, B, C, and E) was assessed using both VFDB and Nullarbor analyses. The full virulome report from Nullarbor and the VFDB may be viewed in Appendices Appendix D: Nullarbor Report and Appendix E: VFDB Output. These two approaches identified different genes with Nullarbor identifying 112 genes across all isolates and the VFDB identifying 95 genes across all isolates; with 73 genes identified in common. A summary of the most clinically relevant virulence factors, for *H. pylori*, are presented in Table 4.7. All isolates possess the genes for the *cagA*, *vacA*, urease enzymes and major OMPs. None of the isolates appear to possess the gene *dupA*, which is involved with the development of duodenal ulcers (Talebi Bezmin Abadi and Perez-Perez 2016). Isolates B, C, and E all possess the *iceA* gene; however, it was not identified in isolate A by these analyses. Absence of the *iceA* gene in isolate A was confirmed by a subsequent BLAST search within the genome. These major virulence factors were identified using both methods.

Table 4.7: Summary of clinically relevant virulence factors identified in isolates A, B, C and E from the VFD and Nullarbor.

Virulence gene	Gene product	Presence/absence of gene in isolates			
		A	B	C	E
cagA	Cytotoxin-associated gene A (CagA)	✓	✓	✓	✓
vacA	Vacuolating cytotoxin (VacA)	✓	✓	✓	✓
dupA	Duodenal ulcer promoting gene (DupA)	×	×	×	×
iceA	Induced by contact with epithelium gene A (IceA)	×	✓	✓	✓
Urease (*ureA*, *ureB*)	Urease alpha and beta subunits (Urease)	✓	✓	✓	✓
OMPs (*babA*, *sabA*, *oipA*)	Blood-group-antigen-binding adhesin (BabA), Sialic-acid-binding adhesin (SabA), Outer inflammatory protein A (OipA)	✓	✓	✓	✓

✓ = present, × = absent, OMP = outer membrane protein.

4.3.6 Metabolic Potential

4.3.6.1 Carbohydrate Metabolism

Pathway analysis indicates that the draft genomes from the *H. pylori* isolates in this study each possess an incomplete glycolysis pathway (Figure 4.6). This suggests that the isolates are unable to convert glucose into important metabolites using this metabolic pathway alone. These results are consistent with previous studies of *H. pylori* glucose metabolism (Dandekar et al. 1999; Mendz and Hazell 1991).

All *H. pylori* isolates appear to lack the standard genes required to uptake extracellular D-glucose. However, a putative glucose/galactose transporter (*gluP*) has been identified within the *H. pylori* 26695 genome (Tomb et al. 1997). Through a BLAST search of all four isolates, this gene was present suggesting that the isolates may use this channel to import extracellular glucose or galactose.

Once D-glucose or galactose enters the cell, it may undergo phosphorylation using glucokinase (EC 2.7.1.2) to form D-glucose-6P. A reversable reaction using glucose-6-phosphate isomerase (EC 5.3.1.9) can then convert D-glucose-6P into D-fructose-6P. However, all isolates lack two key enzymes required to complete the conversion of glucose into pyruvate; 6-phosphofructokinase (EC 2.7.1.11) and pyruvate kinase (EC 2.7.1.40). This lack of enzymes would force the bacterium

to utilise an alternative pathway such as the Entner-Doudoroff or nonoxidative pentose phosphate pathway to complete the breakdown of glucose (green arrows; Figure 4.7). D-glucose-6P enters the Entner-Doudoroff pathway and is converted to D-glucononate-6P. The isolates would then use their enzymes capable of dehydrating and cleaving D-gluconate-6P to produce pyruvate: phosphogluconate dehydratase (EC 4.2.1.12) and 2-dehydro-3-deoxyphosphogluconate aldolase (EC 4.1.2.14). Through the nonoxidative pentose phosphate pathway (blue arrows; Figure 4.7), D-fructose-6P can potentially be converted into D-ribose-5P through a series of reversable reactions. D-ribose-5P may then be used for the synthesis of nucleic acids.

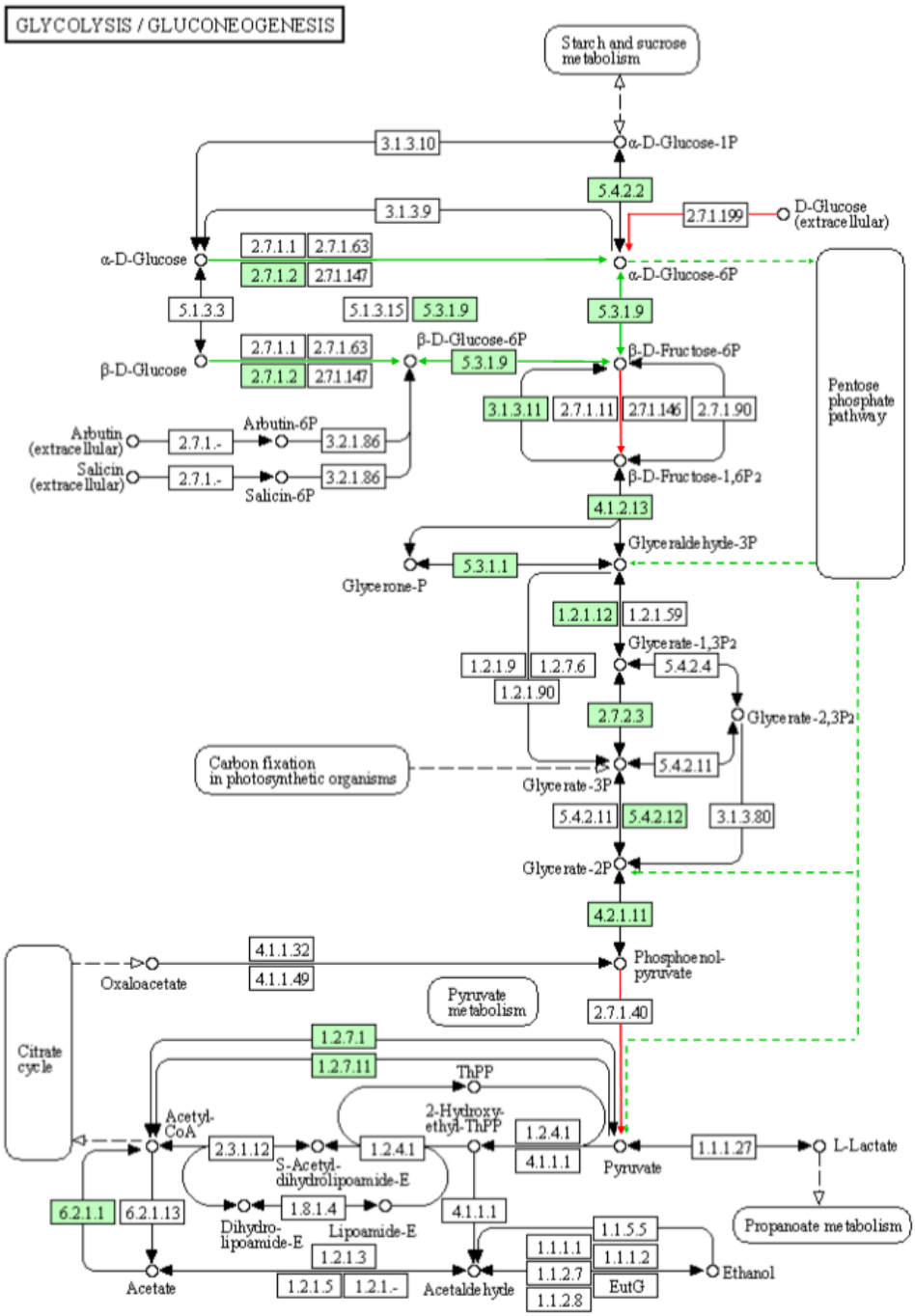

Figure 4.6: Glycolysis pathway for isolates A, B, C, and E. Green EC numbers indicate presence of the enzyme, white EC numbers indicate enzymes were not identified. Green arrows indicate potential flow of the pathway. Dashed arrows indicate that a substrate is either from or entering a different pathway. Red arrows indicate missing genes that inhibit pathway.

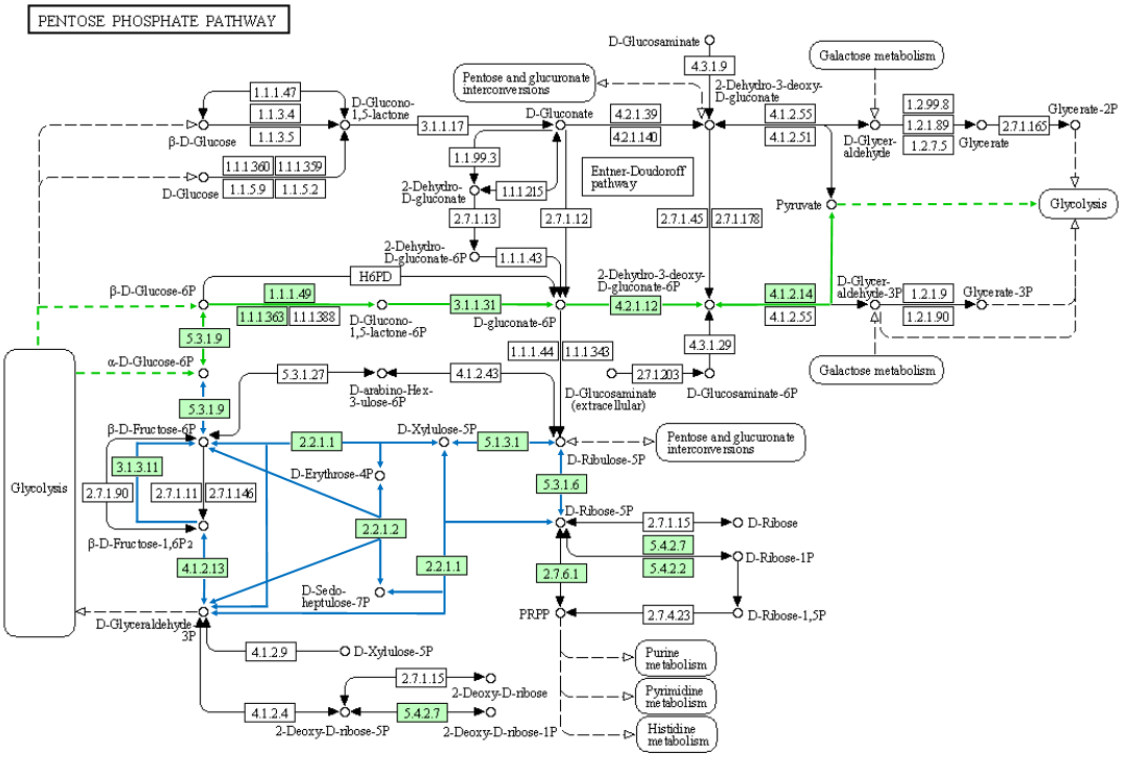

Figure 4.7: Pentose phosphate and Entner-Doudoroff pathways for isolates A, B, C, and E. Green EC numbers indicate presence of the enzyme, white EC numbers indicate enzymes were not identified. Green arrows indicate potential flow of pathway. Blue arrows indicate nonoxidative pentose phosphate pathway. Dashed arrows indicate that a substrate is either from or entering a different pathway.

4.3.6.2 Citric acid cycle

The KEGG citric acid cycle (CAC) identified in the isolates (A, B, C, and E) appears to differ from the 'standard' CAC (Figure 4.8). The absence of the key enzyme succinyl-CoA synthase (EC 6.2.1.5; red arrows; Figure 4.8), required to convert succinyl-CoA into succinate, makes this an incomplete, noncyclic pathway. It has been suggested that *H. pylori* possess a branched CAC, consisting of a dicarboxylic and tricarboxylic arm (green arrows; Figure 4.8), due to the absence of this key enzyme. The isolates possess the genes encoding pyruvate ferredoxin oxidoreductase alpha subunit (EC 1.2.7.1) and 2-oxoacid oxidoreductase (EC 1.2.7.11), to convert pyruvate to acetyl-coenzyme A (CoA). Through the tricarboxylic arm of the pathway, the isolates possess the genes encoding enzymes required to convert oxaloacetate to citrate, isocitrate and α-ketoglutarate. The enzyme α-ketoglutarate dehydrogenase (EC 1.2.4.2), used to convert α-ketoglutarate to succinyl-CoA was not identified in the isolates; however, 2-oxoglutarate synthase (EC 1.2.7.3) and 2-oxoacid oxidoreductase are present instead. In the dicarboxylic branch of the pathway, malate synthase (EC 2.3.3.9) or malate dehydrogenase (EC 1.1.1.37; orange arrows; Figure 4.8) are not present to convert oxaloacetate into malate. However, activity of this enzyme has been observed in other *H. pylori* isolates even though these enzymes have not been identified in the genome of other isolates (Pitson et al. 1999; Tomb et al. 1997). It has been suggested that malate:quinone oxidoreductase (EC 1.1.1.54) is responsible for the

malate dehydrogenase activity exhibited in *H. pylori* (Kather et al. 2000). Malate:quinone oxidoreductase was identified in isolates A, B, C, and E. The isolates do possess genes encoding fumarase (EC 4.2.1.2) and fumarate reductase (EC 1.5.3.4) to convert malate into fumarate and succinate, completing the dicarboxylic branch of the CAC. Overall, the potential CAC identified in isolates A, B, C, and E is similar to that of other *H. pylori* isolates; however, further investigation is required to identify the genes encoding enzymes required to convert oxaloacetate to malate.

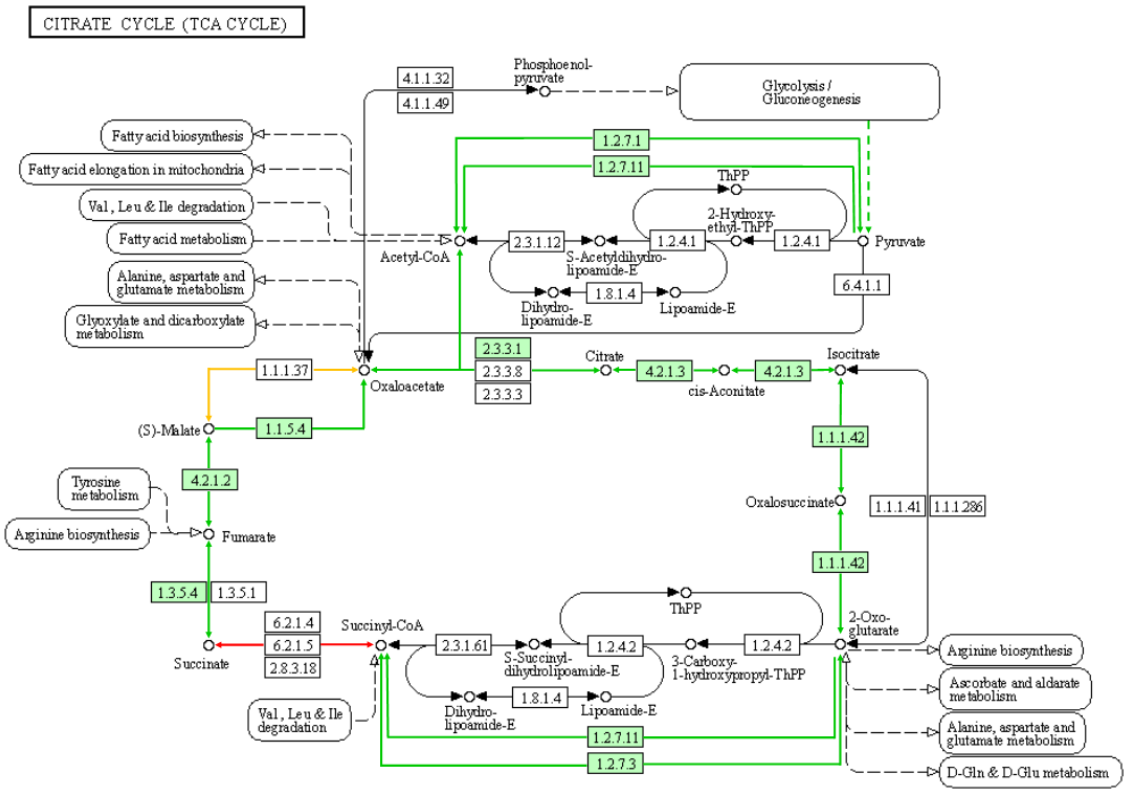

Figure 4.8: Citric acid cycle for isolates A, B, C, and E. Green EC numbers indicate presence of the enzyme, white EC numbers indicate enzymes were not identified within the genomes. Green arrows indicate potential flow of pathway. Dashed arrows indicate that a substrate is either from or entering a different pathway. Orange arrow indicates gene is missing from genome however, activity has been identified in other *H. pylori* isolates. Red arrow indicates missing gene that inhibits pathway.

4.3.6.3 Nitrogen metabolism

Nitrogen is utilised in *H. pylori* to synthesise key components required for the formation of DNA, proteins and other important molecules (Price et al. 2002). In the gastric environment there are two main sources of nitrogen for *H. pylori* to utilise: amino acids and urea (Weeks et al. 2000).

Urea is present in small amounts within the stomach of humans (Weeks et al. 2000). *H. pylori* are able to take up this extracellular urea through an acid activated urea channel, encoded by the *ureI* gene. According to the KEGG annotation, only isolate B, C, and E possess the *ureI* gene; however further analysis through a BLAST search and annotation by VFDB shows that all isolates

(A, B, C, and E) possess this urea channel. The VFDB analysis is described above (Section 4.3.5). All isolates also possess urease (EC 3.5.1.5), which is able to hydrolyse urea to ammonia and carbon dioxide, once inside the cell.

Amino acids are also present within the gastric environment and provide *H. pylori* with a source of both nitrogen and carbon (Komorowska et al. 1981). There are several amino acids that are required for the growth of *H. pylori*; arginine, histidine, isoleucine, leucine, methionine, phenylalanine and valine; with some strains also requiring alanine and serine (Reynolds and Penn 1994). Based on genomic data alone, the isolates from this study are unable to synthesise arginine, histidine, isoleucine, leucine, methionine, phenylalanine, valine, alanine and serine; as they are missing key enzymes in the biosynthetic pathways. However, isolates A, B, C, and E possess a number of genes that aid in the catabolism of some amino acids that produce ammonia as a by-product (Table 4.8).

Table 4.8: Enzymes present in isolates A, B, C, and E that are involved in catabolism of amino acids.

Gene name	Enzyme	EC number	Reaction
ald	Alanine dehydrogenase	1.4.1.1	Converts alanine to pyruvate + ammonia
aspA	Aspartate-ammonia lyase	4.3.1.1	Converts aspartate to fumarate + ammonia
ansB	L-asparaginase II	3.5.1.1	Converts L-asparagine to L-aspartate + ammonia
gdhA	Glutamate dehydrogenase	1.4.1.4	Reversible reaction converting α-ketoglutarate + ammonia to glutamate
glnA	Glutamine synthetase	6.3.1.2	Converts glutamate + ammonia to glutamine
sdaA	L-serine deaminase	4.3.1.17	Converts serine to pyruvate + ammonia
rocF	Arginase	3.5.3.1	Converts arginine to urea
amiF	Formanidase	3.5.1.49	Converts formamide to formate + ammonia
amiE	Aliphatic amidase	3.5.1.4	Hydrolysis of short-chain amides to carboxylic acid + ammonia
dadA	D-amino acid dehydrogenase	1.4.5.1	Oxidation of D-amino acids

4.3.7 Manual Curation of Known and Potential Antibiotic Resistance Determinants from *H. pylori*

To assess whether isolates A, B, C, and E possess any known mechanisms conferring antibiotic resistance, a search of known resistance genes and mutations was conducted. As the isolates were only phenotypically resistant to clarithromycin and metronidazole, only mechanisms associated with resistance to these drugs were included.

4.3.7.1 Clarithromycin Resistance

The mechanism of action for clarithromycin involves the reversible binding of the drug to domain V of the 23S rRNA (Kanoh and Rubin 2010). Domain V of the 23S rRNA gene is therefore of keen interest in resistance to this macrolide. Through alignment of the 23S rRNA genes, from the isolates and their respective reference genomes, the A2147G mutation (red box; Figure 4.9) was identified in isolates B and C, and reference genomes 23-A-EK1 and ML3. The A2146G mutation was not identified in any of the isolates; however, two other mutations were identified in domain V, T2186C (isolates C, E, and ML3; green box, Figure 4.9) and A2227G (isolates C and E; orange box, Figure 4.9).

Of the three mutations identified, two of the isolates (C and E), had multiple mutations within domain V. Isolate C had three mutations, A2147G, T2186C and A2227G; and isolate E had two mutations, T2186C and A2227G. These mutation combinations have been identified before; however, the A2147G, T2186C and A227G combination has been associated with a clarithromycin susceptible strain (Tran et al. 2019).

The G160A mutation within the *infB* gene or the 9 bp insertion or 3 bp deletion within *rplI22* were not identified in any of the isolates or their respective reference genomes (data not shown). These results suggest that target alteration of the 23S rRNA gene is a contributor to the resistance of isolates B and C for clarithromycin resistance. However, the mechanism of clarithromycin resistance for isolate E is still not defined, based on these known mutations, suggesting another mechanism is involved in resistance.

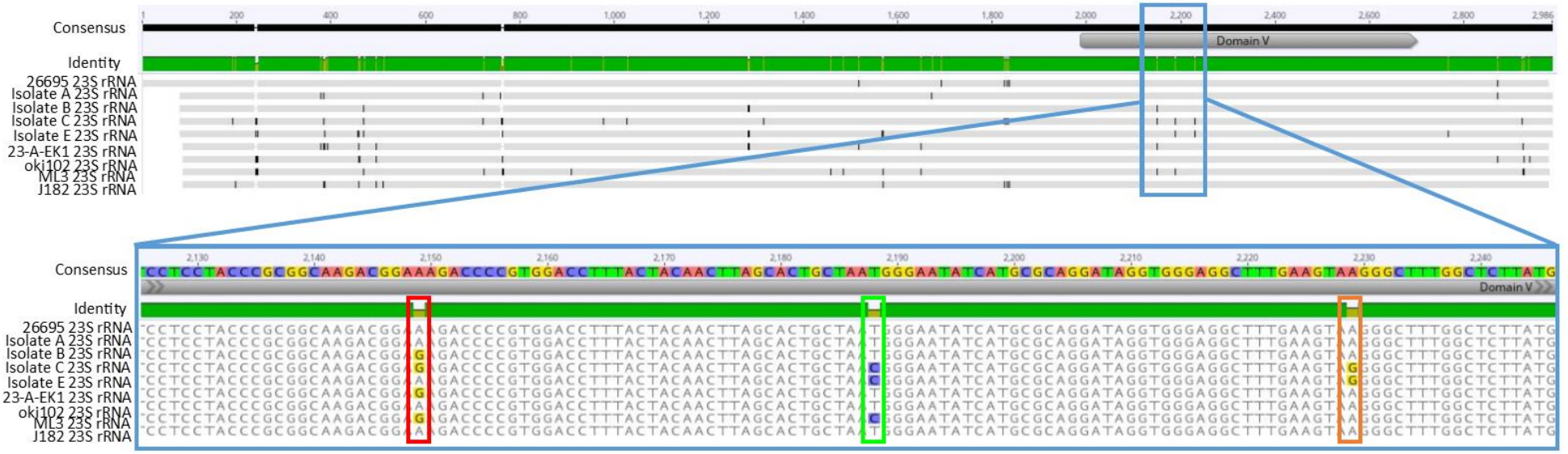

Figure 4.9: MUSCLE multiple sequence alignment of 23S rRNA genes from isolate 26695 (reference genome), isolates (A, B, C, and E) and their respective closest relatives (23-A-EK1, oki102, ML3, and J182). Blue box shows a zoomed in view of the mutations present in domain V of the 23S rRNA. The three mutations present within this domain are A2147G (red box), T2186C (green box), and A2227G (orange box).

4.3.7.2 Metronidazole Resistance

As metronidazole is a prodrug it requires reduction once within the cell to release its bactericidal components. Mutations within key reductases genes in *H. pylori* (*rdxA* and *frxA*) are commonly associated with metronidazole resistance (Marais et al. 2003). A number of mutations were identified in both reductase genes in all isolates and reference genomes. Within isolates B and E, two missense mutations in the *rdxA* gene were identified leading to R16H and A118S (red and blue boxes respectively; Figure 4.10 A) amino acid substitutions in the RdxA protein. Isolate B contained both substitutions whereas isolate E only contained the A118S amino acid substitution. Isolate A also possessed a missense mutation causing at position 118, causing a change from alanine to threonine (A118T); however, this isolate is metronidazole sensitive (blue box; Figure 4.10 A). Mutations in these positions have been identified in other metronidazole resistant strains of *H. pylori* before (Kwon et al. 2000c). No mutations were found in any of the reference genomes in these positions.

Mutations were also identified in the *frxA* gene in isolates A, B, C, E, and the reference genome 23-A-EK1. The three metronidazole resistant strains, isolates B and E, and the reference genome 23-A-EK1 did not share any common mutations within the *frxA* gene. Figure 4.10 B shows the consequences of the mutations within the *frxA* gene. Isolate B possesses a 32 amino acid deletion at the start of FrxA (black box; Figure 4.10 B). A 20 amino acid deletion was observed at position 189 in this protein in strain 23-A-EK1 (yellow box; Figure 4.10 B). Isolate C also had a number of mutations within the *frxA* gene, including creation of a stop codon at amino acid position 207 (orange box; Figure 4.10 B) leading to premature truncation of this protein. However, this isolate is not phenotypically resistant to metronidazole. No major mutations were identified in *frxA* of isolate E.

No premature truncation or significant mutations shared between resistant isolates were identified in Fur or RecA (data not shown).

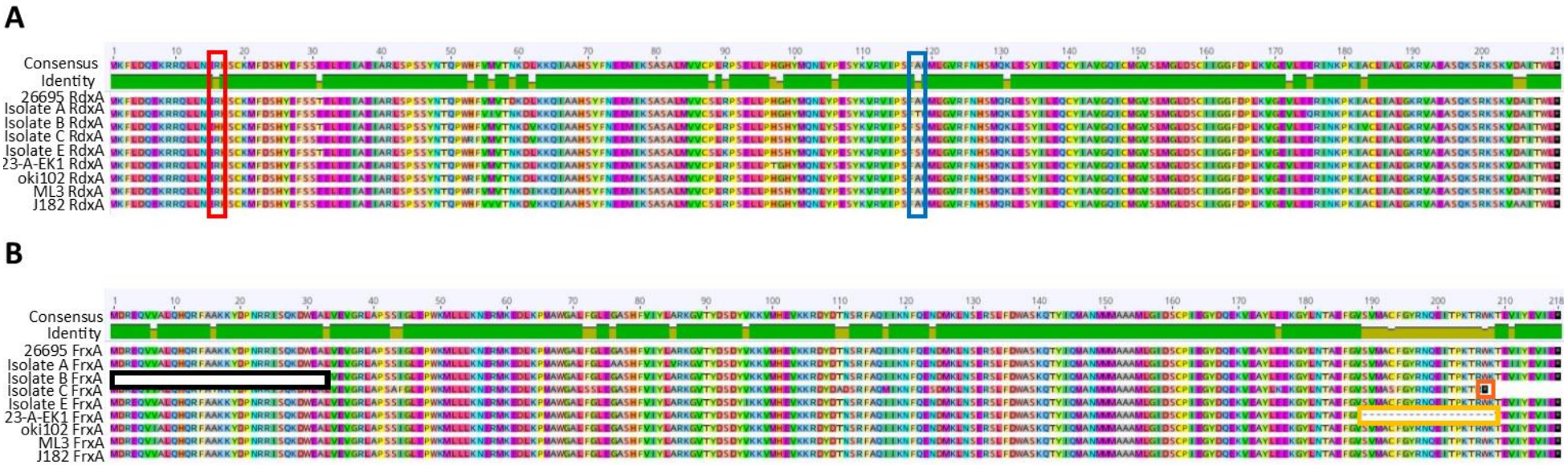

Figure 4.10: MUSCLE protein alignments of RdxA (A) and FrxA (B) from isolates A, B, C, E, closest relative genomes 23-A-EK1, oki102, ML3, and J182, and reference genome 26695. Red box highlights R16H mutation within RdxA. Blue box highlights A118 mutations within RdxA. Black box highlights 32 AA deletion in FrxA of isolate B. Yellow box highlights 20 AA deletion at position 189 in FrxA of 23-A-EK1.

4.3.7.3 Efflux pumps

The efflux pump genes *hefABC*, *hp1184* and *hp1181* were identified in all isolates and their respective reference genomes. Alignment of the translated sequences revealed a number of mutations within all isolates and their respective closest relatives. However, there is limited data available on the resistance profiles of the closest relatives, with only 23-A-EK1 known to be clarithromycin resistant. Therefore, inference on the effects these mutations may have on conferring antibiotic resistance can only be made in reference to the isolates from this study (A, B, C, and E).

Within the RND-type efflux pump genes, hefABC, identified in all isolates, a number of mutations were identified compared to the reference sequence from 26695 (Table 4.9). Overall, resistant isolates had a higher number of missense mutations than the sensitive strain (isolate A). However, in *hefC* both isolate E and 23-A-EK1 shared the same number of missense mutations with isolate A. For each efflux pump gene, mutations shared amongst resistant isolates only were identified. A complete list of the consequences of the missense mutations within each efflux pump may be viewed in Appendix F: Efflux SNP Analysis. Within *hefA* only one missense mutation was present that was common amongst isolates B, C, and E, causing the amino acid change I304L in HefA (outer membrane protein); this mutation was also present in ML3. No mutations were present in only isolates B and E, that share the same resistance profile (> 256 µg/mL MIC towards clarithromycin and metronidazole). Within *hefB*, no mutations were shared between isolates B, C, and E only; however, two missense mutations were shared only be tween isolates B and E leading to amino acid substitutions in HefB (membrane fusion protein), G104S and T108N; both of which were also present in ML3. Within *hefC* one missense mutation was identified in common between isolates B, C, and E, leading to an amino acid substitution in HefC (cytoplasmic pump protein), Q615E, this change was also observed in oki102 and ML3. Two other missense mutations, common amongst isolates B and E were also identified leading to V330I and H614G changes, of which the former was also identified in isolates 23-A-EK1 and oki102. The N177T missense mutation within the *hefA* gene, encoding a protein potentially involved in clarithromycin efflux (HefA), was identified in isolates A, B, C, and E, as well as, 23-A-EK1, ML3 and J182. However, the presence of this mutation in the sensitive isolate (A) suggests that this mutation may not be involved with clarithromycin efflux.

Table 4.9: Number of mutations identified in HefABC of isolates A, B, C, and E and their respective closest relatives.

	Resistance profile	Number of mutations		
		HefA	HefB	HefC
Isolate A	Sensitive	16	2	14
Isolate B	CLR and MTZ resistant	18	4	16
Isolate C	CLR resistant	18	5	30
Isolate E	CLR and MTZ resistant	23	6	14
23-A-EK1	CLR resistant	30	3	14
oki102	Unknown	3	2	24
ML3	Unknown	16	7	33
J182	Unknown	16	3	14

CLR = clarithromycin. MTZ = metronidazole.

4.3.8 Novel Antibiotic Resistance Determinants

Due to differences in known mutations and MICs produced by isolates, as well as an absence of known antibiotic resistance mechanism identified in isolate E, a search for novel resistance determinants was conducted using CARD and OrthoVenn2.

4.3.8.1 CARD

To identify other antibiotic resistance genes related to the nitroimidazole and macrolide, isolates A, B, C, and E were analysed using CARD, through manual analysis as well as in the Nullarbor report. According to the Nullarbor output, isolates A, B, C, and E do not have a resistome. Through manual analysis, the RGI identified 32 genes associated with the drug classes above, across all four isolates. Most of these genes returned a single hit from the isolate genomes. However, *macB*, *msbA*, *adeR*, *tet(58)* and *tet44* all returned multiple hits (Table 4.10). No resistance genes were identified on the plasmids of isolates A, C, and E.

Table 4.10: CARD hits with multiple copies identified.

Isolate	Gene, gene product, and number of hits				
	macB	***msbA***	***adeR***	***tetA(58)***	***tet44***
	Macrolide export ATP-binding/permease (MacB)	Lipid A export ATP-binding/permease protein (MsbA)	Efflux system DNA-binding response regulator (AdeR)	Tetracycline efflux pump (TetA(58))	Ribosomal tetracycline resistance protein (Tet(44))
Isolate A	7	2	2	2	2
Isolate B	7	3	2	2	2
Isolate C	8	3	2	2	2
Isolate E	7	3	2	2	2

Genes identified from the RGI output relating to the drug classes above, are compiled in Appendix G: RGI Output where percentage of amino acid similarity to CARD sequences may be viewed. No perfect hits were identified, only strict and loose hits. Twenty-one genes were identified in association with tetracycline resistance. However, as none of the isolates (A, B, C, and E) were phenotypically resistant to tetracycline, these genes were not studied further. CARD hits relating to macrolide and tetracycline resistance were visualised using a heatmap (Figure 4.11). From this heatmap, no one gene is associated with macrolide or nitroimidazole resistance across all isolates. However, it does highlight two genes with high sequence similarities associated with macrolide and metronidazole resistance; the 23S rRNA mutation (A2147G) and *hp1181*, respectively (dark blue bars; Figure 4.11). These resistance mechanisms have high similarity with submitted sequences in CARD, which was above 97 %. The other resistance mechanisms identified have a range of similarity scores between 23.18 % (*lpeB* (subunit of LpeAB efflux pump, LpeB); isolate B) and 44 % (*lsaC* (ABC-F subfamily protein, LsaC); isolate B; Appendix G: RGI Output).

In total, there were 15 genes identified through CARD analysis, across isolates A, B, C, and E associated with macrolide resistance. The 23S rRNA mutation, A2147G, was identified in both isolates B and C but absent in isolates A and E; this is in agreement with what was observed in Section 4.3.7. No genes were identified that were common amongst isolates B, C, and E only (all resistant towards clarithromycin) that were associated with macrolide resistance. Two genes were identified only in isolates B and E, which share the same MIC towards clarithromycin (> 256 µg/mL), *ermY* and *mexJ*. The *ermY* gene was originally identified in a plasmid from *Staphylococcus aureus* and encodes a protein (ErmY) associated with antibiotic target alteration, which results in methylation of the 23S rRNA gene at position 2058 (*E. coli* 23S rRNA position)

(Matsuoka et al. 1998). The *mexJ* gene encodes the membrane fusion protein (MexJ) of the MexJK efflux pump originally identified in *Pseudomonas aeruginosa* (Chuanchuen et al. 2002). These two genes share low similarity scores at the amino acid level of 25.84 % (isolate B) and 25.11 % (isolate E) with ErmY, and 23.65 % (isolate B) and 25.12 % (isolate E) for MexJ.

Only two genes were identified in the CARD analysis that were associated with nitroimidazole resistance, *msbA* and *hp1181*. These two genes were present in all isolates (A, B, C, and E). The *hp1181* hit had a high sequence similarity score of ~97 %, at the amino acid level, with the HP1181 entry within CARD. This finding is in accordance with Section 4.3.7.3 where *hp1181* was also found in all isolates. However, the *msbA* gene had multiple hits within each isolate (Table 4.10) with a range of percentage similarities, at the amino acid level, from 27.31 % (isolate C) to 32.89 % (isolate B). The *msbA* gene encodes a multidrug resistance transporter protein homolog (MsbA) initially characterised in *E. coli* (Zhou et al. 1998). This gene is essential for cell viability due to its transportation of lipid A, an integral component of the outer membrane of Gram-negative bacteria (Ghanei et al. 2007; Zhou et al. 1998). This transporter was identified in all isolates, likely due to its integral nature.

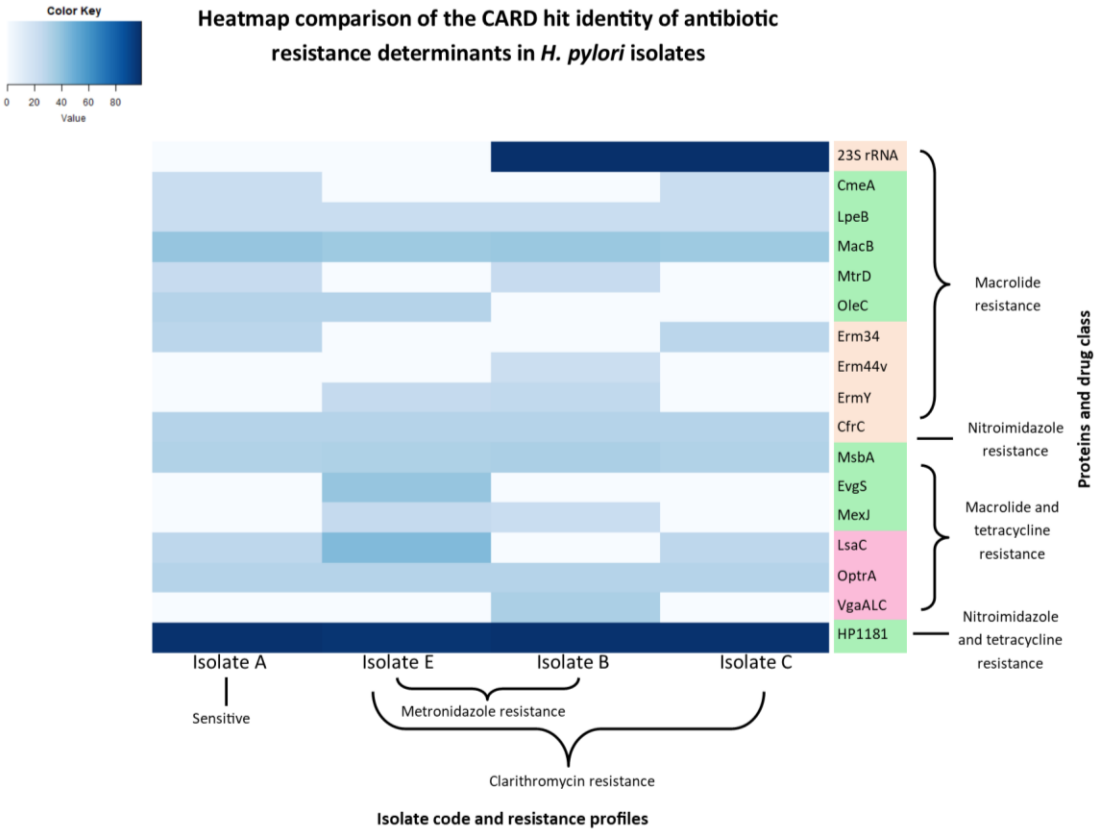

Figure 4.11: Heatmap comparisons of CARD hits and percentage identity of resistance proteins associated with macrolides and nitroimidazoles for isolates A, B, C, and E. The blue colour key indicates the amino acid % similarity between CARD gene hits and the isolates. Protein names are highlighted in coordination with their resistance mechanism. Orange = antibiotic target alteration. Green = antibiotic efflux. Pink = antibiotic target protection.

Further analysis of the genes identified through CARD with InterPro and BLASTp searches was conducted due to the low similarity scores to further elucidate the function of these genes. An analysis of all predicted amino acid sequences of the genes from Figure 4.11 may be viewed in Appendix H: CARD Hit Analysis. From the InterPro search results, determinants associated with antibiotic target alteration were identified associated with methyltrasefearse families, domains and conserved sites. However, both the determinants associated with antibiotic efflux and antibiotic target protection were involved with efflux pump families, domains, and conserved sites. The efflux pump determinants were associated with the RND and MFS superfamilies, whereas, the antibiotic target protection determinants were involved with the ABC superfamily. The BLASTp search, conducted through the NCBI database, provided another layer of evidence towards the function of the genes identified in isolates A, B, C, and E. Overall, the BLASTp results showed a similar pattern to the InterPro results where determinants associated with antibiotic target alteration were identified as methryltransferases and determinants associated with antibiotic efflux or target protection were associated with antibiotic efflux proteins belonging to the RND, MFS or ABC superfamilies (Appendix H: CARD Hit Analysis).

The InterPro and BLASTp results of the two proteins of interest, ErmY and MexJ, shared only between isolates B and E can be viewed in Table 4.11. The ErmY hit was identified as an RNA adenine methyltransferase through InterPro suggesting that this gene is involved in methyltransferase activity in RNAs. However, through the BLASTp search, this gene was identified as the ribosomal RNA small subunit methyltransferase, RsmA, from *H. pylori* with an E-value of 0. This result suggests that this gene does possess methyltransferase activity but is likely involved with small subunit rRNA instead of large subunit rRNA.

The MexJ hit was identified as belonging to the RND efflux, membrane fusion protein family by InterPro. The BLASTp search identified this sequence as HefB, the periplasmic subunit adaptor of an RND-type efflux pump with E-values of 7.00×10^{-157} (Table 4.11; isolate B) and 9.00×10^{-160} (Table 4.11; isolate E). However, from Section 4.3.7.3 the *hefB* gene was identified in all isolates, not just in the high-level clarithromycin and metronidazole resistant strains. The BLASTp analysis of the other hits from CARD identified that CmeA, only present in isolates A and C, was HefB with E-values of 2×10^{-156} and 9×10^{-159}, respectively (Appendix H: CARD Hit Analysis). Based on the InterPro results, both the MexJ and CmeA hits belonged to the RND efflux pump, membrane fusion protein family (IPR006143). However, the MexJ sequences (isolates B and E) had a barrel-sandwich domain (IPR032317), whereas, the CmeA sequences (isolates A and C) had a biotin-lipoyl like domain (IPR039562).

Table 4.11: InterPro and BLASTp searches of CARD gene hits of interest.

CARD hit	Prokka Annotation		InterPro Search	BLASTp Search		
	Locus tag	Description		Description	E-value	Accession number
ErmY	KAEGCKPN_01528	Ribosomal RNA small subunit methyltransferase A	Ribosomal RNA adenine methyltransferase KsgA/Erm (IPR001737) Ribosomal RNA adenine dimethylase (IPR011530) Ribosomal RNA adenine methylase transferase, N-terminal (IPR020598) rRNA adenine dimethylase-like (IPR023165) S-adenosyl-L-methionine-dependent methyltransferase (IPR029063) Ribosomal RNA adenine methylase transferase, conserved site (IPR020596)	*H. pylori* 16S rRNA (adenine(1518)-N(6)/adenine(1519)-N(6))-dimethyltransferase (RsmA)	0	WP_000259408.1
	CBGCIAIG_01460					WP_000259409.1
MexJ	KAEGCKPN_01024	Hypothetical protein	RND efflux pump, membrane fusion protein (IPR006143) RND efflux pump, membrane fusion protein, barrel-sandwich domain (IPR032317)	*H. pylori* efflux RND transporter periplasmic adaptor subunit (HefB)	7.00×10^{-157}	WP_096470046.1
	CBGCIAIG_00869				9.00×10^{-160}	WP_060869471.1
CmeA	INKADMOK_00778	Hypothetical protein	Membrane fusion protein, biotin-lipoyl like domain (IPR039562)	*H. pylori* efflux RND transporter periplasmic adaptor subunit (HefB)	2.00×10^{-156}	WP_108334415.1
	DILABLLM_00883		RND efflux pump, membrane fusion protein (IPR006143), Membrane fusion protein, biotin-lipoyl like domain (IPR039562)		9.00×10^{-159}	WP_058063141.1

INKADMOK locus tag = isolate A. KAEGCKPN locus tag = isolate B. DILABLLM locus tag = isolate C. CBGCIAIG locus tag = isolate E.

4.3.8.2 OrthoVenn2

OrthoVenn2 was used in an attempt to identify other proteins that may be involved in antibiotic resistance. Comparison of isolates A, B, C, and E produced the Venn diagram in Figure 4.12. This Venn diagram shows that the isolates share 1,334 OPCs. From the comparisons between isolate clusters described in Section 4.2.6.2 produced 69 clusters that were subject to further analysis through InterPro and NCBI BLASTp databases to help determine the function of these protein clusters and any potential involvement in antibiotic resistance (Appendix I: OrthoVenn2 Protein Analysis). Of these 69 clusters, none were present on the plasmids of isolates A, C, or E. A number of the proteins were not able to be identified and remain annotated as hypothetical proteins. Other proteins were suggested as possessing activity of methyltransferases, transposases and endonucleases (Table 4.12). Two clusters identified, 1412 and 1448, are of interest due to the InterPro and BLASTp results indicating they belong to the MFS transporter family and outer membrane protein family, respectively (Table 4.12). These clusters were not present in isolate E; cluster 1412 was only identified in isolate B, whereas cluster 1448 was identified in isolates B and C. However, these proteins, likely associated with efflux, may be involved in antibiotic efflux but further research is required to fully elucidate their function and contribution to antibiotic resistance.

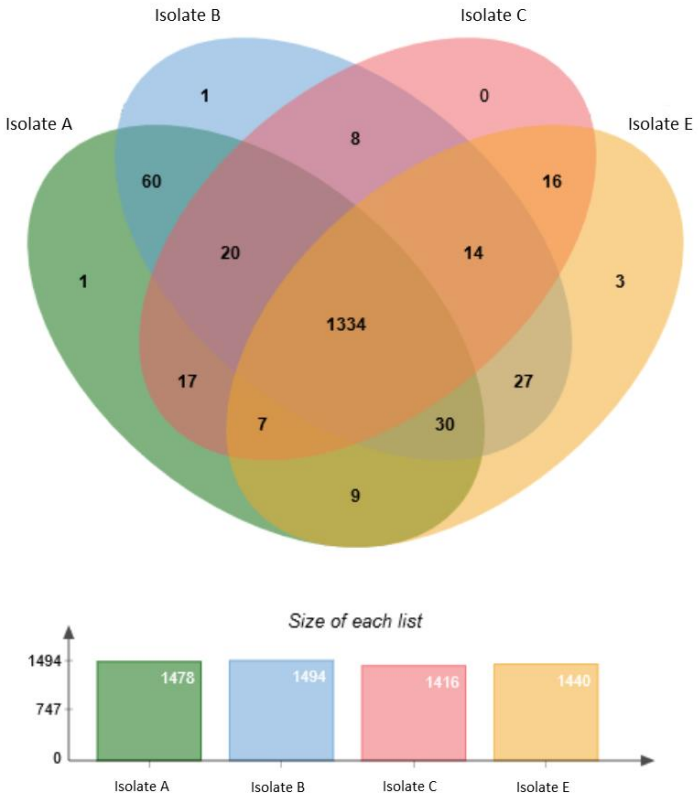

Figure 4.12: Venn diagram produced by OrthoVenn2 for comparison of OPCs from isolates A, B, C, and E.

Table 4.12: InterPro and BLASTp searches of OrthoVenn2 clusters of interest.

Cluster number	Prokka annotation		InterPro Scan		BLASTp		
	Locus tag	Annotation	InterPro number	Description	Description	E value	Accession
Clusters shared between isolates B, C, and E							
1351	DILABLLM_00061 KAEGCKPN_00714 CBGCIAIG_00152	Hypothetical protein	IPR001525	C-5 cytosine methyltransferase	DNA cytosine methyltransferase	1.00E-48	WP_120824255.1
1353	DILABLLM_01453 KAEGCKPN_01530 CBGCIAIG_01457	Hypothetical protein	IPR029063	S-adenosyl-L-methionine-dependent methyltransferase	Hypothetical protein	0	WP_154414142.1
1363	DILABLLM_00541 KAEGCKPN_01210 CBGCIAIG_00762	Hypothetical protein	IPR007979	Type II restriction enzyme NlaIII/ICEA1	HNH endonuclease	8.00E-102	WP_154413989.1
1365	DILABLLM_01058 KAEGCKPN_00872 CBGCIAIG_01487	Hypothetical protein	IPR010446	Beta-1,4-N-acetylgalactosaminyltransferase	Beta-1,4-N-acetylgalactosamyltransferase	0	WP_021304758.1
1367	DILABLLM_00825 KAEGCKPN_00257 CBGCIAIG_00305	Hypothetical protein	IPR022749	N6 adenine-specific DNA methyltransferase, N-terminal domain	Type I restriction-modification system, M subunit	2.00E-103	BAW63937.1
1374	DILABLLM_00308 KAEGCKPN_00205 CBGCIAIG_01532	Hypothetical protein	IPR007409	Restriction endonuclease, type I, HsdR, N-terminal	Type I restriction endonuclease subunit R	8.00E-148	WP_154247746.1
1376	DILABLLM_00823 KAEGCKPN_00260 CBGCIAIG_00309	Type I restriction enzyme EcoR124II R protein	IPR040980	SWI2/SNF2 ATPase	Type I restriction enzyme R protein	0	BAW73314.1

Clusters shared between isolates B and C							
1439	KAEGCKPN_00753 DILABLLM_00027	Type-2 restriction enzyme MboI	IPR021191	Restriction endonuclease, type II, DpmII	Restriction endonuclease	0	WP_053576689.1
1448	KAEGCKPN_01583 DILABLLM_01492	Hypothetical protein	IPR011701	Major facilitator superfamily	MFS transporter	0	WP_042629753.1
Clusters shared between isolate B and E							
1524	KAEGCKPN_00407 CBGCIAIG_01504	Hypothetical protein	IPR004346	CagE, TrbE, VirB component of type IV transporter system	VirB4 family type IV secretion/conjugal transfer ATPase	0	WP_033604619.1
1527	KAEGCKPN_00379 CBGCIAIG_01427	Hypothetical protein	IPR014942	Nucleotidyl transferase AbiEii toxin, Type IV TA system	Nucleotidyl transferase AbiEii/AbiGii toxin family protein	0	WP_140486252.1
1529	KAEGCKPN_00258 CBGCIAIG_00306	Hypothetical protein	IPR003356	DNA methylase, adenine-specific	Type I restriction-modification system, M subunit	1.00E-98	EKE94048.1
1532	KAEGCKPN_00392 CBGCIAIG_01440	Hypothetical protein	IPR014942	Nucleotidyl transferase AbiEii toxin, Type IV TA system	Nucleotidyl transferase AbiEii/AbiGii toxin family protein	1.00E-170	WP_128028548.1
1535	KAEGCKPN_01189 CBGCIAIG_00741	Aldo-keto reductase IolS	IPR023210	NADP-dependent oxidoreductase domain	Aldo/keto reductase	0	WP_001187413.1
1537	KAEGCKPN_00409 CBGCIAIG_01507	Hypothetical protein	IPR007039	Conjugal transfer TrbC/type IV secretion VirB2	TrbC/VirB2 family protein	4.00E-59	WP_000736478.1
1544	KAEGCKPN_00377 CBGCIAIG_01424	Hypothetical protein	IPR005094	Endonuclease relaxase, MobA/VirD2	Relaxase/mobilization nuclease domain-containing protein	0	WP_120903240.1

Clusters present in isolates C and E							
1415	DILABLLM_01457 CBGCIAIG_01583	IS200/IS605 family transposase IS605	IPR010095, IPR001959	Transposase IS605, OrfB, C-terminal. Probable transposase, IS891/IS1136/IS1341	IS200/IS605 family element transposase accessory protein TnpB	0	WP_154569861.1
1423	DILABLLM_00063 CBGCIAIG_00157	Hypothetical protein	IPR003615	HNH nuclease	HNH endonuclease	0	WP_021299744.1
1451	DILABLLM_00064 CBGCIAIG_00159	Modification methylase BspRI	IPR001525	C-5 cytosine methyltransferase	DNA (cytosine-5-)-methyltransferase	0	WP_154428969.1
Clusters only present in isolate B							
1412	KAEGCKPN_01245	Hypothetical protein	IPR002718	Outer membrane protein, Helicobacter	Outer membrane protein family protein	4.00E-68	EJB26909.1
Clusters only present in isolate E							
1409	CBGCIAIG_01113	Hypothetical protein	IPR002201	Glycosyl transferase, family 9	Glycosyltransferase 9 family protein	0	EST40520.1
1411	CBGCIAIG_01580	IS200/IS605 family transposase IS606	IPR001959, IPR010095, IPR021027	Probable transposase, IS891/IS1136/IS1341. Transposase IS605, OrfB, C-terminal. Transposase, putative, helix-turn-helix domain	IS200/IS605 family element transposase accessory protein TnpB	0	WP_108262961.1

KAEGCKPN locus tag = isolate B. DILABLLM locus tag = isolate C. CBGCIAIG locus tag = isolate E.

4.3.9 Presence/Absence of A2147G Mutation Associated with Antibiotic Resistance

From the results presented above many genes were identified that may be involved in clarithromycin and metronidazole resistance in isolates B, C, and E. Of these genes, only the A2147G mutation within the 23S rRNA gene has been well characterised in other research as being associated with clarithromycin resistance. To understand the evolution of this mutation the presence or absence of the mutation was overlaid onto the nucleotide PhyloPhlAn2 tree (Section 3.3.8.1). Within the PhyloPhlAn2 dataset (Section 3.2.4.1, Appendix A: Whole Genome Sequence Data Set), the A2147G mutation was identified in 17 out of 159 *H. pylori* isolates (Figure 4.13). The figure has had minor clades collapsed as they did not contain any isolates with the A2147G mutation; a full image may be viewed in Appendix J: Presence/Absence of A2147G Mutation over Nucleotide PhyloPhlAn2 tree.. One isolate, *H. pylori* strain Hpbs3 (NZ_ CP035106) only had the mutation in one copy of its 23S rRNA gene, the other copy maintained the original nucleotide. In all other isolates the mutation was present in all copies of the 23S rRNA gene. The 17 isolates identified with the A2147G mutation are spread through the nucleotide PhyloPhlAn2 tree.

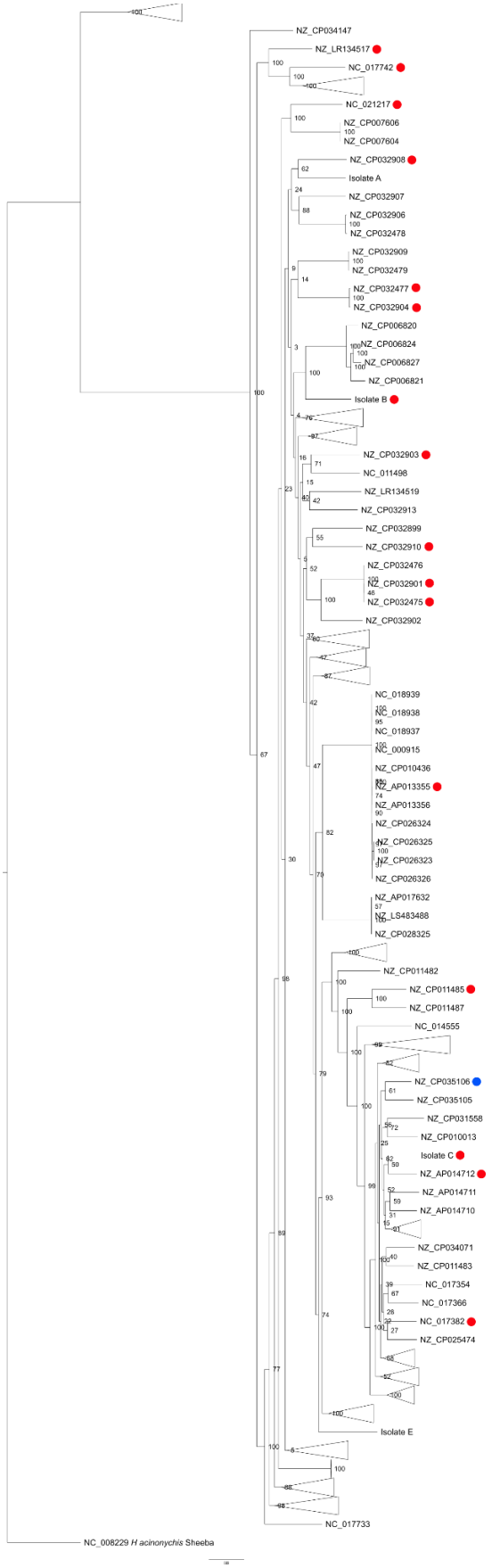

Figure 4.13: Nucleotide sequence PhyloPhlAn2 tree (Section 3.3.8.1) overlaid with presence of A2147G mutation. Red circles next to accession numbers indicate that the A2147G mutation is present in all copies of the 23S rRNA gene. Blue circle indicates that the A2147G mutation was only present in one copy of the 23S rRNA gene. Scale indicates number of substitutions per site.

4.4 Discussion

As antibiotic resistance becomes an increasingly complex and damaging issue to many aspects of modern life, understanding the underlying mechanisms of these phenotypes becomes readily important for the diagnosis and treatment of infections (Laxminarayan et al. 2016). The increasing resistance and subsequent treatment failure of *H. pylori* infections has highlighted the need for further research on how this bacterium develops resistance and what these mechanisms of resistance are. The aim of this chapter was to identify the genomic variation between antibiotic sensitive (isolate A) and antibiotic resistant (isolates B, C, and E) *H. pylori* isolated within NZ to identify genes or mutations that may be associated with resistant phenotypes.

Comparative genomics is an important tool to help understand biological features of many organisms (Xia 2013). The use of WGS in identifying and monitoring antibiotic resistance has become a major tool for clinicians and researchers to use to help with the control and prevention of antibiotic resistant bacteria (Inouye et al. 2014; Schürch and van Schaik 2017).

The PhyloPhlAn2 analysis described in Chapter 3 was used to inform further analysis to identify mechanisms of resistance towards clarithromycin and metronidazole.

4.4.1 Gene Prediction and Annotation

Utilising a variety of programmes to predict and annotate the draft genomes highlighted the different number and descriptions of genes and proteins identified using each programme. These analyses highlight the importance of using different annotation methods to identify potential proteins and to help elucidate the true function of a protein. As well as this, the inability of Nullarbor to identify any tRNAs indicates a potential limitation of the assembly produced by this analysis.

4.4.2 Genomic Features

Genomic characterisation of isolates A, B, C, and E was initially carried out to discern the genomic features of each isolate in this study, followed by comparisons between the resistant and sensitive isolates as well as between the closest relatives of each isolate previously identified in Chapter 3.

4.4.2.1 rRNA Analysis

An interesting feature to note of isolates A, B, C, and E are the number and placement of the rRNA genes. Each isolate only possessed one copy of each rRNA gene (5S, 16S, and 23S rRNA). Typically, *H. pylori* isolates possess two copies of all rRNA genes (Alm et al. 1999; Gerrits et al.

2003; Taylor et al. 1997). However, some isolates have been completely sequenced that only possess one copy of the rRNA genes, strain L7 (NZ_CP011482.1), strain K26A1 (NZ_CP011486.1) and strain DU15 (NZ_CP011483.1). Isolates A, B, C, and E may only possess one copy of each rRNA gene or multiple copies of these genes may have been missed in the assembly. Closing each genome to provide a more comprehensive representation of the true bacterial genomes will provide extra insight into the accurate number of rRNA genes within the genomes.

4.4.2.2 CRISPR Regions

Using CRISPR-Cas Finder (Couvin et al. 2018), some potential CRISPR regions were identified in isolates C and E. However, the small size and low evidence level of these regions indicated that they are not true CRISPR regions. A functional CRISPR-Cas region has not yet been identified in *H. pylori* genomes; however, CRISPR-like loci have been identified and are usually located around the vacA-like paralogue (*vlpC*) gene (Bangpanwimon et al. 2017; García-Zea et al. 2019). The regions identified in isolates C and E do not bear similarity towards the CRISPR-like loci identified in other isolates in sequence, size or location (García-Zea et al. 2019). It is unlikely that the partial CRISPR regions identified in isolates C and E are true CRISPR regions.

4.4.2.3 Phage Analysis

In order to locate any prophages within the genome, PHASTER software (Arndt et al. 2016) was used. An incomplete prophage region was identified in isolate B. Prophage regions are found in approximately 20 % of *H. pylori* strains and contribute to the large genomic diversity observed among this species (Lehours et al. 2011; Vale et al. 2017). The prophage region identified in isolate B is similar to other prophage regions identified within *H. pylori* genomes in terms of size and approximate G+C content (Falush et al. 2003; Fan et al. 2016). The transposase regions identified in this prophage region are also similar to insertion sequences previously identified in *H. pylori* isolates (Vale et al. 2017). Although these transposable elements have been associated with the horizontal spread of genes associated with increased pathogenicity and antibiotic resistance this does not seem to be the case here and thus further analysis on this region was not performed (Frost et al. 2005).

4.4.2.4 Insertion Sequences

Insertion sequences were identified in isolates A, B, C, and E through analysis using ISEScan (Xie and Tang 2017). Although both complete and partial IS were identified, all of the complete sequences belonged to the IS200/605 family. This family of IS were initially identified in *H. pylori* and it is therefore not unexpected to find them in the isolates from this study (Akopyants et al. 1998; Censini et al. 1996).

The position of IS adjacent to antibiotic resistance determinants has been linked with the development of resistance. Metronidazole resistance involving the *rdxA* gene has been linked with a variant of IS605 (Tankovic et al. 2000). Insertion of this sequence within the gene caused premature truncation of the nitroreductase making it unable to reduce the prodrug within the cell. A mini-IS*605* next to *rdxA* has also caused a deletion in the gene, adjacent to its site of insertion, once again leading to premature truncation and metronidazole resistance (Debets-Ossenkopp et al. 1999). Based on the draft and rearranged genomes of isolates A, B, C, and E, the IS, IS605, was not located near the nitroreductase, RdxA. However, this may be a product of the assembly and rearrangement of the contigs, especially where the ISs are overlapping or are near the end contigs, such as IS A1, E3, E4, E5 and E6. Completing the genomes will create a clearer picture of where the true positions of these ISs are within the isolates and the genes surrounding them.

4.4.3 Virulome, Metabolic Potential and COG analysis

The virulence factors possessed by isolates A, B, C, and E were determined using both Nullarbor and manually through the VFDB (Chen et al. 2016). The isolates possessed common virulence factors of *H. pylori* isolates required for colonization of the human stomach. However, none of the isolates possessed *dupA*, suggesting the isolates may be unable to form duodenal ulcers or there is data missing from this assembly. Both Nullarbor and manual analysis through the VFDB identified these common virulence factors in the isolates. However, utilising Nullarbor with standard parameters produced a different report than manual analysis through the VFDB. The discrepancies between these two methods and the different virulence factors identified may be of further interest; however, that analysis is beyond the scope of this study.

The metabolic potential of isolates A, B, C, and E was predicted using KEGG annotations and BLAST analysis. Utilising the draft genomes of each isolate shows that metabolic potential of all isolates is similar to other *H. pylori* isolates. Isolates A, B, C, and E are missing key genes that subsequently cause the standard KEGG glycolysis and CAC pathways to become branched. Although *H. pylori* possess the relevant genes to utilise glucose, this does not seem to be the preferred carbohydrate source for all *H. pylori* isolates (Lee et al. 2017). Instead, amino acids, in particular, asparagine, aspartate and alanine, are believed to be both the core nitrogen and carbon sources for *H. pylori* (Lee et al. 2017). Nitrogen metabolism is integral to *H. pylori*, not only for the synthesis of key bacterial constituents but also for its colonization and survival in the human stomach (Marshall et al. 1990; Price et al. 2002). Urease may be utilised initially by *H. pylori* to help locate the epithelium (Huang et al. 2015). This enzyme then aids in survival by providing a protective cloud of ammonia around the cell by neutralising the acidic environment

(Marshall et al. 1990). Glutamine synthetase and glutamate dehydrogenase also aid in acid tolerance through ammonium assimilation of the ammonium produced by urease (Miller and Maier 2014). However, a complete understanding of nitrogen metabolism, especially amino acid metabolism and the complex interactions of these reactions is not completely understood. This analysis is beyond the scope of this study but is required to fully understand the metabolism of *H. pylori*.

The challenging and complex nature of *H. pylori* metabolism makes it difficult to use standard genomic analysis for identifying the metabolic potential of the isolates. Genomic analysis alone possesses limitations in the absence of phenotypic testing such as genes that are not yet characterised and within a database may be missed leading to incorrect calls of missing genes. This was particularly apparent when using the KEGG database. Several genes such as *glup* and malate:quinone oxidoreductase are not fully incorporated into the standard KEGG pathways. Utilising the standard pathways within the KEGG database provides an incomplete overview of the metabolic potential of *H. pylori* isolates. As well as this, the presence of a gene within the isolates does not mean that this gene is active within the genome; biochemical testing will be required to test the full metabolic potential of the isolates. Further work is required to complete and compile the metabolic pathways of *H. pylori* for a comprehensive genomic analysis.

The COG analysis performed via eggNOG, showed little difference between the isolates from this study (A, B, C, and E) and their respective closest relatives identified by PhyloPhlAn2. The largest COG function grouping was function unknown (S). This suggests that the function of a number of genes within the *H. pylori* genome are still unknown and further work is required to understand and classify these genes.

4.4.4 Mauve Analysis

Utilising Mauve to rearrange and align the contigs of isolates A, B, C, and E against their closest relatives, according to PhyloPhlAn2, provided insight into genomic rearrangement events within *H. pylori*. The alignments show that the isolates share many genomic features with their respective closest relatives. The gaps in the isolates (A, B, C, and E) did not contain any known mechanisms of resistance. However, novel mechanisms may be present in these gaps and further analysis of these genes may be of further interest. Isolates A and E showed higher levels of genome rearrangement compared to isolates B and C. Genomic rearrangement can change the order and orientation of genes during evolution and involve inversions, duplications, transpositions or translocation (Darling et al. 2008; Lara-Ramírez et al. 2011; Noureen et al. 2019). These genomic rearrangements can have an effect on the phenotype of the bacterium (Darling et al. 2008). Genomic inversions are common in *H. pylori* and were observed in isolates

A and E (Lara-Ramírez et al. 2011). However, utilising a draft genome in this analysis creates a limitation as the exact gene order, arrangement of contigs and missing regions are not fully defined. Once again, completing the genome of isolates A, B, C, and E would provide a more comprehensive analysis of the genomic rearrangement events these isolates possess.

4.4.5 Antibiotic Resistance Genes

4.4.5.1 Manual Curation of Known Antibiotic Resistance Genes from *H. pylori*

To identify mechanisms of resistance associated with clarithromycin and metronidazole resistance in isolates B, C, and E two methods were employed. From Section 1.3.5 a number of resistance mechanisms were identified that are associated with resistant strains of *H. pylori*. Analysis of the genomes for isolates A, B, C, and E, identified point mutations with the 23S rRNA gene with a known association with clarithromycin resistance; and point mutations within RdxA, associated with metronidazole resistance (Chua et al. 2019; Mirzaei et al. 2014; Versalovic et al. 1996). CARD and OrthoVenn2 were used to identify other potential, novel mechanisms that have yet to be identified in *H. pylori*.

Clarithromycin

The target of clarithromycin is domain V of the 23S rRNA gene, where the drug reversible binds thereby inhibiting protein synthesis (Kanoh and Rubin 2010). Within this domain, three point mutations were identified amongst all isolates; A2147G, T2186C, and A2227G (Section 4.3.7.1). All three of these mutations have previously been identified in clarithromycin resistant strains of *H. pylori*; however, the position numbers used to be 2143, 2182 and 2223, respectively, which were changed due to re-sequencing of the 23S rRNA gene from *H. pylori* identifying an additional eight nucleotides within the gene (Gong et al. 2020; Mitui et al. 2014; Versalovic et al. 1996). Of these mutations, the A2147G mutation is the most common SNP associated with clarithromycin resistance in *H. pylori* with numerous studies detecting it (Binh et al. 2014; Miftahussurur et al. 2016; Tran et al. 2019; Versalovic et al. 1996). The two other point mutations identified in domain V, T2186C and A2227G, have been identified in clarithromycin resistant strains; however, their clinical relevance has not been well established as these mutations have also been identified in clarithromycin susceptible strains (Burucoa et al. 2005; Mitui et al. 2014; Tran et al. 2019).

Association of certain mutations with higher levels of resistance is still unclear. Binh et al. (2014) have identified the A2147G mutation in high-resistant strains and absent in low-resistant strains. Isolates B and E from this study both showed high resistance towards clarithromycin with an MIC > 256 µg/mL; of these two high-resistant strains, only isolate B possessed the A2147G

mutation. However, this mutation was also present in isolate C, which has an MIC of 24 µg/mL. This large difference between MICs, and absence of the key mutation in isolate E, suggests that the level of resistance cannot be determined from the presence or absence of specific 23S rRNA mutations alone. This has also been found by Lauener et al. (2019), who identified the A2147G mutation in resistant strains with a range of MICs from 1.5 to > 256 µg/mL. This suggests that high-level resistance may be multi-factorial, possibly due to the activation of other mechanisms such as efflux pumps.

Nucleotide position 2147 is located within the peptidyl transferase loop of domain V (Figure 4.14). Due to the high rate of adenine-to-guanine transversions within the peptidyl transferase loop, specifically in positions 2147 and 2146, which are associated with clarithromycin resistance, it has been suggested that the adenine-to-guanine transversion is beneficial for the bacterium (Debets-Ossenkopp et al. 1998; Wang and Taylor 1998). The adenine-to-guanine transversion does not affect the growth rate of the cells as well as creating a higher MIC, making it beneficial for the bacterium to produce and accumulate this mutation (Debets-Ossenkopp et al. 1998; Wang and Taylor 1998).

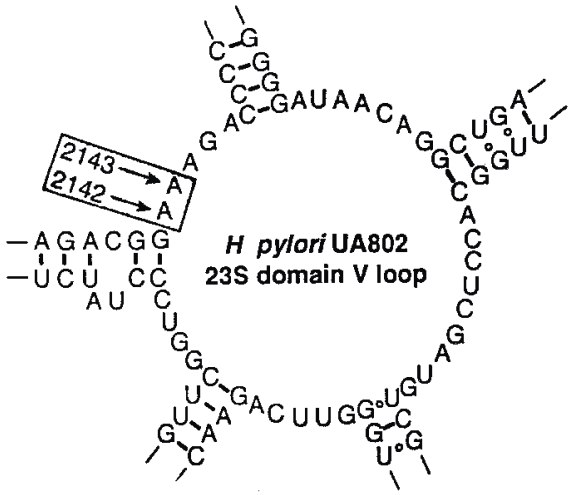

Figure 4.14: Secondary structure model of peptidyl transferase loop within domain V of the 23S rRNA gene of *H. pylori*. Positions of common point mutations conferring clarithromycin resistant strains are within the box. Position numbers have since been revised to 2146 and 2147. Image retrieved from (Taylor et al. 1997).

The exact mechanism of resistance towards clarithromycin caused by the 23S rRNA mutations in the peptidyl transferase loop are yet to be defined in *H. pylori*. However, they have been studied in other bacteria. Mutations within the peptidyl transferase loop of domain V, conferring resistance to macrolides, has been identified in *Mycobacterium avium*, *Mycoplasma pneumoniae* and *Streptococcus pneumoniae* (Lucier et al. 1995; Nash and Inderlied 1995; Tait-Kamradt et al. 2000). Thermodynamic studies of the secondary structure of *M. avium*'s 23S rRNA suggests that mutations within domain V alter the free energy integral to rRNA folding which

may cause a conformational change (Nash and Inderlied 1995). Resistance to erythromycin, the drug clarithromycin is semi-synthetically derived from, is associated with mutations at positions 2057 and 2058 within the peptidyl transferase loop of the 23S rRNA domain V (Douthwaite and Aagaard 1993; Guay et al. 2001). The mutations at positions 2057 and 2058 (corresponding to positions 2146 and 2147 in *H. pylori*) cause conformational changes within the peptidyl transferase loop; inducing a more open conformation thereby reducing the affinity of the drug (Douthwaite and Aagaard 1993). The conformational changes exhibited in other clarithromycin resistant bacteria reducing the binding affinity of macrolides may be similar to what is occurring in *H. pylori*. However, no studies have yet proved this conformational change is occurring in *H. pylori*. As well as this, many of the studies into the effects of domain V mutations on macrolide resistance were performed in the 1990s and early 2000s. With new and updated technologies, more information may be understood through further investigation of the peptidyl transferase loop of clarithromycin resistant *H. pylori* isolates.

The two other mutations present within domain V of isolates C and E, T2186C and A2227G, are not present within the peptidyl-transferase loop. The exact mechanism this mutation causes that elicits resistance to the drug is not currently known. This is likely due to the undefined clinical relevance of these mutations as they have been found in both resistant and sensitive strains of *H. pylori* (Burucoa et al. 2005; Kim et al. 2008a).

Clarithromycin resistance in isolates B and C is likely associated with the A2147G mutation as it has been well established in other resistant strains of *H. pylori*. For isolate E, there is little information about the association of the identified domain V mutations with clarithromycin resistance and further analysis was required to understand the genomic reason for its resistance profile. Between the isolates and their closest relatives (identified through the nucleotide PhyloPhlAn2 tree), the same mutations within the 23S rRNA gene do not appear. This suggests that the mutation is not conserved amongst certain strains of the bacteria and environmental factors may trigger its expression. Alternatively, the PhyloPhlAn2 tree may be inaccurate as this analysis only contained 155 *H. pylori* isolates with relationships being based on 400 genes. However, this is a comprehensive approach and the first phylogenetic analysis of its size on *H. pylori*, suggesting that it is the best representation of this species to date.

Metronidazole

Activation of the prodrug, metronidazole, is required for the drug to exert its bactericidal effect (Sisson et al. 2002). Nitroreductases, RdxA and FrxA are key components for the activation of metronidazole within *H. pylori* cells (Marais et al. 2003). Mutations within these two nitroreductases, especially RdxA, leading to inactivation of the reductase, are integral to the

currently understood mechanisms of metronidazole resistance in *H. pylori* (Marais et al. 2003; Tanih et al. 2011). Several amino acid changes were identified in RdxA of isolates A, B, C, and E; of particular interest are the R16H and A118S substitutions identified in isolates B and E. These substitutions have been identified in other metronidazole resistant *H. pylori* strains (Chua et al. 2019; Kwon et al. 2000c; Mirzaei et al. 2014; Saranathan et al. 2020; Yang et al. 2004).

The mechanism of resistance the R16H mutation causes is not yet well understood. However, it is believed that the mutation may decrease the binding efficiency of the nitroreductase to the drug. The flavin cofactor of RdxA has been established as flavin mononucleotide (FMN), which contributes to the nitroreductive capacity of the nitroreductase (Olekhnovich et al. 2009). Mutations present on the binding side chains of where FMN molecules attach to the RdxA dimer cause a reduction in the affinity of the apoprotein, which is suggested to inhibit the function of the enzyme (Martínez-Júlvez et al. 2012). From the residues associated with reduced affinity and stabilisation of FMN within the RdxA dimer, only the R16H mutation (Figure 4.15) was identified in isolates B and C (Martínez-Júlvez et al. 2012).

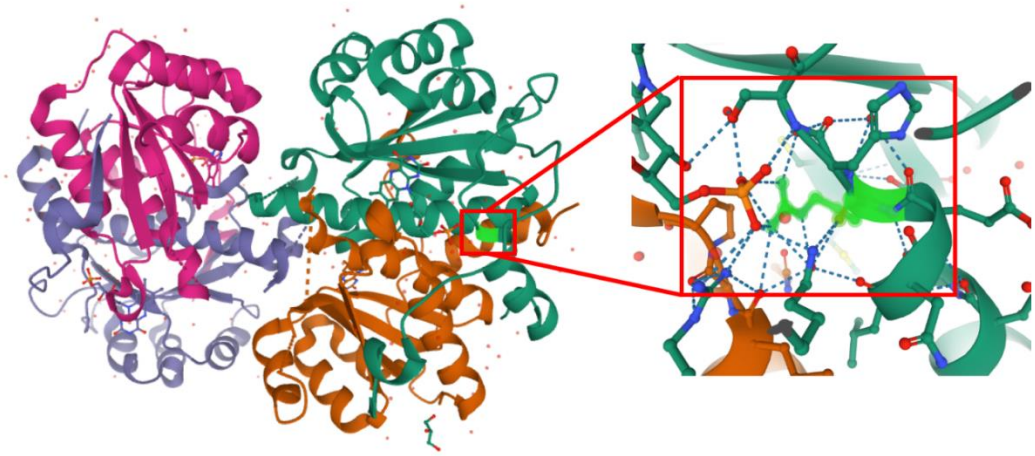

Figure 4.15: 3D Crystal structure of RdxA of *H. pylori*. The red box highlights the R16 residue. The dotted blue lines indicate the interactions of RdxA with FMN. Image adapted from https://www.rcsb.org/structure/3qdl.

The A118S mutation within RdxA of isolates B and E is not a well characterised mutation but has been identified in metronidazole resistant strains of *H. pylori* (Kwon et al. 2000c; Rasheed et al. 2014). It has been suggested that this mutation may cause structural differences in RdxA which may weaken the binding with metronidazole (Chu et al. 2020; Wang et al. 2018). However, these conclusions are based on predicted protein structures only. Further protein structure analysis is required to fully understand the effects this mutation has on the structure of RdxA and its interactions with the prodrug metronidazole in its potential contributions to metronidazole resistance.

The second amino acid substitution at position 118, A118T, was identified in isolate A, which is phenotypically sensitive to metronidazole. This suggests that a change from alanine to threonine in this position is not associated with metronidazole resistance, however, this mutation has been identified in resistant strains (Rasheed et al. 2014). The A118T amino acid substitution may not change the phenotype of the isolate, as it has been identified in both metronidazole resistant and sensitive strains; however, further work into the prevalence of these mutations in metronidazole resistant strains is required to determine the clinical significance of this mutation

Another known mechanism of metronidazole resistance in *H. pylori*, are mutations within the other nitroreductase, FrxA (Marais et al. 2003; Tanih et al. 2011). No common mutation was present in metronidazole resistant isolates (B and E). However, isolate B did possess a 32 amino acid deletion at the start of the protein. This deletion may cause inactivation of the protein; however, this cannot be determined based on *in silico* analysis alone. The other major mutations identified in FrxA in isolates C and 23-A-EK1, may cause inactivation or changes in the structure of the protein; however, they are not directly related to metronidazole resistance as both of these isolates are sensitive to metronidazole. The stop codon at position 207 (isolate C) and the 20 amino acid deletion at position 189 (23-A-EK1), have not been identified in metronidazole resistant *H. pylori* before.

Isolate E is phenotypically resistant to metronidazole at a high concentration. However, it does not possess any mutations within FrxA, and only possesses one amino acid substitution within RdxA, A118S, an under-characterised substitution with suggested involvement in metronidazole resistance. Further work is required to understand the involvement of this substitution in the inactivation of metronidazole and to determine if other mechanisms may be associated with high-level metronidazole resistance in *H. pylori*.

Efflux pumps

The association of clarithromycin and metronidazole resistance and efflux pumps is an area that still requires further research. The HefABC RND-type efflux system involved in multi-drug efflux was found in all isolates (A, B, C, and E). The *hp1184* and *hp1181* efflux pumps, belonging to the MATE and MFS superfamilies, respectively, were also found in all isolates. These efflux pumps may be a contributing factor to the resistance exhibited by isolate E.

As these efflux pumps are present in both sensitive and resistance isolates, it suggests that these pumps are involved in antibiotic resistance through either mutations or regulation of expression (Falsafi et al. 2016). Within HefA, the N177T mutation has been identified in some clarithromycin resistant strains (Iwamoto et al. 2014). However, this mutation was identified in both the sensitive and resistant isolates from this study (A, B, C, and E). This suggests that this mutation

may not be involved in clarithromycin resistance. However, in this study several other mutations were identified that may be involved in clarithromycin or metronidazole resistance as these mutations were absent in the sensitive isolate (A). These mutations have not been reported before in resistant *H. pylori* isolates. Due to the small sample size of this study and the limited metadata available for the closest relatives, further work is required to fully understand the relevance of these mutations to clarithromycin or metronidazole resistance.

Another limitation of this analysis, due to its *in-silico* nature, is that the expression of these genes is unknown. This analysis only provides evidence that these genes are within the isolate (A, B, C, and E) and closest relative genomes; however, it is not known if they are active. This is of particular importance for the MATE and MFS efflux pumps, *hp1184* and *hp1181*. Expression of these efflux pumps is suggested to be linked to a post-transcriptional regulation step which produces a resistance phenotype (Falsafi et al. 2016). Further *in vivo* work is required to determine if these efflux genes are active in isolates A, B, C, and E.

The contribution of efflux pumps to antibiotic resistance in *H. pylori* is still an area that requires further research. Understanding the effects of mutations on the RND-type efflux pumps and understanding the causes of regulation for the MATE and MFS-type efflux pumps is an important step elucidating how these mechanisms operate.

Through identification of genes and mutations commonly associated with clarithromycin and metronidazole resistant *H. pylori*, possible mechanisms of resistance for clarithromycin resistance in isolates B and C were identified as well as metronidazole resistance in isolate B. However, it is still unclear how, at a molecular level, isolate E had developed resistance towards both clarithromycin and metronidazole. Analysis using CARD was performed to identify other genes or mutations of interest that may be associated with macrolides and nitroimidazoles.

4.4.5.2 Novel Antibiotic Resistance Determinants

Using manual extraction and alignment of genes associated with clarithromycin and metronidazole resistance provided insight into how isolates B, C, and E likely confer resistance. However, for isolate E especially, no well-known mechanisms were identified. In an attempt to identify other potential mechanisms of resistance, including any novel determinants, CARD and OrthoVenn2 were used.

CARD

CARD is a manually curated database about the molecular basis of antimicrobial resistance (McArthur et al. 2013). Analysis using this database provided insight into known resistance mechanisms and other potential genes of interest in isolates A, B, C, and E.

This analysis was completed through the Nullarbor report as well as through manual analysis. Between these two approaches, Nullarbor analysis suggested there were no antibiotic resistance mechanisms within the isolates, whereas through manual analysis a number of mechanisms were identified. The discrepancy of results may be partly due to the reduced selection criteria cut-off value for manual analysis. However, the 23S rRNA mutation and HP1181 efflux pump, identified with ~98 % sequence similarity each, were not identified through the Nullarbor analysis. This raises questions about the accuracy of the report produced by Nullarbor.

Through manual analysis of CARD through the RGI, the data highlights that there is not one common mechanism of resistance towards either macrolides or nitroimidazoles in isolates A, B, C, and E. However, it did confirm the presence of the 23S rRNA mutation, A2147G, in isolates B and C; contributing to the clarithromycin resistance exhibited by these isolates. The HP1181 efflux pump was also identified in all isolates (A, B, C, and E) with a high amino acid sequence similarity (~98%). These results are in accordance with what was identified in Section 4.3.7.3. The HefABC and HP1184 efflux pumps were not identified through this analysis as they are not currently present within the CARD database.

In an attempt to identify novel genes of interest for antibiotic resistance in *H. pylori* the selection criteria within the RGI was lowered to include perfect, strict and loose hits; thereby decreasing the sequence similarity cut-off value and increasing the number of homologs identified. The variety in amino acid sequence similarity of the hits recorded by CARD's RGI, ranging from 23 % to 99 %, using the perfect, strict and loose hits selection criteria, required further manual curation to determine the reliability of the hit. This was important to elucidate the true function of the protein sequences identified for the ErmY and MexJ hits. No perfect hits, 100 % sequence similarity, were identified in this analysis highlighting the importance of coupling this genomic analysis with phenotypic analyses. Many genes of interest identified were associated with tetracycline resistance, of which none of the isolates are resistant to. The genes of the antibiotic classes the isolates (A, B, C, and E) were resistant to, macrolides and nitroimidazoles, were studied further due to the low similarity scores associated with these genes and the sequences within CARD.

Due to the low similarity scores associated with most antibiotic resistance determinants identified by CARD, further analysis was required to elucidate the true function of these sequences. Utilising sequence homology searches and identifying protein families and domains aided in this understanding. Through further analysis of the *ermY* gene identified by CARD, suggests that this gene likely encodes RsmA, a 16S rRNA adenine-dimethyltransferase and not a methyltransferase associated with the 23S rRNA mutation. This highlights the limitations of

decreasing the cut-off criteria of determining antibiotic resistance determinants based on open-reading frame prediction and sequence alignment and homology.

The HefABC efflux pump is not currently included within the CARD database, and therefore was not identified through RGI analysis of the isolates (A, B, C, and E). However, through this analysis only the HefB component was identified, and it was identified as two different hits, MexJ and cmeA. Both MexJ and cmeA are the membrane fusion protein component of RND-type efflux proteins identified in *P. aeruginosa* and *C. jejuni*, respectively (Chuanchuen et al. 2002; Lin et al. 2002). As HefB is the membrane fusion protein component of the HefABC efflux pump, this suggests that CARD selected the closest homolog that was present within the database due to the current absence of HefB in the database. Although HefB was identified as a membrane fusion protein homolog, it was still spread across two different CARD hits. The InterPro results showed that there were different domains associated with each CARD hit, membrane fusion protein, biotin-lipoyl like domain (IPR039562; isolates A and C) and RND efflux pump, membrane fusion protein, barrel-sandwich domain (IPR032317; isolates B and E). These two domains are both associated with membrane fusion proteins and identified in RND-type efflux pumps, in *E. coli* (Su et al. 2009; Van Dyk et al. 2004). The difference in domains is likely why CARD identified these sequences as different antibiotic resistance determinants. Further analysis is required to understand the structural differences these different domains might have on the functionality of these efflux pumps and potential association with antibiotic efflux.

The inability of the CARD analysis to distinguish different members of multi-gene families highlights the limitations of this software to identify novel resistance mechanisms, particularly related to efflux pumps. This is likely due to the limited number of entries currently within the database, especially related to *H. pylori*, of which there are currently only two used in the RGI (A2147G 23S rRNA mutation and HP1181 efflux pump).

Although CARD contains many genes associated with antibiotic resistance it is not a comprehensive compilation and only contains well-known and characterised resistance mechanisms (Alcock et al. 2020). This proves a limitation when trying to identify new resistance mechanisms or not well characterised mechanisms of resistance. Mutations other than A2147G within domain V of the 23S rRNA gene are not included within the database, as well as mutations present within nitroreductases. This is likely due to the incomplete characterisation and effect of these mutations on conferring resistance. This limitation highlights the need for further research into the specific effects of these mutations for resistance towards clarithromycin and metronidazole in *H. pylori* isolates.

Overall, CARD provides insight into the presence of known and well characterised antibiotic resistant determinants, such as the 23S rRNA mutation and HP1181 in isolates A, B, C, and E. However, it is not a comprehensive tool for identifying novel antibiotic resistance determinants in *H. pylori* as the data alone is not reliable and further manual curation is required to fully elucidate the function of the protein sequences identified.

OrthoVenn2

Utilising OrthoVenn2 provided a broader insight into protein clusters shared between resistant isolates (B, C, and E) in an attempt to identify novel resistance mechanisms only present in resistant isolates. This analysis suggests that proteins that are shared between or are only present in resistant isolates may be associated with antibiotic resistance. Of the protein clusters identified, clusters 1412 and 1448, belonging to the OMP and MFS families, respectively, were of interest due to their potential involvement in reduced membrane permeability or active efflux.

H. pylori possess a number of outer membrane proteins which make up ~4% of the bacterium's genome (Alm et al. 2000). These proteins are primarily involved in colonization and virulence of this human pathogen and are differentially expressed based on adaptation to the individual host (Odenbreit et al. 2009). However, some outer membrane proteins, HopT (BabB), HofC, and OMP31 are up-regulated in clarithromycin resistance strains (Smiley et al. 2013). The OMP identified in isolate E may be associated with increased membrane permeability as it is present within a resistant isolate from this study; however, this cannot be fully established based on *in silico* analysis alone. Further work is required with a larger dataset to establish whether this OMP is specific to resistant strains of *H. pylori* and if it is involved in reduced membrane permeability leading to antibiotic resistance.

The presence of an MFS efflux pump in isolates B and C suggests that this may be associated with clarithromycin resistance. However, this efflux pump was absent in isolate E, which is also resistant to clarithromycin. Currently within *H. pylori*, only one MFS efflux pump has been characterised, HP1181 (Falsafi et al. 2016). As the BLASTp results did not identify the MFS protein sequence as HP1181, and the absence of this protein in isolates A and E, suggests that this may be a novel MFS efflux protein that may be of further interest.

Comparison of CARD and OrthoVenn2 approaches for identifying novel antibiotic resistance determinants shows the diversity of the two methods. The OrthoVenn2 search provided a wider array of proteins shared between resistant isolates, whereas CARD primarily identified efflux proteins and methyltransferases. Based on BLASTp results of the proteins identified by each method, no two of the same proteins were identified. However, based on the InterPro results,

one family and two homologous superfamilies were identified in both analyses. The MFS (IPR011701), S-adenosyl-L-methionine-dependent methyltransferase (IPR029063) and P-loop containing nucleoside triphosphate hydrolase (IPR027417) categories were identified in both searches. No resistance determinants, including efflux pumps, were identified on the plasmids present in isolates A, C, or E using CARD or OrthoVenn2. This result is in accordance with (Lauener et al. 2019), who did not identify any resistance genes on plasmids. This suggests that antibiotic resistance within *H. pylori* is primarily mediated through SNPs in key genes, rather than plasmid encoded resistance.

Both of these methods provided data that is unreliable on its own and requires further analysis to discern the protein functions and potential involvement in antibiotic resistance. These two methods are also not equipped to identify mutations in specific proteins that may be associated with resistance, unless they are well characterised, like the 23S rRNA A2147G mutation. As further work is done to elucidate the diverse mechanisms contributing to antibiotic resistance in *H. pylori*, these databases will be updated and become a stronger tool for identifying the resistome of isolates.

4.4.6 Presence/Absence of A2147G Mutation Associated with Antibiotic Resistance in PhyloPhlAn2 Dataset

Based on the phylogenetic analysis completed in Chapter 3, it was suggested that resistance to clarithromycin and metronidazole were not deep-rooted traits based on the positions the isolates within the PyloPhlAn2 trees. As the A2147G mutation within the 23S rRNA gene is the only resistance mechanism identified in the isolates that is well defined in its association with clarithromycin resistance it was selected for further analysis to understand its evolution. The spread of the mutation through many different clades in the PhyloPhlAn2 tree also suggests that this mutation is not a deep-rooted trait or a conserved trait amongst certain clades or branches of *H. pylori*.

The spread of this mutation may be due to environmental factors that influence the curation of this mutation or it may be linked with a methyltransferase that regulates the emergence of this mutation; as noted in other bacteria expressing resistance towards macrolides (Matsuoka et al. 1998). The environment the host and bacterium are exposed to may contribute to the expression of resistance mechanisms. Previous use of metronidazole and macrolides has shown an association with resistant *H. pylori* strains; affecting the success of the treatment regimen employed to treat infection (Boltin et al. 2019; Kwon et al. 2019; McMahon et al. 2003). However, the mechanisms associated with these resistant strains have not been established. The association of the A2147G mutation with *H. pylori* strains retrieved from patients with

previous macrolide exposure has not been established but may help elucidate environmental conditions that may influence this mutation.

There may also be bias present within the phylogenetic analysis. The nucleotide PhyloPhlAn2 tree is based on 400 conserved genes that were optimised from 3,737 genomes (Segata et al. 2013). The use of only 400 gene sequences to determine the phylogeny of the isolates in this analysis may contribute to the positioning of the isolates and subsequent positioning of the A2147G mutation throughout the tree showing no direct correlation between genotype and phenotype. Using a single gene or multiple antibiotic resistance specific genes for this phylogenetic analysis may alter the positions and groupings of the isolates with the A2147G mutation, providing another view of the evolutionary history of the resistance mechanism. However, this will also introduce bias as the evolutionary relationships and groupings will only reflect those of the specific mechanism/s included in the analysis. Utilising PhyloPhlAn2, which uses 400 genes, creates a less biased and more reflective representation of the overall phylogeny of these isolates and provides a clearer prediction of the evolution of these antibiotic resistance determinants.

Further analysis is required to fully understand the causes of this mutation arising to help understand the evolution of this mutation and other potential targets for sequence-based antibiotic resistance profiling.

4.4.7 Conclusions

This chapter highlights the complexities of genomic characterisation and antibiotic resistance in *H. pylori* isolates. From genomic characterisation of the four draft genomes from isolates A, B, C, and E, no unique qualities were identified when compared to other *H. pylori* isolates. The genomic features, metabolic potential, virulome and COG distribution were all similar to other *H. pylori* isolates. For antibiotic resistance, of the four isolates included within this study, two had high-level resistance towards clarithromycin and metronidazole (isolates B and E) and one isolate had a moderate level of resistance to clarithromycin (isolate C). No one common mechanism of resistance was identified in association with either clarithromycin or metronidazole resistance.

Clarithromycin resistance of isolate B and C is likely associated with the A2147G mutation within domain V of the 23S rRNA gene. However, this mutation was absent in isolate E. The inability to confirm the resistance profiles determined by Middlemore Hospital (Section 2.3.2.1) suggests that the isolate used in this study may be clarithromycin sensitive, therefore suggesting why no common clarithromycin resistant mutations were identified in key genes. However, this isolate

may also suggest there is another mechanism conferring resistance towards clarithromycin, potentially an RND-type efflux pump identified through CARD analysis. Phenotypic confirmation of the resistance profile for isolate E is required. Metronidazole resistance is likely associated with mutations within the nitroreductase, RdxA, however, further studies are required to fully correlate the contributions of the mutations with the function of this protein. Resistance to this drug may also be associated with the HefABC, HP1181, or HP1184 efflux pumps; however, these pumps were present in all isolates and further studies are required to elucidate whether this pump is active only in resistant strains.

The two methods used to identify novel antibiotic resistance determinants, CARD and OrthoVenn2, helped identify some genes of interest that were common amongst the resistant isolates (B, C, and E). However, more extensive analysis is required to fully establish the function of the proteins and if they have any involvement in antibiotic resistance in these isolates.

The complex and diverse nature of the resistance determinants identified suggests that a single PCR test to identify major mutations only, such as the A2147G mutation, is not an effective test to determine resistance as other mechanisms are contributing to resistance, as noted in isolate E where this major mutation is absent.

Analysis of the isolates (A, B, C, and E) and their respective closest relatives (23-A-EK1, oki102, ML3 and J99, respectively), did not show that the same mutations were present, suggesting the mutations analysed may not be conserved. Furthering the analysis of the A2147G mutation, using the PhyloPhlAn2 dataset and nucleotide tree highlighted the spread of this mutation across the tree, also suggesting that this mutation is not conserved. This limited analysis suggests that environmental factors may contribute to the expression of this mutation.

Overall, this limited analysis on four *H. pylori* isolates from NZ, provides a broad look at common mutations conferring resistance, as well as potential novel mechanisms of resistance towards clarithromycin and metronidazole. However, this analysis does highlight the importance of both phenotypic and genotypic methods to evaluate antibiotic resistance. There are still many questions around establishing and evaluating all potential mechanisms of resistance towards clarithromycin and metronidazole in *H. pylori*.

Chapter 5 Final Discussion

H. pylori are an important human pathogen that are present in approximately 50 % of the world's population (Hooi et al. 2017). Infection can lead to a range of clinical outcomes including gastritis, gastric ulcers and cancer (Kusters et al. 2006). The importance of this pathogen extends beyond the severity of the clinical outcomes into the inability to effectively treat infections due to an increasing rate of antibiotic resistance (Savoldi et al. 2018). This increasing resistance impacts many factors including patient health and economic burden (Laxminarayan et al. 2016). It is therefore, of great clinical and economic importance to understand the mechanisms of antibiotic resistance and how to control it. This current study has contributed to the understanding of antibiotic resistant strains of *H. pylori* in NZ through a comparative genomics approach.

Understanding the rates of resistance and associated mechanisms are important for the effective treatment of this infection. The global spread of this pathogen has caused many subpopulations to arise (Thorell et al. 2017). This makes global recommendations impractical and a thorough understanding of local *H. pylori* populations is imperative in the effective management of both the infection itself and increasing antibiotic resistance rates. The lack of current publications of resistant *H. pylori* strains in NZ as well as an absence of comparative genomic studies of these isolates has created an important gap in knowledge to contribute to the management of this infection within NZ.

The objectives of this study were to establish resistance profiles of *H. pylori* isolates from NZ, investigate the phylogenetic relationships of these isolates within a larger dataset of *H. pylori* isolates and to characterise the genomes and provide insight into known or novel mechanisms of resistance.

5.1 Antibiotic Susceptibility Testing of *H. pylori*

Five isolates provided by Middlemore Hospital underwent antibiotic susceptibility testing to establish their resistance profiles. In-house testing was inconsistent with results provided by Middlemore Hospital, due to variations in antibiotic susceptibility testing methodologies. Resistance profiles provided by Middlemore Hospital using E-test strips showed that isolate A was resistant to clarithromycin, metronidazole, amoxicillin and tetracycline, isolates B and E were resistant to > 256 µg/mL of both clarithromycin and metronidazole, and isolates C and D were both resistant to clarithromycin at 24 µg/mL and 1 µg/mL, respectively. Unfortunately, isolate D did not survive the experimental procedures and was not studied further. Although this chapter was unsuccessful in replicating the results from Middlemore Hospital in-house, it

highlighted the lack of standardisation of antibiotic susceptibility testing of this bacterium. However, the results form Middlemore Hospital were considered reliable and isolates A, B, C, and E were studied further to understand how this resistance arose and what genomic mechanisms are leading to their resistance.

5.2 Phylogenetic Analysis

To understand the relation of these isolates (A-C, and E) with other *H. pylori* isolates and to begin to understand how antibiotic resistance may have arose a phylogenetic analysis was employed. Using a variety of genes to determine the phylogenetic relationships of the *H. pylori* isolates suggested that isolates A, B, C, and E are genetically distinct from each other, likely arise from different sources and through MLST analysis, are unique strains of *H. pylori*. Utilising one gene, 16S rRNA, failed to resolve the interspecies relationships of the dataset due to low bootstrap values, but provided sequence-based evidence, and confirmation, that the isolates (A-C and E) were *H. pylori*. Increasing the number of genes to eight by using the MLST data showed that each isolate (A-C and E) was unique within the *H. pylori* pubMLST database and showed the continental spread of the isolates, even though they were all isolated within NZ. Once again, the number of genes used in the analysis was increased, using PhyloPhlAn2 400 genes were included in this final phylogenetic analysis. This analysis increased the bootstrap values within the internal nodes of the trees compared to the 16S rRNA tree, however, the bootstrap values were still low, suggesting this modern method was still unable to resolve to interspecies relationships of these isolates. The lack of antibiotic resistance metadata associated with the dataset used means that inferences could not be directly made about how antibiotic resistance arose. However, the spread of resistant isolates B, C, and E, throughout all phylogenetic trees suggests that there are no clustered groupings of resistant isolates and antibiotic resistance has likely developed independently and through different methods. The closest neighbours for each isolate (A-C and E), identified through the nucleotide PhyloPhlAn2 tree were used to help characterise and compare the draft genomes of the isolates to provide insight into the potential mechanisms of resistance.

5.3 Genomic Analysis

Using a variety of methods, the draft genomes of isolates A, B, C, and E were characterised and compared against each other and their closest relatives, 23-A-EK1, oki102, ML3, and J182, respectively, identified from the nucleotide PhyloPhlAn2 analysis. Characterisation of the draft genomes from isolates A, B, C, and E, exhibited many similarities to other *H. pylori* isolates in genomic features, COG analysis, metabolic potential, and virulome analysis. Comparison of isolates A, B, C, and E, with their closest relatives (23-A-EK1, oki102, ML3 and J182, respectively),

showed that known mutations conferring resistance were not conserved and no known common mechanism was identified amongst resistant strains. The 23S rRNA mutation, A2147G, was identified in isolates B, C and 23-A-EK1, all phenotypically resistant to clarithromycin, however this mutation was absent in isolate E, also resistant to clarithromycin. Metronidazole resistance is likely linked with the R16H amino acid substitution in RdxA, from a missense mutation within the *rdxA* gene, for isolate B, but once again, this mutation is absent in isolate E. Efflux pumps were identified in all isolates, however, their expression levels are unable to be quantified using *in silico* analysis and further work is required to establish the role these mechanisms play in antibiotic resistance of these isolates (B, C, and E). The search for novel mechanisms highlighted the limitations of CARD for this type of analysis, due to the difficulty in its ability to distinguish between members of multigene families. OrthoVenn2 identified many protein clusters shared between resistant isolates (B, C, and E) only. Two of these are of particular interest, an MFS efflux protein and an OMP, due to the association of these types of proteins with antibiotic resistance (Co and Schiller 2006; Falsafi et al. 2016). The identification of efflux pumps in all isolates suggests that this mechanism may be the common mechanism of resistance, where inconsistencies were observed when looking for the known mutations. Further analysis of the proteins identified by OrthoVenn2 was not completed in this study and further research is required to fully elucidate the potential contribution of these proteins with antibiotic resistance.

5.3.1 Nullarbor

Through this thesis the Nullarbor report has been compared to traditional, manual methods for characterising and analysis the genomes from this study. Nullarbor produces a public health microbiology report consisting of assembly, annotation, MLST, resistome, virulome and SNP analysis (Seemann et al. n.d.). All of these components, expect for SNP analysis, were compared to manual methods in this thesis. Comparison of these methods showed discrepancies in the results, especially in the assembly and resistome analysis, however, MLST analysis correlated well with manual analysis. Using the standard settings and parameters included in Nullarbor produced a report that lacked important information on the resistance profiles of the isolates and the virulome output was different from manual analysis through the VFDB. Therefore, use of this analysis alone is not an accurate evaluation of the bacterium analysed. The inaccuracies of key components in clinically relevant bacteria emphasizes the need for further work to improve this all-encompassing analysis software.

5.4 Future directions

From the summary of results described above, many areas of further research were identified. These areas of future research are presented below.

- **Standardisation of antibiotic susceptibility testing for *H. pylori*.** The lack of standard methodology for assessing antibiotic resistance for this clinically relevant bacterium has led to various methods and breakpoints used for both clinical and research-based resistance determination. Creating standard methods for E-test strips, inoculum size, growth time, media use and breakpoints will help streamline processes and allow comparisons between different studies.
- **Larger sample size of NZ isolates with additional metadata.** A severe limitation of the present study was the small sample size used. Increasing the sample size of isolates and including important metadata of the patient will help create a greater understanding of the effects environmental factors have on the development of antibiotic resistance in *H. pylori*. It will also provide a clearer understanding of the level of antibiotic resistant *H. pylori* in NZ.
- **Genome completion for isolates A, B, C, and E using a third-generation sequencing technology.** Utilising either PacBio or Oxford Nanopore sequencing technology in conjunction with the Illumina sequences from this thesis will help complete the genome. This will provide a more accurate view of the genes and their order within the isolates.
- **Increase phylogenetic resolution.** The methods employed in this study were unable to fully resolve the evolution of the dataset of 159 *H. pylori* isolates. Creating a SNP tree may help resolve the evolutionary relationships of *H. pylori* isolates due to a greater amount of sequence information, compared to multiple gene analyses. Increasing the number of isolates in the phylogenetic analysis will also provide more sequence information and may help increase the bootstrap values throughout the middle of the phylogenetic trees.
- **Implications of migrant populations on local antibiotic resistance rates.** The rate of antibiotic resistance in migrant populations within NZ has not been established. Further work, through large scale analysis of migrant populations and antibiotic susceptibility testing, is needed to establish if this population is at a higher risk of developing antibiotic resistance and the effect of introducing resistance into the NZ population. The effect of these different strains of *H. pylori* on local *H. pylori* populations also requires further work through genomic analysis to determine if recombination is occurring and if other subpopulations are arising leading to geo-specific mutations.

- **Population structures of isolates A, B, C, and E to understand evolutionary histories and geographic implications leading to resistance.** Determining the population structures of the isolates from this study using STRUCTURE software will help determine the geographic origins of the isolates. This will help further understand the evolution of these strains and help determine if there is a connection between geographic location and specific antibiotic resistance mechanisms.
- **Establish to expression of efflux pumps in isolates A, B, C, and E.** To completely understand the role of efflux pumps in clarithromycin and metronidazole resistance in isolates B, C, and E the expression levels of the efflux pumps HefABC, HP1181, and HP1184 must be established. Using Real Time PCR (qPCR) with efflux pump specific primers will quantitively determine the levels of expression for each efflux pump. This will determine the contribution of these pumps to antibiotic resistance and determine if there is increased expression with higher levels of expression.
- **Structural analysis of mutations in key determinants involved in antibiotic resistance.** Predicted tertiary structures of key genes and proteins such as the 23S rRNA gene, RdxA, FrxA, and efflux pumps will help elucidate the conformational effects key mutations and amino acid substitutions have on the structures of these key resistance determinants.
- **Determine the association of MFS and OMP proteins identified in OrthoVenn2 analysis to antibiotic resistance.** From the OrthoVenn2 analysis (Section 4.3.8.2), two protein clusters, only identified in resistant isolates, were identified for further interest, an MFS efflux protein and an OMP. Characterisation of these proteins are required to fully elucidate their function and role in *H. pylori* and potential involvement in antibiotic resistance. This may be completed with further *in silico* and completing transformation or knock-out experiments for each protein of interest.

5.5 Conclusions

In conclusion, this study is the first comparative genomics study of antibiotic resistant *H. pylori* isolates only from NZ. This thesis highlights the complex and defiant nature of *H. pylori* in all aspects; infection, laboratory growth, DNA extraction and genomic variability. Antibiotic susceptibility testing showed that isolates from this study were resistant to clarithromycin and metronidazole. The use of comparative genomics identified the A2147G mutation, associated with clarithromycin resistance, in isolates B and C, and a missense mutation in the *rdx*A gene leading to an R16H amino acid substitution in the protein, likely associated with metronidazole resistance, in isolate B. This approach did not identify any common mechanisms amongst resistant isolates due to a lack of known mechanisms identified in isolate E. This may be due to the inability to verify the clarithromycin resistance identified by Middlemore Hospital,

suggesting that this isolate may be clarithromycin sensitive. Phylogenetic analysis using the 16S rRNA gene, MLST data and PhyloPhlAn2 highlighted the spread and genomic diversity of the isolates (A, B, C, and E). To date, this is the first phylogenetic analysis of *H. pylori* using PhyloPhlAn2 and 159 isolates. Resistant strains (B, C, and E) were not contained within one clade. Further analysis of the presence of the A2147G mutation for clarithromycin resistance in the PhyloPhlAn2 dataset also revealed a widespread occurrence of this mutation throughout the nucleotide tree. These findings suggest that the current known and well-characterised mechanisms of resistance for clarithromycin may not be the only determinants involved in resistance. Identification of several efflux pumps in all isolates suggests that this may be a common mechanism of resistance shared amongst all isolates. However, the contribution of the efflux pumps identified in this study, towards antibiotic resistance could not be determined based on *in silico* analysis alone. Utilising OrthoVenn2 highlighted the presence of efflux pump proteins only present in resistant isolates providing further evidence that this is an important area of future research. Finally, the search for novel resistance mechanisms identified an MFS efflux protein and OMP, only identified in resistant isolates which require further classification and elucidation of their role in antibiotic resistance. The wide array of resistance mechanisms and lack of defined determinants in some resistant strains suggests that a single PCR based method on currently known determinants alone, such as the 23S rRNA mutation, will not provide an accurate representation of antibiotic resistance in *H. pylori* isolates. For clinicians, this finding suggests that phenotypic antibiotic susceptibility testing is still the most reliable method for determining resistance to first line treatment antibiotics as well as levels of resistance.

Overall, this pilot study of antibiotic resistant strains of *H. pylori* from NZ has provided insight into the genomic characteristics of these strains and identified many areas of further research, outlined above (Section 5.4), to fully understand the complexities of this challenging bacterium.

References

Abdalsadeg N, Adam AA, Abdul-Aziz H, Omer WH, Osman HA, Bolad AK. 2012. Comparison of different diagnostic methods of *Helicobacter pylori* infection in Sudanese patients. Al Neelain Med J. 2(4):27-34.

Abdollahi H, Savari M, Zahedi MJ, Darvish MS, Hayatbakhsh AM. 2011. A study of *rdxA* gene deletion in metronidazole resistant and sensitive *Helicobacter pylori* isolates in Kerman, Iran. Jundishapur J Microbiol. 4(2):99–104.

Achtman M, Azuma T, Berg DE, Ito Y, Morelli G, Pan Z-J, Suerbaum S, Thompson SA, Van Der Ende A, Van Doorn L-J. 1999. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. Mol Microbiol. 32(3):459–470.

Adams BL, Bates TC, Oliver JD. 2003. Survival of *Helicobacter pylori* in a natural freshwater environment. Appl Environ Microbiol. 69(12):7462–7466.

Adams MD, Goglin K, Molyneaux N, Hujer KM, Lavender H, Jamison JJ, MacDonald IJ, Martin KM, Russo T, Campagnari AA et al. 2008. Comparative genome sequence analysis of multidrug-resistant *Acinetobacter baumannii*. J Bacteriol. 190(24):8053–8064.

Adelskov J, Patel BKC. 2015. Draft genome sequence of *Paenibacillus* strain P1XP2, a polysaccharide-degrading, thermophilic, facultative anaerobic bacterium isolated from a commercial bioreactor degrading food waste. Genome Announc. 3(1):e01484–01414.

Agudo S, Pérez-Pérez G, Alarcón T, López-Brea M. 2010. High prevalence of clarithromycin-resistant *Helicobacter pylori* strains and risk factors associated with resistance in Madrid, Spain. J Clin Microbiol. 48(10):3703–3707.

Åhman J, Matuschek E, Kahlmeter G. 2020. EUCAST evaluation of 21 brands of Mueller-Hinton dehydrated media for disc diffusion testing. Clin Microbiol Infect. doi:10.1016/j.cmi.2020.01.018.

Ahmed D, Brooks H, McConnell M, Barbezat G. 2004. Antibiotic resistance in *Helicobacter pylori*: is it a problem in New Zealand? N Z Med J. 117(1200):1022.

Ailloud F, Didelot X, Woltemate S, Pfaffinger G, Overmann J, Bader RC, Schulz C, Malfertheiner P, Suerbaum S. 2019. Within-host evolution of *Helicobacter pylori* shaped by niche-specific adaptation, intragastric migrations and selective sweeps. Nat Comm. 10(1):2273–2273.

Akhtereeva AR, Morozova LG, Faizullina RA, Ivanovskaya KA, Pozdeev OK, Valeeva IK, Abdulkhakov SR. 2018. Antibiotic susceptibility assessment of *Helicobacter pylori* isolates by disk-diffusion method. BioNanoScience. 8(3):930-934.

Akopyants NS, Clifton SW, Kersulyte D, Crabtree JE, Youree BE, Reece CA, Bukanov NO, Drazek ES, Roe BA, Berg DE. 1998. Analyses of the cag pathogenicity island of *Helicobacter pylori*. Mol Microbiol. 28(1):37-53.

Akopyanz N, Bukanov NO, Westblom TU, Kresovich S, Berg DE. 1992. DNA diversity among clinical isolates of *Helicobacter pylori* detected by PCR-based RAPD fingerprinting. Nucleic Acids Res. 20(19):5137–5142.

Alarcón-Millán J, Fernández-Tilapa G, Cortés-Malagón EM, Castañón-Sánchez CA, De Sampedro-Reyes J, Cruz-del Carmen I, Betancourt-Linares R, Román-Román A. 2016. Clarithromycin resistance and prevalence of *Helicobacter pylori* virulent genotypes in patients from Southern México with chronic gastritis. Infect Genet Evol. 44:190-198.

Alarcón T, Urruzuno P, Martínez MJ, Domingo D, Llorca L, Correa A, López-Brea M. 2017. Antimicrobial susceptibility of 6 antimicrobial agents in *Helicobacter pylori* clinical isolates by using EUCAST breakpoints compared with previously used breakpoints. Enferm Infecc Microbiol Clin. 35(5):278–282.

Alba C, Blanco A, Alarcón T. 2017. Antibiotic resistance in *Helicobacter pylori*. Curr Opin Infect Dis. 30(5).

Albert TJ, Dailidiene D, Dailide G, Norton JE, Kalia A, Richmond TA, Molla M, Singh J, Green RD, Berg DE. 2005. Mutation discovery in bacterial genomes: metronidazole resistance in *Helicobacter pylori*. Nat Methods. 2(12):951-953.

Albertson N, Wenngren I, Sjöström JE. 1998. Growth and survival of *Helicobacter pylori* in defined medium and susceptibility to Brij 78. J Clin Microbiol. 36(5):1232–1235.

Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-LV, Cheng AA, Liu S et al. 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 48(D1):D517-D525.

Ali A, Naz A, Soares SC, Bakhtiar M, Tiwari S, Hassan SS, Hanan F, Ramos R, Pereira U, Barh D et al. 2015. Pan-genome analysis of human gastric pathogen *H. pylori*: comparative genomics and pathogenomics approaches to identify regions associated with pathogenicity and prediction of potential core therapeutic targets. Biomed Res Int. 2015:17.

Alkim H, Koksal AR, Boga S, Sen I, Alkim C. 2017. Role of bismuth in the eradication of *Helicobacter pylori*. Am J Therapeut. 24(6):e751–e757.

AllSeq. Pacific Biosciences. 2020. [accessed 2020 3 March]. https://allseq.com/knowledge-bank/sequencing-platforms/pacific-biosciences/.

Alm RA, Bina J, Andrews BM, Doig P, Hancock RE, Trust TJ. 2000. Comparative genomics of *Helicobacter pylori*: analysis of the outer membrane protein families. Infect Immun. 68(7):4155-4168.

Alm RA, Ling L-SL, Moir DT, King BL, Brown ED, Doig PC, Smith DR, Noonan B, Guild BC, deJonge BL et al. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. Nature. 397(6715):176–180.

Alvarez-Ortega C, Olivares J, Martinez J. 2013. RND multidrug efflux pumps: what are they good for? Front Microbiol. 4(7).

Anderson VR, Perry CM. 2008. Levofloxacin. Drugs. 68(4):535-565.

Ansorg R, Von Recklinghausen G, Pomarius R, Schmid EN. 1991. Evaluation of techniques for isolation, subcultivation, and preservation of *Helicobacter pylori*. J Clin Microbiol. 29(1):51-53.

Arama SS, Tiliscan C, Negoita C, Croitoru A, Arama V, Mihai CM, Pop F, Garg A. 2016. Efficacy of 7-day and 14-day triple therapy regimens for the eradication of *Helicobacter pylori*: a comparative study in a cohort of Romanian patients. Gastroenterol Res Pract. 2016:7.

Aris-Brosou S. 2005. Determinants of adaptive evolution at the molecular level: the extended complexity hypothesis. Mol Biol Evol. 22(2):200-209.

Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44(W1):W16–W21.

Arslan N, Yılmaz Ö, Demiray-Gürbüz E. 2017. Importance of antimicrobial susceptibility testing for the management of eradication in *Helicobacter pylori* infection. World J Gastroenterol. 23(16):2854–2869.

Aspholm M, Olfat FO, Nordén J, Sondén B, Lundberg C, Sjöström R, Altraja S, Odenbreit S, Haas R, Wadström T et al. 2006. SabA is the *H. pylori* hemagglutinin and is polymorphic in binding to sialylated glycans. PLoS Pathog. 2(10):110.

Asqah MA, Hamoudi NA, Anil S, Jebreen AA, Al-Hamoudi WK. 2009. Is the presence of *Helicobacter pylori* in dental plaque of patients with chronic periodontitis a risk factor for gastric infection? Can J Gastroenterol. 23(3):177-179.

Atapoor S, Safarpoor Dehkordi F, Rahimi E. 2014. Detection of *Helicobacter pylori* in various types of vegetables and salads. Jundishapur J Microbiol. 7(5):e10013-e10013.

Avasthi TS, Devi SH, Taylor TD, Kumar N, Baddam R, Kondo S, Suzuki Y, Lamouliatte H, Mégraud F, Ahmed N. 2011. Genomes of two chronological isolates (*Helicobacter pylori* 2017 and 2018) of the West African *Helicobacter pylori* strain 908 obtained from a single patient. J Bacteriol. 193(13):3385–3386.

Axon AT. 1994. The role of acid inhibition in the treatment of *Helicobacter pylori* infection. Scand J Gastroenterol. 201:16-23.

Azevedo NF, Almeida C, Cerqueira L, Dias S, Keevil CW, Vieira MJ. 2007. Coccoid form of *Helicobacter pylori* as a morphological manifestation of cell adaptation to the environment. Appl Environ Microbiol. 73(10):3423–3427.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M et al. 2008. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genom. 9(1):75.

Baker KH, Hegarty JP. 2001. Presence of *Helicobacter pylori* in drinking water is associated with clinical infection. Scand J Infect Dis. 33(10):744-746.

Balouiri M, Sadiki M, Ibnsouda SK. 2016. Methods for *in vitro* evaluating antimicrobial activity: a review. J Pharm Anal. 6(2):71–79.

Baltrus DA, Amieva MR, Covacci A, Lowe TM, Merrell DS, Ottemann KM, Stein M, Salama NR, Guillemin K. 2009. The complete genome sequence of *Helicobacter pylori* strain G27. J Bacteriol. 191(1):447–448.

Bamford KB, Bickley J, Collins JS, Johnston BT, Potts S, Boston V, Owen RJ, Sloan JM. 1993. *Helicobacter pylori*: comparison of DNA fingerprints provides evidence for intrafamilial infection. Gut. 34(10):1348–1350.

Banatvala N, Davies GR, Abdi Y, Clements L, Rampton DS, Hardie JM, Feldman RA. 1994. High prevalence of *Helicobacter pylori* metronidazole resistance in migrants to east London: relation with previous nitroimidazole exposure and gastroduodenal disease. Gut. 35(11):1562-1566.

Bangpanwimon K, Sottisuporn J, Mittraparp-Arthorn P, Ueaphatthanaphanich W, Rattanasupar A, Pourcel C, Vuddhakul V. 2017. CRISPR-like sequences in *Helicobacter pylori* and application in genotyping. Gut Pathog. 9:65–65.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477.

Baxevanis AD. 2001. Bioinformatics and the internet. Bioinformatics: a practical guide to the analysis of genes and proteins. John Wiley & Sons. p. 1–17.

Begue RE, Gonzales JL, Correa-Gracian H, Tang SC. 1998. Dietary risk factors associated with the transmission of *Helicobacter pylori* in Lima, Peru. Am J Trop Med Hyg. 59(4):637-640.

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 456(7218):53–59.

Bergé M, Mortier-Barrière I, Martin B, Claverys J-P. 2003. Transformation of *Streptococcus pneumoniae* relies on DprA- and RecA-dependent protection of incoming DNA single strands. Mol Microbiol. 50(2):527-536.

Berger SA, Gorea A, Moskowitz M, Santo M, Gilat T. 1993. Effect of inoculum size on antimicrobial susceptibility of *Helicobacter pylori*. Eur J Clin Microbiol Infect Dis. 12(10):782–783.

Berroteran A, Perrone M, Correnti M, Cavazza ME, Tombazzi C, Goncalvez R, Lecuna V. 2002. Detection of *Helicobacter pylori* DNA in the oral cavity and gastroduodenal system of a Venezuelan population. J Med Microbiol. 51(9):764-770.

Bertelli C, Greub G. 2013. Rapid bacterial genome sequencing: methods and applications in clinical microbiology. Clin Microbiol Infect. 19(9):803-813.

Bie M, Wen J, Wang H, Zhou L, Yang L, Pan Y, Luo J, Liu C, Li M, Wang B. 2019. Phylogenetic analysis of clinical strains of *Helicobacter pylori* isolated from patients with gastric diseases in Tibet. Ann Transl Med. 7(14):27.

Bina JE, Alm RA, Uria-Nickelsen M, Thomas SR, Trust TJ, Hancock REW. 2000. *Helicobacter pylori* uptake and efflux: basis for intrinsic susceptibility to antibiotics in vitro. Antimicrob Agents Chemother. 44(2):248–254.

Binh TT, Shiota S, Suzuki R, Matsuda M, Trang TTH, Kwon DH, Iwatani S, Yamaoka Y. 2014. Discovery of novel mutations for clarithromycin resistance in *Helicobacter pylori* by using next-generation sequencing. J Antimicrob Chemother. 69(7):1796-1803.

Binh TT, Suzuki R, Trang TTH, Kwon DH, Yamaoka Y. 2015. Search for novel candidate mutations for metronidazole resistance in *Helicobacter pylori* using next-generation sequencing. Antimicrob Agents Chemother. 59(4):2343–2348.

Binladen J, Thomas M, Gilbert P, Campos PF, Willerslev E. 2007. 5'-tailed sequencing primers improve sequencing quality of PCR products. Biotechniques. 42(2):174–176.

Bio IT World. Six years after acquisition, Roche quietly shutters 454. 2013. Bio IT World; [accessed 2020 29 February]. http://www.bio-itworld.com/2013/10/16/six-years-after-acquisition-roche-quietly-shutters-454.html.

Björkholm B, Sjölund M, Falk PG, Berg OG, Engstrand L, Andersson DI. 2001. Mutation frequency and biological cost of antibiotic resistance in *Helicobacter pylori*. Proc Natl Acad Sci USA. 98(25):14607–14612.

Blair JMA, Webber MA, Baylay AJ, Ogbolu DO, Piddock LJV. 2014. Molecular mechanisms of antibiotic resistance. Nat Rev Microbiol. 13:42.

Blanchard TG, Nedrud JG. 2012. Laboratory maintenance of *Helicobacter* species. Curr Protoc Microbiol. 24(1):8B.1.1–8B.1.19.

Bland MV, Ismail S, Heinemann JA, Keenan JI. 2004. The action of bismuth against *Helicobacter pylori* mimics but is not caused by intracellular iron deprivation. Antimicrob Agents Chemother. 48(6):1983–1988.

Bode G, Mauch F, Malfertheiner P. 1993. The coccoid forms of *Helicobacter pylori.* Criteria for their viability. Epidemiol Infect. 111(3):483–490.

Boehnke KF, Eaton KA, Fontaine C, Brewster R, Wu J, Eisenberg JNS, Valdivieso M, Baker LH, Xi C. 2017. Reduced infectivity of waterborne viable but nonculturable *Helicobacter pylori* strain SS1 in mice. Helicobacter. 22(4):e12391.

Boigegrain R-A, Salhi I, Alvarez-Martinez M-T, Machold J, Fedon Y, Arpagaus M, Weise C, Rittig M, Rouot B. 2004. Release of periplasmic proteins of *Brucella suis* upon acidic shock involves the outer membrane protein Omp25. Infect Immun. 72(10):5693-5703.

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG. 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods. 10(1):57-59.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30(15):2114-2120.

Boltin D, Levi Z, Gingold-Belfer R, Gabay H, Shochat T, Niv Y, Dickman R, Dotan I, Birkenfeld S. 2019. Impact of previous exposure to macrolide antibiotics on *Helicobacter pylori* infection treatment outcomes. Official journal of the American College of Gastroenterology | ACG. 114(6):900-906.

Boren T, Falk P, Roth K, Larson G, Normark S. 1993. Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens. Science. 262(5141):1892–1895.

Bosshard PP, Abels S, Zbinden R, Böttger EC, Altwegg M. 2003. Ribosomal DNA sequencing for identification of aerobic Gram-positive rods in the clinical laboratory (an 18-month evaluation). J Clin Microbiol. 41(9):4134-4140.

Bougnoux M-E, Morand S, d'Enfert C. 2002. Usefulness of multilocus sequence typing for characterization of clinical isolates of *Candida albicans*. J Clin Microbiol. 40(4):1290-1297.

Braga LLBC, Batista MHR, de Azevedo OGR, da Silva Costa KC, Gomes AD, Rocha GA, Queiroz DMM. 2019. *oipA* "on" status of *Helicobacter pylori* is associated with gastric cancer in North-Eastern Brazil. BMC Canc. 19(1):48.

Brandt S, Kwok T, Hartig R, König W, Backert S. 2005. NF-kappaB activation and potentiation of proinflammatory responses by the *Helicobacter pylori* CagA protein. PNAS. 102(26):9300–9305.

Brenciaglia MI, Fornara AM, Scaltrito MM, Dubini F. 2000. *Helicobacter pylori*: cultivability and antibiotic susceptibility of coccoid forms. Int J Antimicrob Agents. 13(4):237–241.

Breurec S, Guillard B, Hem S, Brisse S, Dieye FB, Huerre M, Oung C, Raymond J, Tan TS, Thiberge J-M et al. 2011. Evolutionary history of *Helicobacter pylori* sequences reflect past human migrations in Southeast Asia. PLoS One. 6(7):e22058-e22058.

Brisson-Noël A, Trieu-Cuot P, Courvalin P. 1988. Mechanism of action of spiramycin and other macrolides. J Antimicrob Chemother. 22(Supplement_B):13-23.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 12(1):59-60.

Buck GE, Smith JS. 1987. Medium supplementation for growth of *Campylobacter pyloridis*. J Clin Microbiol. 25(4):597-599.

Buermans HPJ, den Dunnen JT. 2014. Next generation sequencing technology: advances and applications. BBA - Mol Basis Dis. 1842(10):1932–1941.

Bugaytsova JA, Björnham O, Chernov YA, Gideonsson P, Henriksson S, Mendez M, Sjöström R, Mahdavi J, Shevtsova A, Ilver D et al. 2017. *Helicobacter pylori* adapts to chronic infection and gastric disease via pH-responsive BabA-mediated adherence. Cell Host Microbe. 21(3):376-389.

Bulitta JB, Yang JC, Yohonn L, Ly NS, Brown SV, D'Hondt RE, Jusko WJ, Forrest A, Tsuji BT. 2010. Attenuation of colistin bactericidal activity by high inoculum of *Pseudomonas aeruginosa* characterized by a new mechanism-based population pharmacodynamic model. Antimicrob Agents Chemother. 54(5):2051-2062.

Burucoa C, Landron C, Garnier M, Fauchère J-L. 2005. T2182C mutation is not associated with clarithromycin resistance in *Helicobacter pylori*. Antimicrob Agents Chemother. 49(2):868-870.

Busch A, Thomas P, Myrtennäs K, Forsman M, Braune S, Runge M, Tomaso H. 2017. High-quality draft genome sequence of *Francisella tularensis* subsp. *holarctica* strain 08T0073 isolated from a Wild European hare. Genome Announc. 5(12):e01577–01516.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10(1):421.

Camargo MC, Piazuelo MB, Mera RM, Fontham ETH, Delgado AG, Yepez MC, Ceron C, Bravo LE, Bravo JC, Correa P. 2007. Effect of smoking on failure of *H. pylori* therapy and gastric histology in a high gastric cancer risk area of Colombia. Acta Gastroenterol Latinoam. 37(4):238–245.

Canard B, Sarfati RS. 1994. DNA polymerase fluorescent substrates with reversible 3'-tags. Gene. 148(1):1–6.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M et al. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 6(8):1621–1624.

Carson CF, Hammer KA, Riley TV. 1995. Broth micro-dilution method for determining the susceptibility of *Escherichia coli* and *Staphylococcus aureus* to the essential oil of *Melaleuca alternifolia* (tea tree oil). Microbios. 82(332):181-185.

Cederbrant G, Kahlmeter G, Ljungh Â. 1993. The E test for antimicrobial susceptibility testing of *Helicobacter pylori*. J Antimicrob Chemother. 31(1):65–71.

Cederbrant G, Kahlmeter G, Ljungh Å. 1992. Proposed mechanism for metronidazole resistance in *Helicobacter pylori*. J Antimicrob Chemother. 29(2):115-120.

Celli JP, Turner BS, Afdhal NH, Keates S, Ghiran I, Kelly CP, Ewoldt RH, McKinley GH, So P, Erramilli S et al. 2009. *Helicobacter pylori* moves through mucus by reducing mucin viscoelasticity. Proc Natl Acad Sci USA. 106(34):14321–14326.

Cellini L, Allocati N, Angelucci D, Iezzi T, Campli ED, Marzio L, Dainelli B. 1994. Coccoid *Helicobacter pylori* not culturable in vitro reverts in mice. Microbiol Immunol. 38(11):843–850.

Censini S, Lange C, Xiang Z, Crabtree JE, Ghiara P, Borodovsky M, Rappuoli R, Covacci A. 1996. *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. Proc Natl Acad Sci USA. 93(25):14648–14653.

Chan FKL, To KF, Wu JCY, Yung MY, Leung WK, Kwok T, Hui Y, Chan HLY, Chan CSY, Hui E et al. 2002. Eradication of *Helicobacter pylori* and risk of peptic ulcers in patients starting long-term treatment with non-steroidal anti-inflammatory drugs: a randomised trial. Lancet. 359(9300):9-13.

Chan W-Y, Hui P-K, Leung K-M, Chow J, Kwok F, Ng C-S. 1994. Coccoid forms of *Helicobacter pylori* in the human stomach. Am J Clin Pathol. 102(4):503–507.

Chang K-C, Ho S-W, Yang J-C, Wang J-T. 1997. Isolation of a genetic locus associated with metronidazole resistance in *Helicobacter pylori*. Biochem Biophys Res Comm. 236(3):785-788.

Chang W-L, Yeh Y-C, Sheu B-S. 2018. The impacts of *H. pylori* virulence factors on the development of gastroduodenal diseases. J Biomed Sci. 25(1):68.

Chang YW, Ko WJ, Oh CH, Park YM, Oh SJ, Moon JR, Cho J-H, Kim J-W, Jang J-Y. 2019. Clarithromycin resistance and female gender affect *Helicobacter pylori* eradication failure in chronic gastritis. Korean J Intern Med. 34(5):1022–1029.

Chattopadhyay S, Chi PB, Minin VN, Berg DE, Sokurenko EV. 2018. Recombination-independent rapid convergent evolution of the gastric pathogen *Helicobacter pylori*. BMC Genom. 19(1):835.

Chen L, Zheng D, Liu B, Yang J, Jin Q. 2016. VFDB 2016: hierarchical and refined dataset for big data analysis-10 years on. Nucleic Acids Res. 44(D1):D694-D697.

Chewapreecha C, Harris SR, Croucher NJ, Turner C, Marttinen P, Cheng L, Pessia A, Aanensen DM, Mather AE, Page AJ et al. 2014. Dense genomic sampling identifies highways of pneumococcal recombination. Nat Genet. 46(3):305-309.

Chey WD, Wong BCY. 2007. American College of Gastroenterology guideline on the management of *Helicobacter pylori* infection. Am J Gastroenterol Suppl. 102:1808.

Chi T, Zhang A, Zhang X, Li A-D, Zhang H, Zhao Z. 2020. Characteristics of the antibiotic resistance genes in the soil of medical waste disposal sites. Sci Total Environ. 730:139042.

Chisholm SA, Owen RJ. 2003. Mutations in *Helicobacter pylori rdxA* gene sequences may not contribute to metronidazole resistance. J Antimicrob Chemother. 51(4):995–999.

Cho H, Uehara T, Bernhardt TG. 2014. Beta-lactam antibiotics induce a lethal malfunctioning of the bacterial cell wall synthesis machinery. Cell. 159(6):1300-1311.

Choi SS, Chivers PT, Berg DE. 2011. Point mutations in *Helicobacter pylori's fur* regulatory gene that alter resistance to metronidazole, a prodrug activated by chemical reduction. PLoS One. 6(3):e18236.

Chopra I, Roberts M. 2001. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. Microbiol Mol Biol Rev. 65(2):232–260.

Chow TK-F, Lambert JR, Wahlqvist ML, Hsu-Hage BH-H. 1995. *Helicobacter pylori* in Melbourne Chinese immigrants: evidence for oral-oral transmission via chopsticks. J Gastroenterol Hepatol. 10(5):562–569.

Christofilogiannis P. 2001. Current inoculation methods in MIC determination. Aquaculture. 196(3):297–302.

Chu A, Wang D, Guo Q, Lv Z, Yuan Y, Gong Y. 2020. Molecular detection of *H. pylori* antibiotic-resistant genes and molecular docking analysis. The FASEB Journal. 34(1):610-618.

Chua E-G, Debowski AW, Webberley KM, Peters F, Lamichhane B, Loke M-F, Vadivelu J, Tay C-Y, Marshall BJ, Wise MJ. 2019. Analysis of core protein clusters identifies candidate variable sites conferring metronidazole resistance in *Helicobacter pylori*. Gastroenterol Rep. 7(1):42-49.

Chuah Y-Y, Wu D-C, Chuah S-K, Yang J-C, Lee T-H, Yeh H-Z, Chen C-L, Liu Y-H, Hsu P-I. 2017. Real-world practice and expectation of Asia-Pacific physicians and patients in *Helicobacter pylori* eradication (REAP-HP Survey). Helicobacter. 22(3):e12380.

Chuanchuen R, Narasaki CT, Schweizer HP. 2002. The MexJK efflux pump of *Pseudomonas aeruginosa* requires OprM for antibiotic efflux but not for efflux of triclosan. J Bacteriol. 184(18):5036-5044.

Clyne M, Labigne A, Drumm B. 1995. *Helicobacter pylori* requires an acidic environment to survive in the presence of urea. Infect Immun. 63(5):1669–1673.

Co E-MA, Schiller NL. 2006. Resistance mechanisms in an in vitro-selected amoxicillin-resistant strain of *Helicobacter pylori*. Antimicrob Agents Chemother. 50(12):4174–4176.

Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM. 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 42(Database issue):D633–D642.

Cole SP, Harwood J, Lee R, She R, Guiney DG. 2004. Characterization of monospecies biofilm formation by *Helicobacter pylori*. J Bacteriol. 186(10):3124–3132.

Contreras M, Thiberge J-M, Mandrand-Berthelot M-A, Labigne A. 2003. Characterization of the roles of NikR, a nickel-responsive pleiotropic autoregulator of *Helicobacter pylori*. Mol Microbiol. 49(4):947-963.

Coppens F, Castaldo G, Debraekeleer A, Subedi S, Moonens K, Lo A, Remaut H. 2018. Hop-family *Helicobacter* outer membrane adhesins form a novel class of Type 5-like secretion proteins with an interrupted β-barrel domain. Mol Microbiol. 110(1):33-46.

Costerton JW, Lewandowski Z, Caldwell DE, Korber DR, Lappin-Scott HM. 1995. Microbial biofilms. Annu Rev Microbiol. 49(1):711–745.

Coudron PE, Stratton CW. 1995. Factors affecting growth and susceptibility testing of *Helicobacter pylori* in liquid media. J Clin Microbiol. 33(4):1028–1030.

Courvalin P, Soussy CJ. 1996. Technical recommendations for in vitro susceptibility testing. Clin Microbiol Infect. 2:S11–S25.

Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, Rocha EPC, Vergnaud G, Gautheret D, Pourcel C. 2018. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 46(W1):W246–W251.

Cox MM. 1991. The RecA protein as a recombinational repair system. Mol Microbiol. 5(6):1295-1299.

Craig DW, Pearson JV, Szelinger S, Sekar A, Redman M, Corneveaux JJ, Pawlowski TL, Laub T, Nunn G, Stephan DA et al. 2008. Identification of genetic variants using bar-coded multiplexed sequencing. Nat Methods. 5(10):887–893.

Crowe SE. 2019. Treatment regimens for *Helicobacter pylori*. In: Feldman M, editor. UpToDate. Waltham, MA: UpToDate.

Dailidiene D, Bertoli MT, Miciuleviciene J, Mukhopadhyay AK, Dailide G, Pascasio MA, Kupcinskas L, Berg DE. 2002. Emergence of tetracycline resistance in *Helicobacter pylori*: multiple mutational changes in 16S ribosomal DNA and other genetic loci. Antimicrob Agents Chemother. 46(12):3940–3946.

Dalynn Biologicals. McFarland standard. 2014. [accessed 2020 April 17]. http://www.dalynn.com/dyn/ck_assets/files/tech/TM53.pdf.

Dandekar T, Schuster S, Snel B, Huynen M, Bork P. 1999. Pathway alignment: application to the comparative analysis of glycolytic enzymes. Biochemical Journal. 343(1):115-124.

Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14(7):1394–1403.

Darling ACE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One. 5(6):e11147.

Darling AE, Miklós I, Ragan MA. 2008. Dynamics of genome rearrangement in bacterial populations. PLoS Genet. 4(7):e1000128-e1000128.

De Francesco V, Giorgio F, Hassan C, Manes G, Vannella L, Panella C, Ierardi E, Zullo A. 2010. Worldwide *H. pylori* antibiotic resistance: a systematic review. J Gastrointestin Liver Dis. 19(4):409–414.

De Francesco V, Margiotta M, Zullo A, Hassan C, Giorgio F, Burattini O, Stoppino G, Cea U, Pace A, Zotti M et al. 2007. Prevalence of primary clarithromycin resistance in *Helicobacter pylori* strains over a 15 year period in Italy. J Antimicrob Chemother. 59(4):783–785.

De Francesco V, Zullo A, Hassan C, Giorgio F, Rosania R, Ierardi E. 2011. Mechanisms of *Helicobacter pylori* antibiotic resistance: an updated appraisal. World J Gastrointest Pathophysiol. 2(3):35–41.

De Francesco V, Zullo A, Ierardi E, Giorgio F, Perna F, Hassan C, Morini S, Panella C, Vaira D. 2009a. Phenotypic and genotypic *Helicobacter pylori* clarithromycin resistance and therapeutic outcome: benefits and limits. J Antimicrob Chemother. 65(2):327–332.

De Francesco V, Zullo A, Ierardi E, Vaira D. 2009b. Minimal inhibitory concentration (MIC) values and different point mutations in the 23S rRNA gene for clarithromycin resistance in *Helicobacter pylori*. Dig Liver Dis. 41(8):610-611.

de Martel C, Parsonnet J. 2006. *Helicobacter pylori* infection and gender: a meta-analysis of population-based prevalence surveys. Dig Dis Sci. 51(12):2292–2301.

De Vries AC, Van Driel HF, Richardus JH, Ouwendijk M, Van Vuuren AJ, De Man RA, Kuipers EJ. 2008. Migrant communities constitute a possible target population for primary prevention of *Helicobacter pylori*-related complications in low incidence countries. Scand J Gastroenterol. 43(4):403-409.

Debets-Ossenkopp YJ, Brinkman AB, Kuipers EJ, Vandenbroucke-Grauls CM, Kusters JG. 1998. Explaining the bias in the 23S rRNA gene mutations associated with clarithromycin resistance in clinical isolates of *Helicobacter pylori*. Antimicrob Agents Chemother. 42(10):2749-2751.

Debets-Ossenkopp YJ, Pot RGJ, van Westerloo DJ, Goodwin A, Vandenbroucke-Grauls CMJE, Berg DE, Hoffman PS, Kusters JG. 1999. Insertion of mini-IS*605* and deletion of adjacent sequences in the nitroreductase *rdxA* gene cause metronidazole resistance in *Helicobacter pylori* NCTC11637. Antimicrob Agents Chemother. 43(11):2657-2662.

Della Bruna C, Schioppacassi G, Ungheri D, Jabès D, Morvillo E, Sanfilippo A. 1983. LM 427, a new spiropiperidylrifamycin: *in vitro* and *in vivo* studies. J Antibiot. 36(11):1502-1506.

Delmar JA, Su C-C, Yu EW. 2014. Bacterial multidrug efflux transporters. Annual review of biophysics. 43:93-117.

DeLoney CR, Schiller NL. 2000. Characterization of an *in vitro*-selected amoxicillin-resistant strain of *Helicobacter pylori*. Antimicrob Agents Chemother. 44(12):3368-3373.

Desai HG, Gill HH, Shankaran K, Mehta PR, Prabhu SR. 1991. Dental plaque: a permanent reservoir of *Helicobacter pylori*? Scand J Gastroenterol. 26(11):1205–1208.

Destro MT, Leitão MF, Farber JM. 1996. Use of molecular typing methods to trace the dissemination of *Listeria monocytogenes* in a shrimp processing plant. Appl Environ Microbiol. 62(2):705-711.

Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, García-Cobos S, Kooistra-Smid AMD, Raangs EC, Rosema S, Veloo ACM et al. 2017. Application of next generation sequencing in clinical microbiology and infection prevention. J Biotechnol. 243:16-24.

Dewhirst FE, Shen Z, Scimeca MS, Stokes LN, Boumenna T, Chen T, Paster BJ, Fox JG. 2005. Discordant 16S and 23S rRNA gene phylogenies for the genus *Helicobacter*: implications for phylogenetic inference and systematics. J Bacteriol. 187(17):6106–6118.

Didelot X, Nell S, Yang I, Woltemate S, van der Merwe S, Suerbaum S. 2013. Genomic evolution and transmission of *Helicobacter pylori* in two South African families. Proc Natl Acad Sci USA. 110(34):13880-13885.

Dierikx CM, Martodihardjo J, Kuipers EJ, Hensgens CMH, Kusters JG, Suzuki H, De Groot N, van Vliet AHM. 2007. Serum- and animal tissue-free medium for transport and growth of *Helicobacter pylori*. FEMS Immunol Med Microbiol. 50(2):239–243.

Dixon MF, Genta RM, Yardley JH, Correa P. 1996. Classification and grading of gastritis: the updated Sydney system. Am J Surg Pathol. 20(10):1161–1181.

Doig P, Exner MM, Hancock RE, Trust TJ. 1995. Isolation and characterization of a conserved porin protein from *Helicobacter pylori*. J Bacteriol. 177(19):5447-5452.

Doostparast Torshizi A, Wang K. 2018. Next-generation sequencing in drug development: target identification and genetically stratified clinical trials. Drug Discov Today. 23(10):1776-1783.

Dore MP, Osato MS, Realdi G, Mura I, Graham DY, Sepulveda AR. 1999. Amoxycillin tolerance in *Helicobacter pylori*. J Antimicrob Chemother. 43(1):47–54.

Dore MP, Sepulveda AR, El-Zimaity H, Yamaoka Y, Osato MS, Mototsugu K, Nieddu AM, Realdi G, Graham DY. 2001. Isolation of *Helicobacter pylori* from sheep—implications for transmission to humans. Am J Gastroenterol Suppl. 96(5):1396-1401.

Dossumbekova A, Prinz C, Mages J, Lang R, Kusters JG, Van Vliet AHM, Reindl W, Backert S, Saur D, Schmid RM et al. 2006. *Helicobacter pylori* HopH (OipA) and bacterial pathogenicity:

genetic and functional genomic analysis of hopH gene polymorphisms. Int J Infect Dis. 194(10):1346–1355.

Douthwaite S, Aagaard C. 1993. Erythromycin binding is reduced in Rribosomes with conformational alterations in the 23 S rRNA peptidyl transferase loop. J Mol Biol. 232(3):725-731.

Drazek ES, Dubois A, Holmes RK. 1994. Characterization and presumptive identification of *Helicobacter pylori* isolates from rhesus monkeys. J Clin Microbiol. 32(7):1799–1804.

Drini M. 2017. Peptic ulcer disease and non-steroidal anti-inflammatory drugs. Aust Prescr. 40(3):91–93.

Drlica K. 1999. Mechanism of fluoroquinolone action. Curr Opin Microbiol. 2(5):504–508.

Duncan SS, Bertoli MT, Kersulyte D, Valk PL, Tamma S, Segal I, McClain MS, Cover TL, Berg DE. 2013. Genome sequences of three hpAfrica2 strains of *Helicobacter pylori*. Genome Announc. 1(5):e00729–00713.

Duncan SS, Valk PL, Shaffer CL, Bordenstein SR, Cover TL. 2012. J-Western Forms of *Helicobacter pylori cagA* constitute a distinct phylogenetic group with a widespread geographic distribution. J Bacteriol. 194(6):1593-1604.

Dunne C, Dolan B, Clyne M. 2014. Factors that mediate colonization of the human stomach by *Helicobacter pylori*. World J Gastroenterol. 20(19):5610–5624.

Eaton KA, Morgan DR, Krakowka S. 1989. *Campylobacter pylori* virulence factors in gnotobiotic piglets. Infect Immun. 57(4):1119–1125.

Eaton KA, Suerbaum S, Josenhans C, Krakowka S. 1996. Colonization of gnotobiotic piglets by *Helicobacter pylori* deficient in two flagellin genes. Infect Immun. 64(7):2445–2448.

Ebili HO, Hassall JC, Fadhil W, Ham-Karim H, Asiri A, Raposo TP, Agboola AJ, Ilyas M. 2017. "Squirrel" primer-based PCR assay for direct and targeted Sanger sequencing of short genomic segments. J Biomol Tech. 28(3):97–110.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–1797.

Edwards DI. 1980. Mechanisms of selective toxicity of metronidazole and other nitroimidazole drugs. Br J Vener Dis. 56(5):285–290.

Edwards DI. 1986. Reduction of nitroimidazoles *in vitro* and DNA damage. Biochem Pharmacol. 35(1):53-58.

Edwards DI. 1993. Nitroimidazole drugs - action and resistance mechanisms. I. Mechanisms of action. J Antimicrob Chemother. 31(1):9-20.

Egan BJ, Holmes K, O'Connor HJ, O'Morain CA. 2007. *Helicobacter pylori* gastritis, the unifying concept for gastric diseases. Helicobacter. 12(s2):39–44.

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B et al. 2009. Real-time DNA sequencing from single polymerase molecules. Science. 323(5910):133-138.

Ekblom R, Wolf JBW. 2014. A field guide to whole-genome sequencing, assembly and annotation. Evol Appl. 7(9):1026–1042.

Enright MC, Day NPJ, Davies CE, Peacock SJ, Spratt BG. 2000. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. J Clin Microbiol. 38(3):1008-1015.

Escobar ML, Kawakami E. 2004. Evidence of mother-child transmission of *Helicobacter pylori* infection. Arq Gastroenterol. 41:239-244.

Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8(3):186-194.

Falsafi T, Ehsani A, Attaran B, Niknam V. 2016. Association of *hp1181* and *hp1184* genes with the active efflux phenotype in multidrug-resistant isolates of *Helicobacter pylori*. Jundishapur J Microbiol. 9(4):e30726-e30726.

Falush D, Kraft C, Taylor NS, Correa P, Fox JG, Achtman M, Suerbaum S. 2001. Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. Proc Natl Acad Sci USA. 98(26):15056-15061.

Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, Blaser MJ, Graham DY, Vacher S, Perez-Perez GI et al. 2003. Traces of human migrations in *Helicobacter pylori* populations. Science. 299(5612):1582–1585.

Fan X, Li Y, He R, Li Q, He W. 2016. Comparative analysis of prophage-like elements in *Helicobacter sp*. genomes. PeerJ. 4:e2012.

Feldman M, Lee EL. 2016. Gastritis. In: Feldman MMD, Friedman LSMD, Brandt LJMD, editors. Sleisenger and Fordtran's Gastrointestinal and Liver Disease. p. 868–883.

Ferran A, Dupouy V, Toutain P-L, Bousquet-Mélou A. 2007. Influence of inoculum size on the selection of resistant mutants of *Escherichia coli* in relation to mutant prevention concentrations of marbofloxacin. Antimicrob Agents Chemother. 51(11):4163-4166.

Fischer W, Breithaupt U, Kern B, Smith SI, Spicher C, Haas R. 2014. A comprehensive analysis of *Helicobacter pylori* plasticity zones reveals that they are integrating conjugative elements with intermediate integration specificity. BMC Genom. 15(1):310.

Fish DN, Chow AT. 1997. The clinical pharmacokinetics of levofloxacin. Clin Pharmacokinet. 32(2):101–119.

Fontana C, Favaro M, Minelli S, Criscuolo AA, Pietroiusti A, Galante A, Favalli C. 2002. New site of modification of 23S rRNA associated with clarithromycin resistance of *Helicobacter pylori* clinical isolates. Antimicrob Agents Chemother. 46(12):3765-3769.

Fournier P-E, Vallenet D, Barbe V, Audic S, Ogata H, Poirel L, Richet H, Robert C, Mangenot S, Abergel C et al. 2006. Comparative genomics of multidrug resistance in *Acinetobacter baumannii*. PLoS Genet. 2(1):e7.

Fraser AG, Moore L, Hackett M, Hollis B. 1999. *Helicobacter pylori* treatment and antibiotic susceptibility: results of a five-year audit. Austr NZ J Med. 29(4):512–516.

Frost LS, Leplae R, Summers AO, Toussaint A. 2005. Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol. 3(9):722–732.

Fujimura S, Kawamura T, Kato S, Tateno H, Watanabe A. 2002. Detection of *Helicobacter pylori* in cow's milk. Lett Appl Microbiol. 35(6):504-507.

Gaasbeek EJ, van der Wal FJ, van Putten JPM, de Boer P, van der Graaf-van Bloois L, de Boer AG, Vermaning BJ, Wagenaar JA. 2009. Functional characterization of excision repair and RecA-dependent recombinational DNA repair in *Campylobacter jejuni*. J Bacteriol. 191(12):3785-3793.

Gabaldón T. 2008. Large-scale assignment of orthology: back to phylogenetics? Genome Biol. 9(10):235.

Gabaldón T, Koonin EV. 2013. Functional and evolutionary implications of gene orthology. Nat Rev Genet. 14(5):360–366.

Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 43(Database issue):D261-D269.

García-Ortíz M-V, Marsin S, Arana ME, Gasparutto D, Guérois R, Kunkel TA, Radicella JP. 2011. Unexpected role for *Helicobacter pylori* DNA polymerase I as a source of genetic variability. PLoS Genet. 7(6):e1002152.

García-Zea JA, de la Herrán R, Robles Rodríguez F, Navajas-Pérez R, Ruiz Rejón C. 2019. Detection and variability analyses of CRISPR-like loci in the *H. pylori* genome. PeerJ. 7:e6221–e6221.

Garrido L, Toledo H. 2007. Novel genotypes in *Helicobacter pylori* involving domain V of the 23S rRNA gene. Helicobacter. 12(5):505–509.

Gebara ECE, Pannuti C, Faria CM, Chehter L, Mayer MPA, Lima LAPA. 2004. Prevalence of *Helicobacter pylori* detected by polymerase chain reaction in the oral cavity of periodontitis patients. Oral Microbiol Immunol. 19(4):277–280.

Gebert B, Fischer W, Weiss E, Hoffmann R, Haas R. 2003. *Helicobacter pylori* vacuolating cytotoxin inhibits T lymphocyte activation. Science. 301(5636):1099–1102.

Geis G, Leying H, Suerbaum S, Mai U, Opferkuch W. 1989. Ultrastructure and chemical analysis of *Campylobacter pylori* flagella. J Clin Microbiol. 27(3):436–441.

Gemilyan M, Hakobyan G, Benejat L, Allushi B, Melik-Nubaryan D, Mangoyan H, Laur A, Daguerre E, Grigoryan E, Megraud F. 2019. Prevalence of *Helicobacter pylori* infection and antibiotic resistance profile in Armenia. Gut Pathog. 11(1):28.

Gene Ontology C. 2015. Gene Ontology Consortium: going forward. Nucleic Acids Res. 43(Database issue):D1049-D1056.

Gerhard M, Lehn N, Neumayer N, Borén T, Rad R, Schepp W, Miehlke S, Classen M, Prinz C. 1999. Clinical relevance of the *Helicobacter pylori* gene for blood-group antigen-binding adhesin. Proc Natl Acad Sci USA. 96(22):12778-12783.

Gerrits MM, Berning M, Van Vliet AHM, Kuipers EJ, Kusters JG. 2003. Effects of 16S rRNA gene mutations on tetracycline resistance in *Helicobacter pylori*. Antimicrob Agents Chemother. 47(9):2984–2986.

Gerrits MM, Godoy APO, Kuipers EJ, Ribeiro ML, Stoof J, Mendonça S, Van Vliet AHM, Pedrazzoli Jr J, Kusters JG. 2006. Multiple mutations in or adjacent to the conserved penicillin-binding protein motifs of the penicillin-binding protein 1A confer amoxicillin resistance to *Helicobacter pylori*. Helicobacter. 11(3):181–187.

Ghanei H, Abeyrathne PD, Lam JS. 2007. Biochemical characterization of MsbA from *Pseudomonas aeruginosa*. J Biol Chem. 282(37):26939-26947.

Ghosh A, N S, Saha S. 2020. Survey of drug resistance associated gene mutations in *Mycobacterium tuberculosis*, ESKAPE and other bacterial species. Sci Rep. 10(1):8957.

Giannakis M, Chen SL, Karam SM, Engstrand L, Gordon JI. 2008. *Helicobacter pylori* evolution during progression from chronic atrophic gastritis to gastric cancer and its impact on gastric stem cells. PNAS. 105(11):4358-4363.

Gião MS, Azevedo NF, Wilks SA, Vieira MJ, Keevil CW. 2008. Persistence of *Helicobacter pylori* in heterotrophic drinking-water biofilms. Appl Environ Microbiol. 74(19):5898–5904.

Gibbons S, Oluwatuyi M, Kaatz GW. 2003. A novel inhibitor of multidrug efflux pumps in *Staphylococcus aureus*. J Antimicrob Chemother. 51(1):13-17.

Gilani A, Razavilar V, Rokni N, Rahimi E. 2017. VacA and cagA genotypes status and antimicrobial resistance properties of *Helicobacter pylori* strains isolated from meat products in Isfahan province, Iran. Iran J Vet Res. 18(2):97-102.

Gisbert JP, Calvet X. 2012. Update on non-bismuth quadruple (concomitant) therapy for eradication of *Helicobacter pylori*. Clin Exp Gastroenterol. 5:23–34.

Gisbert JP, Calvet X, Bujanda L, Marcos S, Gisbert JL, Pajares JM. 2003. 'Rescue' therapy with rifabutin after multiple *Helicobacter pylori* treatment failures. Helicobacter. 8(2):90-94.

Gisbert JP, Castro-Fernández M, Bermejo F, Pérez-Aisa A, Ducons J, Fernández-Bermejo M, Bory F, Cosme A, Benito L-M, López-Rivas L. 2006. Third-line rescue therapy with levofloxacin after two *H. pylori* treatment failures. Am J Gastroenterol Suppl. 101(2):243.

Gisbert JP, Gisbert JL, Marcos S, Olivares D, Pajares JM. 2005. *Helicobacter pylori* first-line treatment and rescue options in patients allergic to penicillin. Aliment Pharmacol Therapeut. 22(10):1041–1046.

Glocker E, Stueger H-P, Kist M. 2007. Quinolone resistance in *Helicobacter pylori* isolates in Germany. Antimicrob Agents Chemother. 51(1):346–349.

Glupczynski Y, Broutet N, Cantagrel A, Andersen L, Alarcon T, López-Brea M, Mégraud F. 2002. Comparison of the Etest and agar dilution method for antimicrobial suceptibility testing of *Helicobacter pylori*. Eur J Clin Microbiol Infect Dis. 21(7):549–552.

Glupczynski Y, Labbé M, Hansen W, Crokaert F, Yourassowsky E. 1991. Evaluation of the E test for quantitative antimicrobial susceptibility testing of *Helicobacter pylori*. J Clin Microbiol. 29(9):2072–2075.

Godoy APO, Reis FC, Ferraz LFC, Gerrits MM, Mendonça S, Kusters JG, Ottoboni LMM, Ribeiro ML, Pedrazzoli Jr J. 2007. Differentially expressed genes in response to amoxicillin in *Helicobacter pylori* analyzed by RNA arbitrarily primed PCR. FEMS Immunol Med Microbiol. 50(2):226-230.

Goldman P, Koch RL, Yeung T-C, Chrystal EJT, Beaulieu BB, McLafferty MA, Sudlow G. 1986. Comparing the reduction of nitroimidazoles in bacteria and mammalian tissues and relating it to biological activity. Biochem Pharmacol. 35(1):43-51.

Gomollón F, Ducóns JA, Ferrero M, Cabezudo JG, Guirao R, Simón MA, Montoro M. 1999. Quadruple therapy is effective for eradicating *Helicobacter pylori* after failure of triple proton-pump inhibitor–based therapy: a detailed, prospective analysis of 21 consecutive cases. Helicobacter. 4(4):222–225.

Gong EJ, Ahn JY, Kim JM, Lee SM, Na HK, Lee JH, Jung KW, Choi KD, Kim DH, Song HJ et al. 2020. Genotypic and phenotypic resistance to clarithromycin in *Helicobacter pylori* strains. J Clin Med. 9(6):1930.

Gontcharov AA, Marin B, Melkonian M. 2004. Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and *rbc*L sequence comparisons in the Zygnematophyceae (Streptophyta). Mol Biol Evol. 21(3):612-624.

Goodman KJ, Correa P. 2000. Transmission of *Helicobacter pylori* among siblings. Lancet. 355(9201):358–362.

Goodman KJ, Correa P, Aux HJT, Ramirez H, DeLany JP, Pepinosa OG, Quiñones ML, Parra TC. 1996. *Helicobacter pylori* infection in the Colombian Andes: a population-based study of transmission pathways. Am J Epidemiol. 144(3):290–299.

Goodwin A, Kersulyte D, Sisson G, Veldhuyzen van Zanten SJO, Berg DE, Hoffman PS. 1998. Metronidazole resistance in *Helicobacter pylori* is due to null mutations in a gene (*rdxA*) that encodes an oxygen-insensitive NADPH nitroreductase. Mol Microbiol. 28(2):383–393.

Goodwin CS, Armstrong JA, Chilvers T, Peters M, Collins MD, Sly L, McConnell W, Harper WES. 1989. Transfer of *Campylobacter pylori* and *Campylobacter mustelae* to *Helicobacter* gen. nov. as *Helicobacter pylori* comb. nov. and *Helicobacter mustelae* comb. nov., respectively. Int J Syst Evol Microbiol. 39(4):397–405.

Goodwin CS, Blincow ED, Warren JR, Waters TE, Sanderson CR, Easton L. 1985a. Evaluation of cultural techniques for isolating *Campylobacter pyloridis* from endoscopic biopsies of gastric mucosa. J Clin Pathol. 38(10):1127–1131.

Goodwin CS, McCulloch RK, Armstrong JA, Wee SH. 1985b. Unusual cellular fatty acids and distinctive ultrastructure in a new spiral bacterium (*Campylobacter Pyloridis*) from the human gastric mucosa. J Med Microbiol. 19(2):257–267.

Gorbach SL. 1990. Bismuth therapy in gastrointestinal diseases. Gastroenterology. 99(3):863-875.

Graham DY, Calvet X. 2012. Guide regarding choice of second-line therapy to obtain a high cumulative cure rate. Helicobacter. 17(4):243–245.

Graham DY, Fischbach L. 2010. *Helicobacter pylori* treatment in the era of increasing antibiotic resistance. Gut. 59(8):1143–1153.

Graham DY, Lew GM, Malaty HM, Evans DG, Evans DJ, Klein PD, Alpert LC, Genta RM. 1992. Factors influencing the eradication of *Helicobacter pylori* with triple therapy. Gastroenterology. 102(2):493–496.

Graham DY, Lu H, Yamaoka Y. 2007. A report card to grade *Helicobacter pylori* therapy. Helicobacter. 12(4):275–278.

Graham DY, Malaty HM, Evans DG, Evans DJ, Jr., Klein PD, Adam E. 1991. Epidemiology of *Helicobacter pylori* in an asymptomatic population in the United States: effect of age, race, and socioeconomic status. Gastroenterology. 100(6):1495–1501.

Gramley WA, Asghar A, Frierson HF, Powell SM. 1999. Detection of *Helicobacter pylori* DNA in fecal samples from infected individuals. J Clin Microbiol. 37(7):2236-2240.

Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, Stryke D, Bouquet J, Somasekar S, Linnen JM et al. 2015. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med. 7(1):99.

Grevskott DH, Salvà-Serra F, Moore ERB, Marathe NP. 2020. Nanopore sequencing reveals genomic map of CTX-M-type extended-spectrum β-lactamases carried by *Escherichia*

coli strains isolated from blue mussels (*Mytilus edulis*) in Norway. BMC Microbiol. 20(1):134.

Grignon B, Tankovic J, Mégraud F, Glupczynski Y, Husson MO, Conroy MC, Emond JP, Loulergue J, Raymond J, Fauchère JL. 2002. Validation of diffusion methods for macrolide susceptibility testing of *Helicobacter pylori*. Microb Drug Resist. 8(1):61-66.

Grübel P, Hoffman JS, Chong FK, Burstein NA, Mepani C, Cave DR. 1997. Vector potential of houseflies (*Musca domestica*) for *Helicobacter pylori*. J Clin Microbiol. 35(6):1300–1303.

Guay DRP, Gustavson LE, Devcich KJ, Zhang J, Cao G, Olson CA. 2001. Pharmacokinetics and tolerability of extended-release clarithromycin. Clin Ther. 23(4):566-577.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59(3):307–321.

Hachem CY, Clarridge JE, Reddy R, Flamm R, Evans DG, Tanaka SK, Graham DY. 1996. Antimicrobial susceptibility testing of *Helicobacter pylori* comparison of E-test, broth microdilution, and disk diffusion for ampicillin, clarithromycin, and metronidazole. Diagn Microbiol Infect Dis. 24(1):37–41.

Hammerum AM, Hansen F, Skov MN, Stegger M, Andersen PS, Holm A, Jakobsen L, Justesen US. 2015. Investigation of a possible outbreak of carbapenem-resistant *Acinetobacter baumannii* in Odense, Denmark using PFGE, MLST and whole-genome-based SNPs. J Antimicrob Chemother. 70(7):1965-1968.

Han F, Liu S, Ho B, Yan Z, Yan X. 2007. Alterations in *rdxA* and *frxA* genes and their upstream regions in metronidazole-resistant *Helicobacter pylori* isolates. Microbiol Res. 158(1):38–44.

Hänninen M-L, Hannula M. 2007. Spontaneous mutation frequency and emergence of ciprofloxacin resistance in *Campylobacter jejuni* and *Campylobacter coli*. J Antimicrob Chemother. 60(6):1251–1257.

Hansomburana P, Anantapanpong S, Sirinthornpunya S, Chuengyong K, Rojborwonwittaya J. 2012. Prevalence of single nucleotide mutation in clarithromycin resistant gene of *Helicobacter pylori*: a 32-months prospective study by using hybridization real time polymerase chain reaction. J Med Assoc Thai. 95 Suppl 3:S28-35.

Hao Q, Li Y, Zhang Z-J, Liu Y, Gao H. 2004. New mutation points in 23S rRNA gene associated with *Helicobacter pylori* resistance to clarithromycin in northeast China. World J Gastroenterol. 10(7):1075–1077.

Happel A-U, Kullin B, Gamieldien H, Wentzel N, Zauchenberger CZ, Jaspan HB, Dabee S, Barnabas SL, Jaumdally SZ, Dietrich J et al. 2020. Exploring potential of vaginal *Lactobacillus*

isolates from South African women for enhancing treatment for bacterial vaginosis. PLoS Pathog. 16(6):e1008559.

Harper CMG, Dangler CA, Xu S, Feng Y, Shen Z, Sheppard B, Stamper A, Dewhirst FE, Paster BJ, Fox JG. 2000. Isolation and characterization of a *Helicobacter* sp. from the gastric mucosa of dolphins, *Lagenorhynchus acutus* and *Delphinus delphis*. Appl Environ Microbiol. 66(11):4751–4757.

Harrington CT, Lin EI, Olson MT, Eshleman JR. 2013. Fundamentals of Pyrosequencing. Arch Pathol Lab Med. 137(9):1296–1303.

Harris AG, Hazell SL, Netting AG. 2000. Use of digoxigenin-labelled ampicillin in the identification of penicillin-binding proteins in *Helicobacter pylori*. J Antimicrob Chemother. 45(5):591-598.

Hartzen SH, Andersen LP, Bremmelgaard A, Colding H, Arpi M, Kristiansen J, Justesen T, Espersen F, Frimodt-Møller N, Bonnevie O. 1997. Antimicrobial susceptibility testing of 230 *Helicobacter pylori* strains: importance of medium, inoculum, and incubation time. Antimicrob Agents Chemother. 41(12):2634–2639.

Hatakeyama M. 2014. *Helicobacter pylori* CagA and gastric cancer: a paradigm for hit-and-run carcinogenesis. Cell Host Microbe. 15(3):306–316.

Hawkey PM. 1998. The origins and molecular basis of antibiotic resistance. Br Med J. 317(7159):657–660.

Hazell SL, Graham DY. 1990. Unsaturated fatty acids and viability of *Helicobacter* (*Campylobacter*) *pylori*. J Clin Microbiol. 28(5):1060–1061.

He L, Deng T, Luo H. 2015. Meta-analysis of sequential, concomitant and hybrid therapy for *Helicobacter pylori* eradication. Intern Med. 54(7):703–710.

Heep M, Lehn N, Brandstätter B, Rieger U, Senzenberger S, Wehrl W. 2002. Detection of rifabutin resistance and association of *rpoB* mutations with resistance to four rifamycin derivatives in *Helicobacter pylori*. Eur J Clin Microbiol Infect Dis. 21(2):143–145.

Hemmatinezhad B, Momtaz H, Rahimi E. 2016. *VacA*, *cagA*, *iceA* and *oipA* genotypes status and antimicrobial resistance properties of *Helicobacter pylori* isolated from various types of ready to eat foods. Ann Clin Microbiol Antimicrob. 15:2-2.

Hendriksen RS, Bortolaia V, Tate H, Tyson GH, Aarestrup FM, McDermott PF. 2019. Using genomics to track global antimicrobial resistance. Front Public Health. 7(242).

Hennig EE, Allen JM, Cover TL. 2006. Multiple chromosomal loci for the *babA* gene in *Helicobacter pylori*. Infect Immun. 74(5):3046–3051.

Hirata K, Suzuki H, Nishizawa T, Tsugawa H, Muraoka H, Saito Y, Matsuzaki J, Hibi T. 2010. Contribution of efflux pumps to clarithromycin resistance in *Helicobacter pylori*. J Gastroenterol Hepatol. 25(s1):S75–S79.

Hodkinson BP, Grice EA. 2015. Next-generation sequencing: a review of technologies and tools for wound microbiome research. Adv Wound Care (New Rochelle). 4(1):50–58.

Hoffmann S, Otto C, Kurtz S, Sharma CM, Khaitovich P, Vogel J, Stadler PF, Hackermüller J. 2009. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. PLoS Comput Biol. 5(9):e1000502-e1000502.

Hofreuter D, Haas R. 2002. Characterization of two cryptic *Helicobacter pylori* plasmids: a putative source for horizontal gene transfer and gene shuffling. J Bacteriol. 184(10):2755–2766.

Hollister EB, Brooks JP, Gentry TJ. 2015. Bioinformation and 'omic approaches for characterization of environmental microorganisms. In: Pepper IL, Gerba CP, Gentry TJ, editors. Environmental Microbiology (Third Edition). San Diego: Academic Press. p. 483-505.

Hombach M, Maurer FP, Pfiffner T, Böttger EC, Furrer R. 2015. Standardization of operator-dependent variables affecting precision and accuracy of the disk diffusion method for antibiotic susceptibility testing. J Clin Microbiol. 53(12):3864–3869.

Hooi JKY, Lai WY, Ng WK, Suen MMY, Underwood FE, Tanyingoh D, Malfertheiner P, Graham DY, Wong VWS, Wu JCY et al. 2017. Global prevalence of *Helicobacter pylori* infection: systematic review and meta-analysis. Gastroenterology. 153(2):420–429.

Hooper DC. 1999. Mechanisms of fluoroquinolone resistance. Drug Resist Updates. 2(1):38–55.

Hopkins RJ, Vial PA, Ferreccio C, Ovalle J, Prado P, Sotomayor V, Russell RG, Wasserman SS, Morris JG, Jr. 1993. Seroprevalence of *Helicobacter pylori* in Chile: vegetables may serve as one route of transmission. Int J Infect Dis. 168(1):222-226.

Horiki N, Omata F, Uemura M, Suzuki S, Ishii N, Iizuka Y, Fukuda K, Fujita Y, Katsurahara M, Ito T et al. 2009. Annual change of primary resistance to clarithromycin among *Helicobacter pylori* isolates from 1996 through 2008 in Japan. Helicobacter. 14(5):438–442.

Horiuchi T, Ohkusa T, Watanabe M, Kobayashi D, Miwa H, Eishi Y. 2001. *Helicobacter pylori* DNA in drinking water in Japan. Microbiol Immunol. 45(7):515–519.

Howden CW, Hunt RH. 1998. Guidelines for the management of *Helicobacter pylori* infection. Am J Gastroenterol Suppl. 93:2330.

Hoy B, Löwer M, Weydig C, Carra G, Tegtmeyer N, Geppert T, Schröder P, Sewald N, Backert S, Schneider G et al. 2010. *Helicobacter pylori* HtrA is a new secreted virulence factor that cleaves E-cadherin to disrupt intercellular adhesion. EMBO Rep. 11(10):798–804.

Hsiang J, Selvaratnam S, Taylor S, Yeoh J, Tan Y-M, Huang J, Patrick A. 2013. Increasing primary antibiotic resistance and ethnic differences in eradication rates of *Helicobacter pylori* infection in New Zealand-a new look at an old enemy. N Z Med J. 126(1384):64–76.

Hsu P-I, Wu D-C, Wu J-Y, Graham DY. 2011. Modified sequential *Helicobacter pylori* therapy: proton pump inhibitor and amoxicillin for 14 days with clarithromycin and metronidazole added as a quadruple (hybrid) therapy for the final 7 days. Helicobacter. 16(2):139–145.

Huang J-Q, Sridhar S, Hunt RH. 2002. Role of *Helicobacter pylori* infection and non-steroidal anti-inflammatory drugs in peptic-ulcer disease: a meta-analysis. Lancet. 359(9300):14-22.

Huang JQ, Hunt R. 1996. pH, healing rate and symptom relief in acid-related diseases. Yale J Biol Med. 69(2):159.

Huang Julie Y, Sweeney Emily G, Sigal M, Zhang HC, Remington SJ, Cantrell Michael A, Kuo Calvin J, Guillemin K, Amieva Manuel R. 2015. Chemodetection and destruction of host urea allows *Helicobacter pylori* to locate the epithelium. Cell Host Microbe. 18(2):147-156.

Huang S, Vieira S, Bunk B, Riedel T, Spröer C, Overmann J. 2016a. First complete genome sequence of a subdivision 6 *Acidobacterium* strain. Genome Announc. 4(3):e00469–00416.

Huang X, Deng Z, Zhang Q, Li W, Wang B, Li M. 2016b. Relationship between the *iceA* gene of *Helicobacter pylori* and clinical outcomes. Ther Clin Risk Manag. 12:1085–1092.

Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast genome-gide functional annotation through orthology assignment by eggNOG-Mapper. Mol Biol Evol. 34(8):2115-2122.

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen Lars J et al. 2018. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47(D1):D309-D314.

Hultén K, Gibreel A, Sköld O, Engstrand L. 1997. Macrolide resistance in *Helicobacter pylori*: mechanism and stability in strains from clarithromycin-treated patients. Antimicrob Agents Chemother. 41(11):2550–2553.

Hultén K, Han SW, Enroth H, Klein PD, Opekun AR, Gilman RH, Evans DG, Engstrand L, Graham DY, El-Zaatari FA. 1996. *Helicobacter pylori* in the drinking water in Peru. Gastroenterology. 110(4):1031–1035.

Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 11:119-119.

Ihamauki T, Kekki M, Sipponen P, Siurala M. 1985. The sequelae and curse of cronic gstritis during a 30- to 34-yar boptic fllow-up sudy. Scand J Gastroenterol. 20(4):485–491.

Illumina. 2017. An introduction to next-generation sequencing technology. Illumina; [accessed 2 March 2020]. www.illumina.com/technology/next-generation-sequencing.html.

Illumina. Read length recommendations. 2020. [accessed 2020 2 March]. https://sapac.illumina.com/science/technology/next-generation-sequencing/plan-experiments/read-length.html?langsel=/nz/.

Ilver D, Arnqvist A, Ögren J, Frick I-M, Kersulyte D, Incecik ET, Berg DE, Covacci A, Engstrand L, Borén T. 1998. *Helicobacter pylori* adhesin binding fucosylated histo-blood group antigens revealed by retagging. Science. 279(5349):373–377.

Imamura S, Kita M, Yamaoka Y, Yamamoto T, Ishimaru A, Konishi H, Wakabayashi N, Mitsufuji S, Okanoue T, Imanishi J. 2003. Vector potential of cockroaches for *Helicobacter pylori* infection. Am J Gastroenterol Suppl. 98(7):1500–1503.

Ings RMJ, McFadzean JA, Ormerod WE. 1974. The mode of action of metronidazole in *Trichomonas vaginalis* and other micro-organisms. Biochem Pharmacol. 23(9):1421-1429.

Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ, Tomita T, Zobel J, Holt KE. 2014. SRST2: rapid genomic surveillance for public health and hospital microbiology labs. Genome Med. 6(11):90-90.

Iovene MR, Romano M, Pilloni AP, Giordano B, Montella F, Caliendo S, Tufano MA. 1999. Prevalence of antimicrobial resistance in eighty clinical isolates of *Helicobacter pylori*. Chemotherapy. 45(1):8–14.

Itskoviz D, Boltin D, Leibovitzh H, Tsadok Perets T, Comaneshter D, Cohen A, Niv Y, Levi Z. 2017. Smoking increases the likelihood of *Helicobacter pylori* treatment failure. Dig Liver Dis. 49(7):764–768.

Iwamoto A, Tanahashi T, Okada R, Yoshida Y, Kikuchi K, Keida Y, Murakami Y, Yang L, Yamamoto K, Nishiumi S et al. 2014. Whole-genome sequencing of clarithromycin resistant *Helicobacter pylori* characterizes unidentified variants of multidrug resistant efflux pump genes. Gut Pathog. 6(1):27.

Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT et al. 2018. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol. 36(4):338–345.

Jain M, Olsen HE, Paten B, Akeson M. 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol. 17(1):239.

Jain R, Rivera MC, Lake JA. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. Proc Natl Acad Sci USA. 96(7):3801-3806.

Jenks PJ, Ferrero RL, Labigne A. 1999. The role of the *rdxA* gene in the evolution of metronidazole resistance in *Helicobacter pylori*. J Antimicrob Chemother. 43(6):753-758.

Jensen LJ, Julien P, Kuhn M, von Mering C, Muller J, Doerks T, Bork P. 2008. eggNOG: automated construction and annotation of orthologous groups of genes. Nucleic Acids Res. 36(Database issue):D250-D254.

Jensen PJ, Feldman M, LaMont J, Grover S. 2019. Acute and chronic gastritis due to *Helicobacter pylori*. In: Lamont JT, editor. UpToDate. UpToDate, Waltham, MA. (Accessed on December 7, 2019).

Jeong J-Y, Mukhopadhyay AK, Dailidiene D, Wang Y, Velapatiño B, Gilman RH, Parkinson AJ, Nair GB, Wong BCY, Lam SK et al. 2000. Sequential inactivation of *rdxA* (HP0954) and *frxA* (HP0642) nitroreductase genes causes moderate and high-level metronidazole resistance in *Helicobacter pylori*. J Bacteriol. 182(18):5082–5090.

Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN et al. 2017. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res. 45(D1):D566-D573.

Jiang Q, Hiratsuka K, Taylor DE. 1996. Variability of gene order in different *Helicobacter pylori* strains contributes to genome diversity. Mol Microbiol. 20(4):833–842.

Jiang X, Doyle MP. 1998. Effect of environmental and substrate factors on survival and growth of *Helicobacter pylori*. J Food Protect. 61(8):929–933.

Jiang X, Doyle MP. 2000. Growth supplements for *Helicobacter pylori*. J Clin Microbiol. 38(5):1984–1987.

Johnson JM, Church GM. 1999. Alignment and structure prediction of divergent protein families: periplasmic and outer membrane proteins of bacterial efflux pumps. J Mol Biol. 287(3):695-715.

Johnson JS, Spakowicz DJ, Hong B-Y, Petersen LM, Demkowicz P, Chen L, Leopold SR, Hanson BM, Agresta HO, Gerstein M et al. 2019. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat Comm. 10(1):5029.

Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 3:124-124.

Jones DM, Curry A, Fox AJ. 1985. An ultrastructural study of the gastric Campylobacter-like organism '*Campylobacter pyloridis*'. Microbiology. 131(9):2335–2341.

Jorgensen MA, Manos J, Mendz GL, Hazell SL. 1998. The mode of action of metronidazole in *Helicobacter pylori*: futile cycling or reduction? J Antimicrob Chemother. 41(1):67-75.

Kamada T, Haruma K, Komoto K, Mihara M, Chen X, Yoshihara M, Sumii K, Kajiyama G, Tahara K, Kawamura Y. 1999. Effect of smoking and histological gastritis severity on the rate of *H. pylori* eradication with omeprazole, amoxicillin, and clarithromycin. Helicobacter. 4(3):204–210.

Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28(1):27–30.

Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2013. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 42(D1):D199-D205.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2015. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44(D1):D457–D462.

Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 428(4):726-731.

Kangatharalingam N, Amy PS. 1994. *Helicobacter pylori* comb. nov. exhibits facultative acidophilism and obligate microaerophilism. Appl Environ Microbiol. 60(6):2176–2179.

Kanoh S, Rubin BK. 2010. Mechanisms of action and clinical application of macrolides as immunomodulatory medications. Clin Microbiol Rev. 23(3):590–615.

Kapperud G, Nesbakken T, Aleksic S, Mollaret HH. 1990. Comparison of restriction endonuclease analysis and phenotypic typing methods for differentiation of *Yersinia enterocolitica* isolates. J Clin Microbiol. 28(6):1125-1131.

Kargar M, Baghernejad M, Doosti A. 2010. Role of NADPH-insensitive nitroreductase gene to metronidazole resistance of *Helicobacter pylori* strains. DARU. 18(2):137–140.

Kargar M, Baghernejad M, Doosti A, Ghorbani-Dalini S. 2011. Clarithromycin resistance and 23S rRNA mutations in *Helicobacter pylori* isolates in Iran. Afr J Microbiol Res. 5(8):853–856.

Kassim A, Omuse G, Premji Z, Revathi G. 2016. Comparison of Clinical Laboratory Standards Institute and European Committee on Antimicrobial Susceptibility Testing guidelines for the interpretation of antibiotic susceptibility at a University teaching hospital in Nairobi, Kenya: a cross-sectional study. Ann Clin Microbiol Antimicrob. 15:21–21.

Kather B, Stingl K, van der Rest ME, Altendorf K, Molenaar D. 2000. Another unusual type of citric acid cycle enzyme in *Helicobacter pylori*: the malate:quinone oxidoreductase. J Bacteriol. 182(11):3204-3209.

Kennemann L, Didelot X, Aebischer T, Kuhn S, Drescher B, Droege M, Reinhardt R, Correa P, Meyer TF, Josenhans C et al. 2011. *Helicobacter pylori* genome evolution during human infection. Proc Natl Acad Sci USA. 108(12):5033–5038.

Kersulyte D, Bertoli MT, Tamma S, Keelan M, Munday R, Geary J, Veldhuyzen van Zanten S, Goodman KJ, Berg DE. 2015. Complete genome sequences of two *Helicobacter pylori* strains from a Canadian Arctic Aboriginal community. Genome Announc. 3(2):e00209–00215.

Kersulyte D, Lee W, Subramaniam D, Anant S, Herrera P, Cabrera L, Balqui J, Barabas O, Kalia A, Gilman RH et al. 2009. *Helicobacter pylori*'s plasticity zones are novel transposable elements. PLoS One. 4(9):e6859.

Kersulyte D, Velapatiño B, Mukhopadhyay AK, Cahuayme L, Bussalleu A, Combe J, Gilman RH, Berg DE. 2003. Cluster of type IV secretion genes in *Helicobacter pylori*'s plasticity zone. J Bacteriol. 185(13):3764–3772.

Khademi F, Faghri J, Moghim S, Esfahani BN, Fazeli H, Poursina F, Adibi P, Madhi M, Safaei HG. 2014. The study of mutation in 23S rRNA resistance gene of *Helicobacter pylori* to clarithromycin in patients with gastrointestinal disorders in Isfahan - Iran. Adv Biomed Res. 3:98-98.

Khan R, Nahar S, Mukhopadhyay AK, Berg DE, Ahmad MM, Okamoto K, Nair GB, Rahman M. 2008. Isolation of tetracycline-resistant clinical *Helicobacter pylori* without mutations in 16S rRNA gene in Bangladesh. Microbiol Immunol. 52(10):508–511.

Khan R, Nahar S, Sultana J, Ahmad MM, Rahman M. 2004. T2182C Mutation in 23S rRNA is associated with clarithromycin resistance in *Helicobacter pylori* isolates obtained in Bangladesh. Antimicrob Agents Chemother. 48(9):3567–3569.

Khan ZA, Siddiqui MF, Park S. 2019. Current and emerging methods of antibiotic susceptibility testing. Diagnostics. 9(2):49.

Kieboom J, Dennis JJ, de Bont JAM, Zylstra GJ. 1998. Identification and molecular characterization of an efflux pump involved in *Pseudomonas putida* S12 solvent tolerance. J Biol Chem. 273(1):85-91.

Kilianski A, Haas JL, Corriveau EJ, Liem AT, Willis KL, Kadavy DR, Rosenzweig CN, Minot SS. 2015. Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer. GigaScience. 4(1).

Kim BG, Lee DH, Ye BD, Lee KH, Kim BW, Kim SG, Kim SW, Kim SK, Kim JJ, Kim HY et al. 2007. Comparison of 7-day and 14-day proton pump inhibitor-containing triple therapy for *Helicobacter pylori* eradication: neither treatment duration provides acceptable eradication rate in Korea. Helicobacter. 12(1):31–35.

Kim BJ, Kim JG. 2013. Substitutions in penicillin-binding protein 1 in amoxicillin-resistant *Helicobacter pylori* strains isolated from Korean patients. Gut Liver. 7(6):655-660.

Kim D, Song L, Breitwieser FP, Salzberg SL. 2016. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. 26(12):1721-1729.

Kim J, Kim N, Kim Y-J, Kim I, Chee YJ, Lee C, Jung H. 2008a. Gene mutations of 23S rRNA associated with clarithromycin resistance in *Helicobacter pylori* strains isolated from Korean patients. J Microbiol Biotechnol. 18 9:1584-1589.

Kim JM, Kim JS, Kim N, Jung HC, Song IS. 2005. Distribution of fluoroquinolone MICs in *Helicobacter pylori* strains from Korean patients. J Antimicrob Chemother. 56(5):965–967.

Kim JM, Kim JS, Kim N, Kim Y-J, Kim IY, Chee YJ, Lee C-H, Jung HC. 2008b. Gene mutations of 23S rRNA associated with clarithromycin resistance in *Helicobacter pylori* strains isolated from Korean patients. J Microbiol Biotechnol. 18(9):1584–1589.

Kim KS, Kang JO, Eun CS, Han DS, Choi TY. 2002. Mutations in the 23S rRNA gene of *Helicobacter pylori* associated with clarithromycin resistance. J Korean Med Sci. 17(5):599-603.

Kim SJ, Kim Y-J, Ko KS. 2018. Genomic analysis of consecutive *Acinetobacter baumannii* strains from a single patient. Front Microbiol. 9(2840).

King A. 2001. Recommendations for susceptibility tests on fastidious organisms and those requiring special handling. J Antimicrob Chemother. 48(suppl 1):77–80.

Kirst HA, Sides GD. 1989. New directions for macrolide antibiotics: structural modifications and in vitro activity. Antimicrob Agents Chemother. 33(9):1413–1418.

Kivi M, Tindberg Y, Sörberg M, Casswall TH, Befrits R, Hellström PM, Bengtsson C, Engstrand L, Granström M. 2003. Concordance of *Helicobacter pylori* strains within families. J Clin Microbiol. 41(12):5604–5608.

Klein PD, Opekun AR, Smith EO, Klein PD, Graham DY, Graham DY, Gaillour A. 1991. Water source as risk factor for *Helicobacter pylori* infection in Peruvian children. Lancet. 337(8756):1503-1506.

Kocazeybek B, Sakli MK, Yuksel P, Demirci M, Caliskan R, Sarp TZ, Saribas S, Demiryas S, Kalayci F, Cakan H et al. 2019. Comparison of new and classical point mutations associated with clarithromycin resistance in *Helicobacter pylori* strains isolated from dyspeptic patients and their effects on phenotypic clarithromycin resistance. J Med Microbiol. 68(4):566-573.

Komorowska M, Szafran H, Popiela T, Szafran Z. 1981. Free amino acids of human gastric juice. Acta Physiol Pol. 32(5):559-567.

Kostrzynska M, Betts JD, Austin JW, Trust TJ. 1991. Identification, characterization, and spatial localization of two flagellin species in *Helicobacter pylori* flagella. J Bacteriol. 173(3):937–946.

Krzyżek P, Biernat MM, Gościniak G. 2019. Intensive formation of coccoid forms as a feature strongly associated with highly pathogenic *Helicobacter pylori* strains. Folia Microbiol. 64(3):273–281.

Kuipers EJ, Peña AS, Festen HPM, Meuwissen SGM, Uyterlinde AM, Roosendaal R, Pals G, Nelis GF. 1995. Long-term sequelae of *Helicobacter pylori* gastritis. Lancet. 345(8964):1525–1528.

Kulathila R, Kulathila R, Indic M, van den Berg B. 2011. Crystal structure of *Escherichia coli* CusC, the outer membrane component of a heavy metal efflux pump. PLoS One. 6(1):e15610.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 35(6):1547–1549.

Kumar V, Sun P, Vamathevan J, Li Y, Ingraham K, Palmer L, Huang J, Brown JR. 2011. Comparative genomics of *Klebsiella pneumoniae* strains with different antibiotic resistance profiles. Antimicrob Agents Chemother. 55(9):4267–4276.

Kunin CM. 1996. Antimicrobial activity of rifabutin. Clin Infect Dis. 22(Supplement_1):S3-S14.

Kuo C-H, Lu C-Y, Shih H-Y, Liu C-J, Wu M-C, Hu H-M, Hsu W-H, Yu F-J, Wu D-C, Kuo F-C. 2014. CYP2C19 polymorphism influences *Helicobacter pylori* eradication. World J Gastroenterol. 20(43):16029–16036.

Kusters JG, Gerrits MM, Van Strijp JA, Vandenbroucke-Grauls CM. 1997. Coccoid forms of *Helicobacter pylori* are the morphologic manifestation of cell death. Infect Immun. 65(9):3672–3679.

Kusters JG, van Vliet AHM, Kuipers EJ. 2006. Pathogenesis of *Helicobacter pylori* infection. Clin Microbiol Rev. 19(3):449.

Kwon DH, Dore MP, Kim JJ, Kato M, Lee M, Wu JY, Graham DY. 2003. High-level β-lactam resistance associated with acquired multidrug resistance in *Helicobacter pylori*. Antimicrob Agents Chemother. 47(7):2169–2178.

Kwon DH, El-Zaatari FA, Kato M, Osato MS, Reddy R, Yamaoka Y, Graham DY. 2000a. Analysis of *rdxA* and involvement of additional genes encoding NAD(P)H flavin oxidoreductase (FrxA) and ferredoxin-like protein (FdxB) in metronidazole resistance of *Helicobacter pylori*. Antimicrob Agents Chemother. 44(8):2133–2142.

Kwon DH, Hulten K, Kato M, Kim JJ, Lee M, El-Zaatari FA, Osato MS, Graham DY. 2001. DNA sequence analysis of *rdxA* and *frxA* from 12 pairs of metronidazole-sensitive and -resistant clinical *Helicobacter pylori* isolates. Antimicrob Agents Chemother. 45(9):2609-2615.

Kwon DH, Kim JJ, Lee M, Yamaoka Y, Kato M, Osato MS, El-Zaatari FAK, Graham DY. 2000b. Isolation and characterization of tetracycline-resistant clinical isolates of *Helicobacter pylori*. Antimicrob Agents Chemother. 44(11):3203–3205.

Kwon DH, Peña JA, Osato MS, Fox JG, Graham DY, Versalovic J. 2000c. Frameshift mutations in *rdxA* and metronidazole resistance in North American *Helicobacter pylori* isolates. J Antimicrob Chemother. 46(5):793-796.

Kwon YH, Jeon SW, Nam SY, Lee HS, Park JH. 2019. Efficacy of tailored therapy for *Helicobacter pylori* eradication based on clarithromycin resistance and survey of previous antibiotic exposure: a single-center prospective pilot study. Helicobacter. 24(4):e12585.

La Teana A, Gualerzi CO, Dahlberg AE. 2001. Initiation factor IF 2 binds to the alpha-sarcin loop and helix 89 of *Escherichia coli* 23S ribosomal RNA. RNA. 7(8):1173-1179.

Labenz J, Blum AL, Bolten WW, Dragosics B, Rösch W, Stolte M, Koelz HR. 2002. Primary prevention of diclofenac associated ulcers and dyspepsia by omeprazole or triple therapy in *Helicobacter pylori* positive patients: a randomised, double blind, placebo controlled, clinical trial. Gut. 51(3):329-335.

Lambert JR, Midolo P. 1997. The actions of bismuth in the treatment of *Helicobacter pylori* infection. Aliment Pharmacol Therapeut. 11(S1):27–33.

Lamichhane B, Chua E-G, Wise MJ, Laming C, Marshall BJ, Tay C-Y. 2019. The complete genome and methylome of *Helicobacter pylori* hpNEAfrica strain HP14039. Gut Pathog. 11(1):7.

Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc Natl Acad Sci USA. 82(20):6955-6959.

Lara-Ramírez EE, Segura-Cabrera A, Guo X, Yu G, García-Pérez CA, Rodríguez-Pérez MA. 2011. New implications on genomic adaptation derived from the *Helicobacter pylori* genome comparison. PLoS One. 6(2):e17300-e17300.

Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM et al. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol. 50(4):1355-1361.

Lauener FN, Imkamp F, Lehours P, Buissonnière A, Benejat L, Zbinden R, Keller PM, Wagner K. 2019. Genetic determinants and prediction of antibiotic resistance phenotypes in *Helicobacter pylori*. J Clin Med. 8(1):53.

Laxminarayan R, Matsoso P, Pant S, Brower C, Røttingen J-A, Klugman K, Davies S. 2016. Access to effective antimicrobials: a worldwide challenge. Lancet. 387(10014):168–175.

Lee HR, Han KS, Yoo BC, Park SM, Cha YJ. 1993. Prevalence of *Helicobacter pylori* infection in patients with peptic ulcer diseases and non-ulcer dyspepsia. Korean J Intern Med. 8(2):73–77.

Lee S-B, Yang J-W, Kim C-S. 2008. The association between conjunctival MALT lymphoma and *Helicobacter pylori*. Br J Ophthalmol. 92(4):534–536.

Lee SM, Kim N, Kwon YH, Nam RH, Kim JM, Park JY, Lee YS, Lee DH. 2018. *rdxA*, *frxA*, and efflux pump in metronidazole-resistant *Helicobacter pylori*: their relation to clinical outcomes. J Gastroenterol Hepatol. 33(3):681–688.

Lee WC, Goh KL, Loke MF, Vadivelu J. 2017. Elucidation of the metabolic network of Helicobacter pylori J99 and Malaysian clinical strains by phenotype microarray. Helicobacter. 22(1):e12321.

Lehours P, Vale FF, Bjursell MK, Melefors O, Advani R, Glavas S, Guegueniat J, Gontier E, Lacomme S, Alves Matos A et al. 2011. Genome sequencing reveals a phage in *Helicobacter pylori*. mBio. 2(6):e00239–00211.

Leipzig J. 2016. A review of bioinformatic pipeline frameworks. Brief Bioinform. 18(3):530–536.

Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44(W1):W242-W245.

Leung W-K, Siu KLK, Kwok CKL, Chan S-Y, Sung R, Sung JJY. 1999. Isolation of *Helicobacter pylori* from vomitus in children and its implication in gastro-oral transmission. Am J Gastroenterol Suppl. 94(10):2881–2884.

Levine C, Hiasa H, Marians KJ. 1998. DNA gyrase and topoisomerase IV: biochemical activities, physiological roles during chromosome replication, and drug sensitivities. BBA - Gen Subjects. 1400(1):29–43.

Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 31(10):1674-1676.

Li L, Stoeckert CJ, Jr., Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13(9):2178-2189.

Li XZ, Nikaido H, Poole K. 1995. Role of MxA-MexB-OprM in antibiotic efflux in *Pseudomonas aeruginosa*. Antimicrob Agents Chemother. 39(9):1948-1953.

Li Y, Dannelly HK. 2006. Inactivation of the putative tetracycline resistance gene HP1165 in *Helicobacter pylori* led to loss of inducible tetracycline resistance. Arch Microbiol. 185(4):255–262.

Liao YL, Guo G, Mao XH, Xie QH, Zhang WJ, Liu XF, Zou QM. 2009. Core genome haplotype diversity and *vacA* allelic heterogeneity of Chinese *Helicobacter pylori* strains. Curr Microbiol. 59(2):123-129.

Lim JH, Lee DH, Choi C, Lee ST, Kim N, Jeong SH, Kim JW, Hwang JH, Park YS, Lee SH et al. 2013. Clinical outcomes of two-week sequential and concomitant therapies for *Helicobacter pylori* eradication: a randomized pilot study. Helicobacter. 18(3):180–186.

Lin J, Michel LO, Zhang Q. 2002. CmeABC functions as a multidrug efflux system in *Campylobacter jejuni*. Antimicrob Agents Chemother. 46(7):2124-2131.

Lin J, Sahin O, Michel LO, Zhang Q. 2003. Critical role of multidrug efflux pump CmeABC in bile resistance and in vivo colonization of *Campylobacter jejuni*. Infect Immun. 71(8):4250-4259.

Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P, Falush D, Stamer C, Prugnolle F, van der Merwe SW et al. 2007. An African origin for the intimate association between humans and *Helicobacter pylori*. Nature. 445(7130):915–918.

Linz B, Windsor HM, Gajewski JP, Hake CM, Drautz DI, Schuster SC, Marshall BJ. 2013. *Helicobacter pylori* genomic microevolution during naturally occurring transmission between adults. PLoS One. 8(12):e82187–e82187.

Linz B, Windsor HM, McGraw JJ, Hansen LM, Gajewski JP, Tomsho LP, Hake CM, Solnick JV, Schuster SC, Marshall BJ. 2014. A mutation burst during the acute phase of *Helicobacter pylori* infection in humans and rhesus macaques. Nat Comm. 5:4165.

Liu B, Zheng D, Jin Q, Chen L, Yang J. 2019. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res. 47(D1):D687-D692.

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. 2012. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012:251364–251364.

Liu Y, Yue H, Li A, Wang J, Jiang B, Zhang Y, Bai Y. 2009. An epidemiologic study on the correlation between oral *Helicobacter pylori* and gastric *H. pylori*. Curr Microbiol. 58(5):449.

Liu Z-Q, Zheng P-Y, Yang P-C. 2008. Efflux pump gene *hefA* of *Helicobacter pylori* plays an important role in multidrug resistance. World J Gastroenterol. 14(33):5217–5222.

Lõivukene K, Maaroos HI, Kolk H, Kull I, Labotkin K, Mikelsaar M. 2002. Prevalence of antibiotic resistance of *Helicobacter pylori* isolates in Estonia during 1995–2000 in comparison to the consumption of antibiotics used in treatment regimens. Clin Microbiol Infect. 8(9):598–603.

Löscher T, Alberer M. 2013. Clinical presentation and management of travelers' diarrhea. In: Keystone JS, Freedman DO, Kozarsky PE, Connor BA, Nothdurft HD, editors. Travel Medicine (Third Edition). London: Saunders. p. 197–206.

Lu H, Hsu P-I, Graham DY, Yamaoka Y. 2005. Duodenal ulcer promoting gene of *Helicobacter pylori*. Gastroenterology. 128(4):833–848.

Lu Y, Shen Y, Warren W, Walter R. 2016. Next generation sequencing in aquatic models. Next Generation Sequencing-Advances, Applications and Challenges. p. 61–79.

Lucier TS, Heitzman K, Liu SK, Hu PC. 1995. Transition mutations in the 23S rRNA of erythromycin-resistant isolates of *Mycoplasma pneumoniae*. Antimicrob Agents Chemother. 39(12):2770-2773.

Lundgren A, Suri-Payer E, Enarsson K, Svennerholm A-M, Lundin BS. 2003. *Helicobacter pylori* specific CD4+ CD25high regulatory T cells suppress memory T-cell responses to *H. pylori* in infected individuals. Infect Immun. 71(4):1755–1762.

Ma D, Cook DN, Alberti M, Pon NG, Nikaido H, Hearst JE. 1995. Genes *acrA* and *acrB* encode a stress-induced efflux system of *Escherichia coli*. Mol Microbiol. 16(1):45-55.

Mackay WG, Gribbon LT, Barer MR, Reid DC. 1998. Biofilms in drinking water systems: a possible reservoir for *Helicobacter pylori*. J Appl Microbiol. 85(S1):52S–59S.

Magalhães A, Reis CA. 2010. *Helicobacter pylori* adhesion to gastric epithelial cells is mediated by glycan receptors. Braz J Med Biol Res. 43:611–618.

Magnet S, Courvalin P, Lambert T. 2001. Resistance-nodulation-cell division-type efflux pump involved in aminoglycoside resistance in *Acinetobacter baumannii* strain BM4454. Antimicrob Agents Chemother. 45(12):3375-3380.

Mahdavi J, Sondén B, Hurtig M, Olfat FO, Forsberg L, Roche N, Angstrom J, Larsson T, Teneberg S, Karlsson K-A et al. 2002. *Helicobacter pylori* SabA adhesin in persistent infection and chronic inflammation. Science. 297(5581):573–578.

Mahmoud M, Zywicki M, Twardowski T, Karlowski WM. 2019. Efficiency of PacBio long read correction by 2nd generation Illumina sequencing. Genomics. 111(1):43–49.

Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA et al. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. PNAS. 95(6):3140-3145.

Malaty HM, Graham DY. 1994. Importance of childhood socioeconomic status on the current prevalence of *Helicobacter pylori* infection. Gut. 35(6):742–745.

Malfertheiner P, Megraud F, O'Morain CA, Gisbert JP, Kuipers EJ, Axon AT, Bazzoli F, Gasbarrini A, Atherton J, Graham DY et al. 2017. Management of *Helicobacter pylori* infection—the Maastricht V/Florence consensus report. Gut. 66(1):6–30.

Mamori S, Higashida A, Kawara F, Ohnishi K, Takeda A, Senda E, Ashida C, Yamada H. 2010. Age-dependent eradication of *Helicobacter pylori* in Japanese patients. World J Gastroenterol. 16(33):4176–4179.

Marais A, Bilardi C, Cantet F, Mendz GL, Mégraud F. 2003. Characterization of the genes *rdxA* and *frxA* involved in metronidazole resistance in *Helicobacter pylori*. Microbiol Res. 154(2):137–144.

Marcus EA, Inatomi N, Nagami GT, Sachs G, Scott DR. 2012. The effects of varying acidity on *Helicobacter pylori* growth and the bactericidal efficacy of ampicillin. Aliment Pharmacol Therapeut. 36(10):972–979.

Marcus EA, Sachs G, Scott DR. 2015. Colloidal bismuth subcitrate impedes proton entry into *Helicobacter pylori* and increases the efficacy of growth-dependent antibiotics. Aliment Pharmacol Therapeut. 42(7):922–933.

Mardis ER. 2017. DNA sequencing technologies: 2006–2016. Nat Protocol. 12(2):213–218.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 437(7057):376–380.

Marians KJ, Hiasa H. 1997. Mechanism of quinolone action: a drug-induced structural perturbation of the DNA precedes strand cleavage by topoisomerase IV. J Biol Chem. 272(14):9401–9409.

Marshall B. 2008. *Helicobacter pylori*: a Nobel pursuit? Can J Gastroenterol. 22(11):895–896.

Marshall B, Goodwin CS. 1987. Revised nomenclature of *Campylobacter pyloridis*. Int J Syst Evol Microbiol. 37(1):68–68.

Marshall B, Warren JR. 1984. Unidentified curves bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet. 323(8390):1311–1315.

Marshall BJ, Barrett LJ, Prakash C, McCallum RW, Guerrant RL. 1990. Urea protects *Helicobacter* (*Campylobacter*) *pylori* from the bactericidal effect of acid. Gastroenterology. 99(3):697-702.

Martínez-Júlvez M, Rojas AL, Olekhnovich I, Espinosa Angarica V, Hoffman PS, Sancho J. 2012. Structure of RdxA--an oxygen-insensitive nitroreductase essential for metronidazole activation in *Helicobacter pylori*. The FEBS journal. 279(23):4306-4317.

Martiny AC, Treseder K, Pusch G. 2013. Phylogenetic conservatism of functional traits in microorganisms. ISME J. 7(4):830-838.

Mascellino MT, Porowska B, De Angelis M, Oliva A. 2017. Antibiotic susceptibility, heteroresistance, and updated treatment strategies in *Helicobacter pylori* infection. Drug Des Devel Ther. 11:2209–2220.

Matsuoka M, Endou K, Kobayashi H, Inoue M, Nakajima Y. 1998. A plasmid that encodes three genes for resistance to macrolide antibiotics in *Staphylococcus aureus*. FEMS Microbiol Lett. 167(2):221-227.

Matteo MJ, Granados G, Olmos M, Wonaga A, Catalano M. 2008. *Helicobacter pylori* amoxicillin heteroresistance due to point mutations in PBP-1A in isogenic isolates. J Antimicrob Chemother. 61(3):474-477.

Mattsson A, Quiding-Järbrink M, Lönroth H, Hamlet A, Ahlstedt I, Svennerholm A-M. 1998a. Antibody-secreting cells in the stomachs of symptomatic and asymptomatic Helicobacter pylori - infected subjects. Infection and Immunity. 66(6):2705-2712.

Mattsson A, Quiding-Järbrink M, Lönroth H, Hamlet A, Ahlstedt I, Svennerholm A. 1998b. Antibody-secreting cells in the stomachs of symptomatic and asymptomatic *Helicobacter pylori*-infected subjects. Infect Immun. 66(6):2705–2712.

Mattsson A, Tinnert A, Hamlet A, Lönroth H, Bölin I, Svennerholm AM. 1998c. Specific antibodies in sera and gastric aspirates of symptomatic and asymptomatic *Helicobacter pylori*-infected subjects. Clin Diagn Lab Immunol. 5(3):288–293.

Matuschek E, Brown DFJ, Kahlmeter G. 2014. Development of the EUCAST disk diffusion antimicrobial susceptibility testing method and its implementation in routine microbiology laboratories. Clin Microbiol Infect. 20(4):O255–O266.

Mayhall CG, Apollo E. 1980. Effect of storage and changes in bacterial growth phase and antibiotic concentrations on antimicrobial tolerance in *Staphylococcus aureus*. Antimicrob Agents Chemother. 18(5):784–788.

McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L et al. 2013. The comprehensive antibiotic resistance database. Antimicrob Agents Chemother. 57(7):3348-3357.

McClain MS, Beckett AC, Cover TL. 2017. *Helicobacter pylori* vacuolating toxin and gastric cancer. Toxins. 9(10):316.

McColl KEL. 2010. *Helicobacter pylori* infection. New Engl J Med. 362(17):1597–1604.

McDermott PF, Tyson GH, Kabera C, Chen Y, Li C, Folster JP, Ayers SL, Lam C, Tate HP, Zhao S. 2016. Whole-genome sequencing for detecting antimicrobial resistance in nontyphoidal *Salmonella*. Antimicrob Agents Chemother. 60(9):5515–5520.

McMahon BJ, Hennessy TW, Bensler JM, Bruden DL, Parkinson AJ, Morris JM, Reasonover AL, Hurlburt DA, Bruce MG, Sacco F et al. 2003. The relationship among previous antimicrobial use, antimicrobial resistance, and treatment outcomes for *Helicobacter pylori* infections. Ann Intern Med. 139(6):463-469.

McNulty C, and the PHLS Helicobacter Working Group:, Owen R, Tompkins D, Hawtin P, McColl K, Price A, Smith G, Teare L. 2002. *Helicobacter pylori* susceptibility testing by disc diffusion. J Antimicrob Chemother. 49(4):601–609.

Mégraud F. 1997. Resistance of Helicobacter pylori to antibiotics. Aliment Pharmacol Therapeut. 11(S1):43-53.

Mégraud F. 2004. *H pylori* antibiotic resistance: prevalence, importance, and advances in testing. Gut. 53(9):1374-1384.

Mégraud F, Coenen S, Versporten A, Kist M, Lopez-Brea M, Hirschl AM, Andersen LP, Goossens H, Glupczynski Y. 2013. *Helicobacter pylori* resistance to antibiotics in Europe and its relationship to antibiotic consumption. Gut. 62(1):34–42.

Mégraud F, Lehours P. 2007. *Helicobacter pylori* detection and antimicrobial susceptibility testing. Clin Microbiol Rev. 20(2):280–322.

Mehrabadi JF, Sirous M, Daryani NE, Eshraghi S, Akbari B, Shirazi MH. 2011. Assessing the role of the RND efflux pump in metronidazole resistance of *Helicobacter pylori* by RT-PCR assay. J Infect Dev Ctries. 5(02):88–93.

Mendoza-Elizalde S, Cortés-Márquez AC, Zuñiga G, Cerritos R, Valencia-Mayoral P, Sánchez AC, Olivares-Clavijo H, Velázquez-Guadarrama N. 2019. Inference from the analysis of

genetic structure of *Helicobacter pylori* strains isolates from two paediatric patients with recurrent infection. BMC Microbiol. 19(1):184.

Mendz GL, Hazell SL. 1991. Evidence for a pentose phosphate pathway in *Helicobacter pylori*. FEMS Microbiol Lett. 84(3):331-336.

Mhaskar RS, Ricardo I, Azliyati A, Laxminarayan R, Amol B, Santosh W, Boo K. 2013. Assessment of risk factors of *Helicobacter pylori* infection and peptic ulcer disease. J Glob Infect Dis. 5(2):60-67.

Miehlke S, Kirsch C, Agha-Amiri K, Günther T, Lehn N, Malfertheiner P, Stolte M, Ehninger G, Bayerdörffer E. 2000. The *Helicobacter pylori vacA* s1, m1 genotype and *cagA* is associated with gastric carcinoma in Germany. Int J Canc. 87(3):322–327.

Miftahussurur M, Fauzia KA, Nusi IA, Setiawan PB, Syam AF, Waskito LA, Doohan D, Ratnasari N, Khomsan A, Adnyana IK et al. 2020. E-test versus agar dilution for antibiotic susceptibility testing of *Helicobacter pylori*: a comparison study. BMC Res Notes. 13(1):22.

Miftahussurur M, Syam AF, Nusi IA, Makmun D, Waskito LA, Zein LH, Akil F, Uwan WB, Simanjuntak D, Wibawa IDN et al. 2016. Surveillance of *Helicobacter pylori* antibiotic susceptibility in Indonesia: different resistance types among regions and with novel genetic mutations. PLoS One. 11(12):e0166199.

Mignard S, Flandrois JP. 2006. 16S rRNA sequencing in routine bacterial identification: a 30-month experiment. J Microbiol Meth. 67(3):574–581.

Milani C, Hevia A, Foroni E, Duranti S, Turroni F, Lugli GA, Sanchez B, Martín R, Gueimonde M, van Sinderen D et al. 2013. Assessing the fecal microbiota: an optimized ion torrent 16S rRNA gene-based analysis protocol. PLoS One. 8(7):e68739–e68739.

Miller EF, Maier RJ. 2014. Ammonium metabolism enzymes aid *Helicobacter pylori* acid resistance. J Bacteriol. 196(17):3074-3081.

Minas K, McEwan NR, Newbold CJ, Scott KP. 2011. Optimization of a high-throughput CTAB-based protocol for the extraction of qPCR-grade DNA from rumen fluid, plant and bacterial pure cultures. FEMS Microbiol Lett. 325(2):162–169.

Mirzaei N, Poursina F, Moghim S, Rahimi E, Safaei HG. 2014. The mutation of the *rdxA* gene in metronidazole-resistant *Helicobacter pylori* clinical isolates. Adv Biomed Res. 3:90-90.

Mishra KK, Srivastava S, Garg A, Ayyagari A. 2006. Antibiotic susceptibility of *Helicobacter pylori* clinical isolates: comparative evaluation of disk-diffusion and Etest methods. Curr Microbiol. 53(4):329.

Misiewicz JJ, Harris AW, Bardhan KD, Levi S, O'Morain C, Cooper BT, Kerr GD, Dixon MF, Langworthy H, Piper D. 1997. One week triple therapy for *Helicobacter pylori*: a multicentre comparative study. Gut. 41(6):735-739.

Mitchell HM, Li YY, Hu PJ, Liu Q, Chen M, Du GG, Wang ZJ, Lee A, Hazell SL. 1992. Epidemiology of *Helicobacter pylori* in Southern China: identification of early childhood as the critical period for acquisition. Int J Infect Dis. 166(1):149–153.

Mitui M, Patel A, Leos NK, Doern CD, Park JY. 2014. Novel *Helicobacter pylori* sequencing test identifies high rate of clarithromycin resistance. J Pediatr Gastroenterol Nutr. 59(1):6-9.

Miwa H, Nagahara A, Kurosawa A, Ohkusa T, Ohkura R, Hojo M, Enomoto N, Sato N. 2003. Is antimicrobial susceptibility testing necessary before second-line treatment for *Helicobacter pylori* infection? Aliment Pharmacol Therapeut. 17(12):1545–1551.

Momtaz H, Dabiri H, Souod N, Gholami M. 2014. Study of *Helicobacter pylori* genotype status in cows, sheep, goats and human beings. BMC Gastroenterol. 14(1):61.

Montano V, Didelot X, Foll M, Linz B, Reinhardt R, Suerbaum S, Moodley Y, Jensen JD. 2015. Worldwide population structure, long-term demography, and local adaptation of *Helicobacter pylori*. Genetics. 200(3):947–963.

Moodley Y, Linz B. 2009. *Helicobacter pylori* sequences reflect past human migrations. Microbial Pathogenomics. Karger Publishers. p. 62–74.

Moodley Y, Linz B, Bond RP, Nieuwoudt M, Soodyall H, Schlebusch CM, Bernhöft S, Hale J, Suerbaum S, Mugisha L et al. 2012. Age of the association between *Helicobacter pylori* and man. PLoS Pathog. 8(5):e1002693–e1002693.

Moodley Y, Linz B, Yamaoka Y, Windsor HM, Breurec S, Wu J-Y, Maady A, Bernhöft S, Thiberge J-M, Phuanukoonnon S et al. 2009. The peopling of the Pacific from a bacterial perspective. Science. 323(5913):527-530.

Moore JM, Salama NR. 2005. Mutational analysis of metronidazole resistance in *Helicobacter pylori*. Antimicrob Agents Chemother. 49(3):1236–1237.

Morgan DR, Crowe SE. 2016. *Helicobacter pylori* infection. In: Feldman MMD, Friedman LSMD, Brandt LJMD, editors. Sleisenger and Fordtran's Gastrointestinal and Liver Disease. p. 856–867.e855.

Morozova O, Marra MA. 2008. Applications of next-generation sequencing technologies in functional genomics. Genomics. 92(5):255–264.

Morris A, Nicholson G. 1987. Ingestion of *Campylobacter pyloridis* causes gastritis and raised fasting gastric pH. Am J Gastroenterol Suppl. 82(3).

Mostafa HH, Taffner SM, Wang J, Malek A, Hardy DJ, Pecora ND. 2019. Genome sequence of a *Facklamia hominis* isolate from a patient with urosepsis. Microbiol Resour Announc.8(17)doi:10.1128/MRA.00100-19.

Mungazi SG, Chihaka OB, Muguti GI. 2018. Prevalence of *Helicobacter pylori* in asymptomatic patients at surgical outpatient department: Harare hospitals. Ann Med Surg (Lond). 35:153-157.

Muñoz-Ramírez ZY, Mendez-Tenorio A, Kato I, Bravo MM, Rizzato C, Thorell K, Torres R, Aviles-Jimenez F, Camorlinga M, Canzian F et al. 2017. Whole genome sequence and phylogenetic analysis show *Helicobacter pylori* strains from Latin America have followed a unique evolution pathway. Front Cell Infect Mi. 7(50).

Murray BO, Dawson RA, Alsharaf LM, Anne Winter J. 2020. Protective effects of *Helicobacter pylori* membrane vesicles against stress and antimicrobial agents. Microbiology.

Mutters NT, Meyle E. 2013. Amoxicillin: pharmaceutical properties and clinical applications. In: Leary WA, Kruse LM, editors. Amoxicillin. Hauppauge, United States: Nova Science Publishers, Incorporated. p. 83–100.

Nash KA, Inderlied CB. 1995. Genetic basis of macrolide resistance in *Mycobacterium avium* isolated from patients with disseminated disease. Antimicrob Agents Chemother. 39(12):2625-2630.

Naylor NR, Pouwels KB, Hope R, Green N, Henderson KL, Knight GM, Atun R, Robotham JV, Deeny SR. 2019. The health and cost burden of antibiotic resistant and susceptible *Escherichia coli* bacteraemia in the English hospital setting: a national retrospective cohort study. PLoS One. 14(9):e0221944.

Nell S, Eibach D, Montano V, Maady A, Nkwescheu A, Siri J, Elamin WF, Falush D, Linz B, Achtman M et al. 2013. Recent acquisition of *Helicobacter pylori* by Baka pygmies. PLoS Genet. 9(9):e1003775-e1003775.

Neu HC. 1982. Penicillins: new insights into their mechanisms of activity and clinical use. Bull N Y Acad Med. 58(8):681–695.

New Zealand Formulary. Amoxicillin. 2020a. [accessed 2020 11 February]. https://nzf.org.nz/nzf_3025.

New Zealand Formulary. Recommended regimens for *Helicobacter pylori* eradication in adults. 2020b. [accessed 2020 10 January]. https://nzf.org.nz/nzf_732.

New Zealand Guidelines Group. Management of dyspepsia and heartburn. 2004. Ministry of Health; [accessed 2020 11 February]. https://www.health.govt.nz/system/files/documents/publications/dyspepsia_guideline_web201.pdf.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32(1):268-274.

Nguyen LT, Uchida T, Tsukamoto Y, Trinh TD, Ta L, Mai HB, Le HS, Ho DQD, Hoang HH, Matsuhisa T et al. 2010. Clinical relevance of cagPAI intactness in *Helicobacter pylori* isolates from Vietnam. Eur J Clin Microbiol Infect Dis. 29(6):651-660.

Nightingale SD, Cameron DW, Gordin FM, Sullam PM, Cohn DL, Chaisson RE, Eron LJ, Sparti PD, Bihari B, Kaufman DL et al. 1993. Two controlled trials of rifabutin prophylaxis against *Mycobacterium avium* complex infection in AIDS. New Engl J Med. 329(12):828-833.

Noda S, Aihara C, Yuki M, Ohkuma M. 2018. Draft genome sequence of *Lactococcus* sp. strain NtB2 (JCM 32569), isolated from the gut of the higher termite *Nasutitermes takasagoensis*. Genome Announc. 6(24):e00445–00418.

Nomura A, Stemmermann GN, Chyou P-H, Perez-Perez GI, Blaser MJ. 1994. *Helicobacter pylori* infection and the risk for duodenal and gastric ulceration. Ann Intern Med. 120(12):977–981.

Noureen M, Tada I, Kawashima T, Arita M. 2019. Rearrangement analysis of multiple bacterial genomes. BMC Bioinformatics. 20(Suppl 23):631-631.

Nukui Y, Chino T, Tani C, Sonobe K, Aiso Y, Tohda S, Koike R, Saito R. 2020. Molecular epidemiologic and clinical analysis of *Helicobacter cinaedi* bacteremia in Japan. Helicobacter. 25(1):e12675.

Nurgalieva ZZ, Malaty HM, Graham DY, Almuchambetova R, Machmudova A, Kapsultanova D, Osato MS, Hollinger FB, Zhangabylov A. 2002. *Helicobacter pylori* infection in Kazakhstan: effect of water source and household hygiene. Am J Trop Med Hyg. 67(2):201-206.

O'Connor JPA, Taneike I, O'Morain C. 2009. Improving compliance with *Helicobacter pylori* eradication therapy: when and how? Therap Adv Gastroenterol. 2(5):273–279.

Odenbreit S, Püls J, Sedlmaier B, Gerland E, Fischer W, Haas R. 2000. Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. Science. 287(5457):1497–1500.

Odenbreit S, Swoboda K, Barwig I, Ruhl S, Borén T, Koletzko S, Haas R. 2009. Outer membrane protein expression profile in *Helicobacter pylori* clinical isolates. Infect Immun. 77(9):3782-3790.

Ogata M, Araki K, Ogata T. 1998. An electron microscopic study of *Helicobacter pylori* in the surface mucous gel layer. Histol Histopathol. 13(2):347–358.

Ogata SK, Gales AC, Kawakami E. 2015. Antimicrobial susceptibility testing for *Helicobacter pylori* isolates from Brazilian children and adolescents: comparing agar dilution, E-test, and disk diffusion. Braz J Microbiol. 45(4):1439–1448.

Oh DH, Lee DH, Kang KK, Park YS, Shin CM, Kim N, Yoon H, Hwang JH, Jeoung SH, Kim JW et al. 2014. The efficacy of hybrid therapy as first-line regimen for *Helicobacter pylori* infection compared with sequential therapy. J Gastroenterol Hepatol. 29(6):1171–1176.

Ohkusa T, Miwa H, Endo SHU, Okayasu I, Sato N. 2004. *Helicobacter pylori* is a fragile bacteria when stored at low and ultra-low temperatures. J Gastroenterol Hepatol. 19(2):200-204.

Ojetti V, Migneco A, Zocco MA, Nista EC, Gasbarrini G, Gasbarrini A. 2004. Beta-lactamase inhibitor enhances *Helicobacter pylori* eradication rate. J Intern Med. 255(1):125–129.

Okamoto T, Yoshiyama H, Nakazawa T, Park I-D, Chang M-W, Yanai H, Okita K, Shirai M. 2002. A change in PBP1 is involved in amoxicillin resistance of clinical isolates of *Helicobacter pylori*. J Antimicrob Chemother. 50(6):849–856.

Okusu H, Ma D, Nikaido H. 1996. AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of *Escherichia coli* multiple-antibiotic-resistance (Mar) mutants. J Bacteriol. 178(1):306-308.

Olekhnovich IN, Goodwin A, Hoffman PS. 2009. Characterization of the NAD(P)H oxidase and metronidazole reductase activities of the RdxA nitroreductase of *Helicobacter pylori*. The FEBS journal. 276(12):3354-3364.

Olivier BJ, Bond RP, van Zyl WB, Delport M, Slavik T, Ziady C, Terhaar Sive Droste JS, Lastovica A, van der Merwe SW. 2006. Absence of *Helicobacter pylori* within the oral cavities of members of a healthy South African community. J Clin Microbiol. 44(2):635-636.

Olivieri R, Bugnoli M, Armellini D, Bianciardi S, Rappuoli R, Bayeli PF, Abate L, Esposito E, de Gregorio L, Aziz J et al. 1993. Growth of *Helicobacter pylori* in media containing cyclodextrins. J Clin Microbiol. 31(1):160–162.

Osato MS, Ayub K, Le H-H, Reddy R, Graham DY. 1998. Houseflies are an unlikely reservoir or vector for *Helicobacter pylori*. J Clin Microbiol. 36(9):2786–2788.

Osato MS, Reddy R, Reddy SG, Penland RL, Graham DY. 2001. Comparison of the Etest and the NCCLS-approved agar dilution method to detect metronidazole and clarithromycin resistant *Helicobacter pylori*. Int J Antimicrob Agents. 17(1):39–44.

Ottemann KM, Lowenthal AC. 2002. *Helicobacter pylori* uses motility for initial colonization and to attain robust infection. Infect Immun. 70(4):1984–1990.

Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M et al. 2013. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 42(D1):D206–D214.

Oxford Nanopore. Nanopore overview. 2020. [accessed 2020 3 March]. https://nanoporetech.com/products/comparison.

Ozaydin N, Turkyilmaz SA, Cali S. 2013. Prevalence and risk factors of *Helicobacter pylori* in Turkey: a nationally-representative, cross-sectional, screening with the 13C-Urea breath test. BMC Publ Health. 13(1):1215.

Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 31(22):3691-3693.

Panayotopoulou EG, Sgouras DN, Papadakos KS, Petraki K, Breurec S, Michopoulos S, Mantzaris G, Papatheodoridis G, Mentis A, Archimandritis A. 2010. CagA and VacA polymorphisms are associated with distinct pathological features in *Helicobacter pylori*-infected adults with peptic ulcer and non-peptic ulcer disease. J Clin Microbiol. 48(6):2237-2239.

Papatheodoridis GV, Sougioultzis S, Archimandritis AJ. 2006. Effects of Helicobacter pylori and nonsteroidal anti-inflammatory drugs on peptic ulcer disease: a systematic review. Clin Gastroenterol Hepatol. 4(2):130-142.

Parente JM, da Silva BB, Palha-Dias MP, Zaterka S, Nishimura NF, Zeitune JM. 2006. *Helicobacter pylori* infection in children of low and high socioeconomic status in northeastern Brazil. Am J Trop Med Hyg. 75(3):509-512.

Park C-G, Kim S, Lee E-J, Jeon H-S, Han S. 2018. Clinical relevance of point mutations in the 23S rRNA gene in *Helicobacter pylori* eradication: a prospective, observational study. Medicine (Baltimore). 97(33):e11835-e11835.

Park JY, Dunbar KB, Mitui M, Arnold CA, Lam-Himlin DM, Valasek MA, Thung I, Okwara C, Coss E, Cryer B et al. 2016. *Helicobacter pylori* clarithromycin resistance and treatment failure are common in the USA. Dig Dis Sci. 61(8):2373–2380.

Park SR, Mackay WG, Reid DC. 2001. *Helicobacter* sp. recovered from drinking water biofilm sampled from a water distribution system. Water Res. 35(6):1624–1626.

Parsonnet J, Friedman GD, Vandersteen DP, Chang Y, Vogelman JH, Orentreich N, Sibley RK. 1991. *Helicobacter pylori* infection and the risk of gastric carcinoma. New Engl J Med. 325(16):1127–1131.

Patel JB. 2001. 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. Mol Diagn. 6(4):313–321.

Paziak-Domańska B, Chmiela M, Jarosińska A, Rudnicka W. 2000. Potential role of CagA in the inhibition of T cell reactivity in *Helicobacter pylori* infections. Cell Immunol. 202(2):136–139.

Pearson JP, Van Delden C, Iglewski BH. 1999. Active efflux and diffusion are involved in transport of *Pseudomonas aeruginosa* cell-to-cell signals. J Bacteriol. 181(4):1203-1210.

Peek RM, Thompson SA, Donahue JP, Tham KT, Atherton JC, Blaser MJ, Miller GG. 1998. Adherence to gastric epithelial cells induces expression of a *Helicobacter pylori* gene, iceA, that is associated with clinical outcome. Proc Assoc Am Physicians. 110(6):531–544.

Penfold SS, Lastovica AJ, Elisha BG. 1988. Demonstration of plasmids in *Campylobacter pylori*. Int J Infect Dis. 157(4):850–851.

Pereira R, Oliveira J, Sousa M. 2020. Bioinformatics and computational tools for next-generation sequencing analysis in clinical genetics. J Clin Med. 9(1):132.

Perez-Perez GI, Olivares AZ, Foo FY, Foo S, Neusy AJ, Ng C, Holzman RS, Marmor M, Blaser MJ. 2005. Seroprevalence of *Helicobacter pylori* in New York City populations originating in East Asia. J Urban Health. 82(3):510-516.

Perri, Festa, Clemente, Quitadamo, Andriulli. 2000. Rifabutin-based 'rescue therapy' for *Helicobacter pylori* infected patients after failure of standard regimens. Aliment Pharmacol Therapeut. 14(3):311-316.

Pesci EC, Pickett CL. 1994. Genetic organization and enzymatic activity of a superoxide dismutase from the microaerophilic human pathogen, *Helicobacter pylori*. Gene. 143(1):111-116.

Pfeifer SP. 2017. From next-generation resequencing reads to a high-quality variant data set. Heredity. 118(2):111-124.

Pitson SM, Mendz GL, Srinivasan S, Hazell SL. 1999. The tricarboxylic acid cycle of *Helicobacter pylori*. European Journal of Biochemistry. 260(1):258-267.

Plummer M, de Martel C, Vignat J, Ferlay J, Bray F, Franceschi S. 2016. Global burden of cancers attributable to infections in 2012: a synthetic analysis. Lancet Glob Health. 4(9):e609–e616.

Porsch-Ozcürümez M, Doppl W, Hardt PD, Schnell-Kretschmer H, Tuncay M, Akinci A, Bilgin Y, Klör HU. 2003. Impact of migration on *Helicobacter pylori* seroprevalence in the offspring of Turkish immigrants in Germany. Turk J Pediatr Dis. 45(3):203-208.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS One. 5(3):e9490.

Price ND, Papin JA, Palsson BØ. 2002. Determination of redundancy and systems properties of the metabolic network of *Helicobacter pylori* using genome-scale extreme pathway analysis. Genome Res. 12(5):760-769.

Primrose SB, Twyman R. 2009. Comparative genomics. Principles of Genome Analysis and Genomics. John Wiley & Sons. p. 113 – 132.

Qin Z, Lin W-t, Zhu S, Franco AT, Liu J. 2017. Imaging the motility and chemotaxis machineries in *Helicobacter pylori* by cryo-electron tomography. J Bacteriol. 199(3):e00695–00616.

Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genom. 13(1):341.

Quek C, Pham ST, Tran KT, Pham BT, Huynh LV, Luu NBL, Le TKT, Quek K, Pham VH. 2016. Antimicrobial susceptibility and clarithromycin resistance patterns of *Helicobacter pylori* clinical isolates in Vietnam. F1000Res. 5:671-671.

Queralt N, Araujo R. 2007. Analysis of the survival of *H. pylori* within a laboratory-based aquatic model system using molecular and classical techniques. Microb Ecol. 54(4):771-777.

Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, Bore JA, Koundouno R, Dudas G, Mikhail A et al. 2016. Real-time, portable genome sequencing for Ebola surveillance. Nature. 530(7589):228–232.

Qureshi NN, Gallaher B, Schiller NL. 2014. Evolution of amoxicillin resistance of *Helicobacter pylori in vitro*: characterization of resistance mechanisms. Microb Drug Resist. 20(6):509–516.

R Core Team. 2019. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Randel A. 2018. *H. pylori* infection: ACG updates treatment recommendations. Am Fam Physician. 97(2):135–137.

Rasheed F, Campbell BJ, Alfizah H, Varro A, Zahra R, Yamaoka Y, Pritchard DM. 2014. Analysis of clinical isolates of *Helicobacter pylori* in Pakistan reveals high degrees of pathogenicity and high frequencies of antibiotic resistance. Helicobacter. 19(5):387-399.

Raymond J, Thiberg J-M, Chevalier C, Kalach N, Bergeret M, Labigne A, Dauga C. 2004. Genetic and transmission analysis of *Helicobacter pylori* strains within a family. Emerg Infect Dis. 10(10):1816–1821.

Read TD, Salzberg SL, Pop M, Shumway M, Umayam L, Jiang L, Holtzapple E, Busch JD, Smith KL, Schupp JM et al. 2002. Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. Science. 296(5575):2028–2033.

Reller LB, Weinstein M, Jorgensen JH, Ferraro MJ. 2009. Antimicrobial susceptibility testing: a review of general principles and contemporary practices. Clin Infect Dis. 49(11):1749–1755.

Reynolds DJ, Penn CW. 1994. Characteristics of *Helicobacter pylori* growth in a defined medium and determination of its amino acid requirements. Microbiology. 140(10):2649-2656.

Rhead JL, Letley DP, Mohammadi M, Hussein N, Mohagheghi MA, Eshagh Hosseini M, Atherton JC. 2007. A new *Helicobacter pylori* vacuolating cytotoxin determinant, the intermediate region, is associated with gastric cancer. Gastroenterology. 133(3):926–936.

Rhoads A, Au KF. 2015. PacBio sequencing and its applications. GPB. 13(5):278–289.

Ribeiro ML, Gerrits MM, Benvengo YHB, Berning M, Godoy APO, Kuipers EJ, Mendonça S, van Vliet AHM, Pedrazzoli J, Jr., Kusters JG. 2004. Detection of high-level tetracycline resistance in clinical isolates of *Helicobacter pylori* using PCR-RFLP. FEMS Immunol Med Microbiol. 40(1):57–61.

Ribeiro ML, Vitiello L, Miranda MCB, Benvengo YHB, Godoy APO, Mendonca S, Pedrazzoli J. 2003. Mutations in the 23S rRNA gene are associated with clarithromycin resistance in *Helicobacter pylori* isolates in Brazil. Ann Clin Microbiol Antimicrob. 2(1):11.

Richardson EJ, Watson M. 2012. The automatic annotation of bacterial genomes. Brief Bioinform. 14(1):1–12.

Ricker N, Qian H, Fulthorpe RR. 2012. The limitations of draft assemblies for understanding prokaryotic adaptation and evolution. Genomics. 100(3):167–175.

Rimbara E, Noguchi N, Kawai T, Sasatsu M. 2007. Correlation between substitutions in penicillin-binding protein 1 and amoxicillin resistance in *Helicobacter pylori*. Microbiol Immunol. 51(10):939–944.

Rimbara E, Noguchi N, Kawai T, Sasatsu M. 2008a. Mutations in penicillin-binding proteins 1, 2 and 3 are responsible for amoxicillin resistance in *Helicobacter pylori*. J Antimicrob Chemother. 61(5):995-998.

Rimbara E, Noguchi N, Kawai T, Sasatsu M. 2008b. Novel mutation in 23S rRNA that confers low-level resistance to clarithromycin in *Helicobacter pylori*. Antimicrob Agents Chemother. 52(9):3465-3466.

Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. 2009. Reordering contigs of draft genomes using the Mauve aligner. Bioinformatics. 25(16):2071-2073.

Rollason TP, Stone J, Rhodes JM. 1984. Spiral organisms in endoscopic biopsies of the human stomach. J Clin Pathol. 37(1):23–26.

Romaniuk PJ, Zoltowska B, Trust TJ, Lane DJ, Olsen GJ, Pace NR, Stahl DA. 1987. *Campylobacter pylori*, the spiral bacterium associated with human gastritis, is not a true *Campylobacter* sp. J Bacteriol. 169(5):2137–2141.

Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P. 1996. Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem. 242(1):84–89.

Rothenbacher D, Bode G, Berg G, Knayer U, Gonser T, Adler G, Brenner H. 1999. *Helicobacter pylori* among preschool children and their parents: evidence of parent-child transmission. Int J Infect Dis. 179(2):398–402.

Rowland M, Daly L, Vaughan M, Higgins A, Bourke B, Drumm B. 2006. Age-specific incidence of *Helicobacter pylori*. Gastroenterology. 130(1):65–72.

RStudio Team. 2020. RStudio: Integrated Development for R. Boston, MA: RStudio, PBC.

Rusk N. 2011. Torrents of sequence. Nat Methods. 8(1):44–44.

Russell AD. 2004. Types of antibiotics and synthetic antimicrobial agents. Hugo and Russell's Pharmaceutical Microbiology. p. 152–186.

Saad RJ, Schoenfeld P, Kim HM, Chey WD. 2006. Levofloxacin-based triple therapy versus bismuth-based quadruple therapy for persistent *Helicobacter pylori* infection: a meta-analysis. Am J Gastroenterol Suppl. 101(3):488.

Sabat AJ, Budimir A, Nashev D, Sá-Leão R, van Dijl JM, Laurent F, Grundmann H, Friedrich AW, Markers obotESGoE. 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Eurosurveillance. 18(4):20380.

Sabat AJ, van Zanten E, Akkerboom V, Wisselink G, van Slochteren K, de Boer RF, Hendrix R, Friedrich AW, Rossen JWA, Kooistra-Smid AMD. 2017. Targeted next-generation sequencing of the 16S-23S rRNA region for culture-independent bacterial identification - increased discrimination of closely related species. Sci Rep. 7(1):3434.

Saeidi E, Sheikhshahrokh A. 2016. *vacA* genotype status of *Helicobacter pylori* isolated from foods with animal origin. Biomed Res Int. 2016:8701067-8701067.

Saier MH, Jr. 1998. Molecular phylogeny as a basis for the classification of transport proteins from bacteria, archaea and eukarya. Adv Microb Physiol. 40:81-136.

Saito N, Konishi K, Sato F, Kato M, Takeda H, Sugiyama T, Asaka M. 2003. Plural transformation-processes from spiral to coccoid *Helicobacter pylori* and its viability. J Infect. 46(1):49–55.

Salama N, Guillemin K, McDaniel TK, Sherlock G, Tompkins L, Falkow S. 2000. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. Proc Natl Acad Sci USA. 97(26):14668–14673.

Salaün L, Audibert C, Le Lay G, Burucoa C, Fauchère J-L, Picard B. 1998. Panmictic structure of *Helicobacter pylori* demonstrated by the comparative study of six genetic markers. FEMS Microbiol Lett. 161(2):231-239.

Salipante SJ, SenGupta DJ, Cummings LA, Land TA, Hoogestraat DR, Cookson BT. 2015. Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology. J Clin Microbiol. 53(4):1072-1079.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. PNAS. 74(12):5463–5467.

Saranathan R, Levi MH, Wattam AR, Malek A, Asare E, Behin DS, Pan DH, Jacobs WR, Jr., Szymczak WA. 2020. *Helicobacter pylori* infections in the Bronx, New York: surveying antibiotic susceptibility and strain lineage by whole-genome sequencing. J Clin Microbiol. 58(3):e01591-01519.

Sasaki K, Tajiri Y, Sata M, Fujii Y, Matsubara F, Zhao M, Shimizu S, Toyonaga A, Tanikawa K. 1999. *Helicobacter pylori* in the natural environment. Scand J Infect Dis. 31(3):275–279.

Satou K, Shiroma A, Teruya K, Shimoji M, Nakano K, Juan A, Tamotsu H, Terabayashi Y, Aoyama M, Teruya M et al. 2014. Complete genome sequences of eight *Helicobacter pylori* strains with different virulence factor genotypes and methylation profiles, isolated from patients with diverse gastrointestinal diseases on Okinawa Island, Japan, determined using PacBio single-molecule real-time technology. Genome Announc. 2(2):e00286–00214.

Savoldi A, Carrara E, Graham DY, Conti M, Tacconelli E. 2018. Prevalence of antibiotic resistance in *Helicobacter pylori*: a systematic review and meta-analysis in World Health Organization regions. Gastroenterology. 155(5):1372–1382.

Schmidt TP, Perna AM, Fugmann T, Böhm M, Jan H, Haller S, Götz C, Tegtmeyer N, Hoy B, Rau TT et al. 2016. Identification of E-cadherin signature motifs functioning as cleavage sites for *Helicobacter pylori* HtrA. Sci Rep. 6(1):23264.

Schouls LM, Schot CS, Jacobs JA. 2003. Horizontal transfer of segments of the 16S rRNA genes between species of the *Streptococcus anginosus* group. J Bacteriol. 185(24):7241-7246.

Schubert M, Lindgreen S, Orlando L. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes. 9:88–88.

Schürch AC, van Schaik W. 2017. Challenges and opportunities for whole-genome sequencing–based surveillance of antibiotic resistance. Ann New York Acad Sci. 1388(1):108–120.

Schuster SC. 2008. Next-generation sequencing transforms today's biology. Nat Methods. 5(1):16–18.

Schwarz S, Morelli G, Kusecek B, Manica A, Balloux F, Owen RJ, Graham DY, van der Merwe S, Achtman M, Suerbaum S. 2008. Horizontal versus familial transmission of *Helicobacter pylori*. PLoS Pathog. 4(10):e1000180.

Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 30(14):2068–2069.

Seemann T, Goncalves da Silva A, Bulach D, Schultz M, Kwong J, Howden B. Nullarbor. n.d. Github; [accessed 2020 March 3]. https://github.com/tseemann/nullarbor.

Segata N, Börnigen D, Morgan XC, Huttenhower C. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat Comm. 4:2304–2304.

Sen K, Acosta J, Lye DJ. 2011. Effects of prolonged chlorine exposures upon PCR detection of *Helicobacter pylori* DNA. Curr Microbiol. 62(3):727-732.

Sennblad B, Lagergren J. 2009. Probabilistic orthology analysis. Syst Biol. 58(4):411–424.

Sepulveda AR, Yao Y, Yan W, Park DI, Kim JJ, Gooding W, Abudayyeh S, Graham DY. 2010. CpG methylation and reduced expression of O6-methylguanine DNA methyltransferase is associated with *Helicobacter pylori* infection. Gastroenterology. 138(5):1836–1844.e1834.

Shahamat M, Mai UE, Paszko-Kolva C, Yamamoto H, Colwell RR. 1991. Evaluation of liquid media for growth of *Helicobacter pylori*. J Clin Microbiol. 29(12):2835–2837.

Shahamat M, Paszko-Kolva C, Mai UE, Yamamoto H, Colwell RR. 1992. Selected cryopreservatives for long term storage of *Helicobacter pylori* at low temperatures. J Clin Pathol. 45(8):735-736.

Shakir SM, Otiso J, Keller G, Heule HV, Kohner P, Cole N, Schuetz AN, Richter SS, Couturier MR. 2018. Multicenter evaluation of the Etest vs. agar dilution for susceptibility testing of *Helicobacter pylori*. Open Forum Infect Dis. 5:605.

Shakya M, Ahmed SA, Davenport KW, Flynn MC, Lo C-C, Chain PSG. 2020. Standardized phylogenetic and molecular evolutionary analysis applied to species across the microbial tree of life. Sci Rep. 10(1):1723.

She F-F, Lin J-Y, Liu J-Y, Huang C, Su D-H. 2003. Virulence of water-induced coccoid *Helicobacter pylori* and its experimental infection in mice. World J Gastroenterol. 9(3):516-520.

Sheh A, Chaturvedi R, Merrell DS, Correa P, Wilson KT, Fox JG. 2013. Phylogeographic origin of *Helicobacter pylori* determines host-adaptive responses upon coculture with gastric epithelial cells. Infect Immun. 81(7):2468-2477.

Shiota S, Watada M, Matsunari O, Iwatani S, Suzuki R, Yamaoka Y. 2012. *Helicobacter pylori iceA*, clinical outcomes, and correlation with *cagA*: a meta-analysis. PLoS One. 7(1):e30354–e30354.

Shmuely H, Passaro D, Figer A, Niv Y, Pitlik S, Samra Z, Koren R, Yahav J. 2001. Relationship between *Helicobacter pylori* CagA status and colorectal cancer. Am J Gastroenterol Suppl. 96(12):3406–3410.

Shrestha P, Cooper BS, Coast J, Oppong R, Do Thi Thuy N, Phodha T, Celhay O, Guerin PJ, Wertheim H, Lubell Y. 2018. Enumerating the economic cost of antimicrobial resistance per antibiotic consumed to inform the evaluation of interventions affecting their use. Antimicrob Resist In. 7(1):98.

Si-Tuan N, Ngoc HM, Nhat LD, Nguyen C, Pham HQ, Huong NT. 2020. Genomic features, whole-genome phylogenetic and comparative genomic analysis of extreme-drug-resistant ventilator-associated-pneumonia *Acinetobacter baumannii* strain in a Vietnam hospital. Infect Genet Evol. 80:104178.

Silva Rossi-Aguiar VP, Navarro-Rodriguez T, Mattar R, Siqueira de Melo Peres MP, Correa Barbuti R, Silva FM, Carrilho FJ, Eisig JN. 2009. Oral cavity is not a reservoir for *Helicobacter pylori* in infected patients with functional dyspepsia. Oral Microbiol Immunol. 24(3):255–259.

Siroy A, Molle V, Lemaître-Guillier C, Vallenet D, Pestel-Caron M, Cozzone AJ, Jouenne T, Dé E. 2005. Channel formation by CarO, the carbapenem resistance-associated outer membrane protein of *Acinetobacter baumannii*. Antimicrob Agents Chemother. 49(12):4876-4883.

Sisson G, Goodwin A, Raudonikiene A, Hughes NJ, Mukhopadhyay AK, Berg DE, Hoffman PS. 2002. Enzymes associated with reductive activation and action of nitazoxanide,

nitrofurans, and metronidazole in *Helicobacter pylori*. Antimicrob Agents Chemother. 46(7):2116–2123.

Smiley R, Bailey J, Sethuraman M, Posecion N, Showkat Ali M. 2013. Comparative proteomics analysis of sarcosine insoluble outer membrane proteins from clarithromycin resistant and sensitive strains of *Helicobacter pylori*. J Microbiol. 51(5):612-618.

Smith JM, Smith NH, O'Rourke M, Spratt BG. 1993. How clonal are bacteria? PNAS. 90(10):4384-4388.

Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SBH, Hood LE. 1986. Fluorescence detection in automated DNA sequence analysis. Nature. 321(6071):674–679.

Smith MA, Edwards DI. 1995. Redox potential and oxygen concentration as factors in the susceptibility of *Helicobacter pylori* to nitroheterocyclic drugs. J Antimicrob Chemother. 35(6):751-764.

Smith MA, Edwards DI. 1997. Oxygen scavenging, NADH oxidase and metronidazole resistance in *Helicobacter pylori*. J Antimicrob Chemother. 39(3):347-353.

Smith SM, O'Morain C, McNamara D. 2014. Antimicrobial susceptibility testing for *Helicobacter pylori* in times of increasing antibiotic resistance. World J Gastroenterol. 20(29):9912–9921.

Sobala GM, Crabtree JE, Dixon MF, Schorah CJ, Taylor JD, Rathbone BJ, Heatley RV, Axon AT. 1991. Acute *Helicobacter pylori* infection: clinical features, local and systemic immune response, gastric mucosal histology, and gastric juice ascorbic acid concentrations. Gut. 32(11):1415–1418.

Soltis DE, Soltis PS. 2003. The role of phylogenetics in comparative genetics. Plant Physiol. 132(4):1790-1800.

Song H-Y, Li Y. 2013. Can eradication rate of gastric *Helicobacter pylori* be improved by killing oral *Helicobacter pylori*? World J Gastroenterol. 19(39):6645–6650.

Song Z-Q, Zhou L-Y. 2016. Hybrid, sequential and concomitant therapies for *Helicobacter pylori* eradication: a systematic review and meta-analysis. World J Gastroenterol. 22(19):4766–4775.

Sørensen UB, Larsen JL. 1986. Serotyping of *Vibrio anguillarum*. Appl Environ Microbiol. 51(3):593-597.

Souto R, Colombo APV. 2008. Detection of *Helicobacter pylori* by polymerase chain reaction in the subgingival biofilm and saliva of non-dyspeptic periodontal Patients. J Periodontol. 79(1):97-103.

Srinivasan R, Karaoz U, Volegova M, MacKichan J, Kato-Maeda M, Miller S, Nadarajan R, Brodie EL, Lynch SV. 2015. Use of 16S rRNA gene for identification of a broad range of clinically relevant bacterial pathogens. PLoS One. 10(2):e0117617-e0117617.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 22(21):2688-2690.

Stark RM, Gerwig GJ, Pitman RS, Potts LF, Williams NA, Greenman J, Weinzweig IP, Hirst TR, Millar MR. 1999. Biofilm formation by *Helicobacter pylori*. Lett Appl Microbiol. 28(2):121–126.

Stats NZ Tatauranga Aotearoa. International travel: December 2019. 2020. [accessed 2020 June 28]. https://www.stats.govt.nz/information-releases/international-travel-december-2019.

Stein L. 2001. Genome annotation: from sequence to biology. Nat Rev Genet. 2(7):493–503.

Stone GG, Shortridge D, Versalovic J, Beyer J, Flamm RK, Graham DY, Ghoneim AT, Tanaka SK. 1997. A PCR-oligonucleotide ligation assay to determine the prevalence of 23S rRNA gene mutations in clarithromycin-resistant *Helicobacter pylori*. Antimicrob Agents Chemother. 41(3):712–714.

Stratton CW, Warner RR, Coudron PE, Lilly NA. 1999. Bismuth-mediated disruption of the glycocalyx- cell wall of *Helicobacter pylori*: ultrastructural evidence for a mechanism of action for bismuth salts. J Antimicrob Chemother. 43(5):659–666.

Su C-C, Yang F, Long F, Reyon D, Routh MD, Kuo DW, Mokhtari AK, Van Ornam JD, Rabe KL, Hoy JA et al. 2009. Crystal structure of the membrane fusion protein CusB from *Escherichia coli*. J Mol Biol. 393(2):342-355.

Su H, Tissera K, Jang S, Choi YH, Kim A, Cho Y-J, Li M, Gunawardhana N, Merrell DS, Ge L et al. 2019a. Evolutionary mechanism leading to the multi-*cagA* genotype in *Helicobacter pylori*. Sci Rep. 9(1):11203–11203.

Su M, Satola SW, Read TD. 2019b. Genome-based prediction of bacterial antibiotic resistance. J Clin Microbiol. 57(3):e01405–01418.

Suarez G, Nathans D. 1965. Inhibition of aminoacyl-sRNA binding to ribosomes by tetracycline. Biochem Biophys Res Comm. 18(5):743–750.

Suerbaum S. 1995. The complex flagella of gastric *Helicobacter* species. Trends Microbiol. 3(5):168–170.

Suerbaum S. 2000. Genetic variability within *Helicobacter pylori*. International Journal of Medical Microbiology. 290(2):175-181.

Suerbaum S, Michetti P. 2002. *Helicobacter pylori* infection. New Engl J Med. 347(15):1175–1186.

Suerbaum S, Smith JM, Bapumia K, Morelli G, Smith NH, Kunstmann E, Dyrek I, Achtman M. 1998. Free recombination within *Helicobacter pylori*. Proc Natl Acad Sci USA. 95(21):12619–12624.

Suzuki T, Matsuo K, Ito H, Sawaki A, Hirose K, Wakai K, Sato S, Nakamura T, Yamao K, Ueda R et al. 2006. Smoking increases the treatment failure for *Helicobacter pylori* eradication. Am J Med. 119(3):217–224.

Swerdlow H, Gesteland R. 1990. Capillary gel electrophoresis for rapid, high resolution DNA sequencing. Nucleic Acids Res. 18(6):1415–1419.

Szymczak A, Ferenc S, Majewska J, Miernikiewicz P, Gnus J, Witkiewicz W, Dąbrowska K. 2020. Application of 16S rRNA gene sequencing in *Helicobacter pylori* detection. PeerJ. 8:e9099.

Tacconelli E, Carrara E, Savoldi A, Harbarth S, Mendelson M, Monnet DL, Pulcini C, Kahlmeter G, Kluytmans J, Carmeli Y et al. 2018. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. Lancet Infect Dis. 18(3):318–327.

Tait-Kamradt A, Davies T, Cronan M, Jacobs MR, Appelbaum PC, Sutcliffe J. 2000. Mutations in 23S rRNA and ribosomal protein L4 account for resistance in pneumococcal strains selected *in vitro* by macrolide passage. Antimicrob Agents Chemother. 44(8):2118-2125.

Talebi Bezmin Abadi A, Perez-Perez G. 2016. Role of *dupA* in virulence of *Helicobacter pylori*. World J Gastroenterol. 22(46):10118-10123.

Tamayo E, Montes M, Fernández-Reyes M, Lizasoain J, Ibarra B, Mendarte U, Zapata E, Mendiola J, Pérez-Trallero E. 2017. Clarithromycin resistance in *Helicobacter pylori* and its molecular determinants in Northern Spain, 2013–2015. J Glob Antimicrob Re. 9:43-46.

Tang X, Shen Y, Hu R, Yang T, Benghezal M, Li H, Tang H. 2020. Re-assessment of the disk diffusion technique for routine antimicrobial susceptibility testing for *Helicobacter pylori*. Helicobacter. 25(4):e12703.

Tanih NF, Ndip LM, Ndip RN. 2011. Characterisation of the genes encoding resistance to metronidazole (*rdxA* and *frxA*) and clarithromycin (the 23S-rRNA genes) in South African isolates of Helicobacter pylori. Ann Trop Med Parasit. 105(3):251-259.

Tankovic J, Lamarque D, Delchier JC, Soussy CJ, Labigne A, Jenks PJ. 2000. Frequent association between alteration of the *rdxA* gene and metronidazole resistance in French and North African isolates of *Helicobacter pylori*. Antimicrob Agents Chemother. 44(3):608-613.

Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28(1):33–36.

Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. Science. 278(5338):631–637.

Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Ciufo S, Li W. 2013. Prokaryotic genome annotation pipeline. The NCBI Handbook 2nd edition. National Center for Biotechnology Information (US).

Tay CY, Mitchell H, Dong Q, Goh K-L, Dawes IW, Lan R. 2009. Population structure of *Helicobacter pylori* among ethnic groups in Malaysia: recent acquisition of the bacterium by the Malay population. BMC Microbiol. 9(1):126.

Taylor DE, Ge Z, Purych D, Lo T, Hiratsuka K. 1997. Cloning and sequence analysis of two copies of a 23S rRNA gene from *Helicobacter pylori* and association of clarithromycin resistance with 23S rRNA mutations. Antimicrob Agents Chemother. 41(12):2621–2628.

Taylor NS, Fox JG, Akopyants NS, Berg DE, Thompson N, Shames B, Yan L, Fontham E, Janney F, Hunter FM. 1995. Long-term colonization with single and multiple strains of *Helicobacter pylori* assessed by DNA fingerprinting. J Clin Microbiol. 33(4):918-923.

Tee W, Fairley S, Smallwood R, Dwyer B. 1991. Comparative evaluation of three selective media and a nonselective medium for the culture of *Helicobacter pylori* from gastric biopsies. J Clin Microbiol. 29(11):2587–2589.

Teymournejad O, Mobarez AM, Hassan ZM, Talebi Bezmin abadi A. 2017. Binding of the *Helicobacter pylori* OipA causes apoptosis of host cells via modulation of Bax/Bcl-2 levels. Sci Rep. 7(1):8036.

The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters, version 10.0. 2020. [accessed 2020 April 17]. http://www.eucast.org.

The UniProt Consortium. 2014. UniProt: a hub for protein information. Nucleic Acids Res. 43(D1):D204-D212.

Therrien C, Levesque RC. 2000. Molecular basis of antibiotic resistance and β-lactamase inhibition by mechanism-based inactivators: perspectives and future directions. FEMS Microbiol Rev. 24(3):251–262.

Thomas JE, Gibson GR, Darboe MK, Weaver LT, Dale A. 1992. Isolation of *Helicobacter pylori* from human faeces. Lancet. 340(8829):1194–1195.

Thompson SA, Blaser MJ. 1995. Isolation of the *Helicobacter pylori recA* gene and involvement of the recA region in resistance to low pH. Infect Immun. 63(6):2185-2193.

Thorell K, Yahara K, Berthenet E, Lawson DJ, Mikhail J, Kato I, Mendez A, Rizzato C, Bravo MM, Suzuki R et al. 2017. Rapid evolution of distinct *Helicobacter pylori* subpopulations in the Americas. PLoS Genet. 13(2):e1006546-e1006546.

Tian R-M, Cai L, Zhang W-P, Cao H-L, Qian P-Y. 2015. Rare events of intragenus and intraspecies horizontal transfer of the 16S rRNA gene. Genome Biol Evol. 7(8):2310-2320.

Tilton RC, Lieberman L, Gerlach EH. 1973. Microdilution antibiotic susceptibility test: examination of certain variables. Appl Microbiol. 26(5):658–665.

Tocher JH, Edwards DI. 1992. The interaction of reduced metronidazole with DNA bases and nucleosides. Int J Radiat Oncol Biol Phys. 22(4):661-663.

Tomb J-F, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA et al. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature. 388(6642):539–547.

Toracchio S, Aceto GM, Mariani-Costantini R, Battista P, Marzio L. 2004. Identification of a novel mutation affecting domain V of the 23S rRNA gene in *Helicobacter pylori*. Helicobacter. 9(5):396-399.

Tran VH, Ha TMT, Le PTQ, Phan TN, Tran TNH. 2019. Characterisation of point mutations in domain V of the 23S rRNA gene of clinical *Helicobacter pylori* strains and clarithromycin-resistant phenotype in central Vietnam. J Glob Antimicrob Re. 16:87–91.

Travers KJ, Chin C-S, Rank DR, Eid JS, Turner SW. 2010. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res. 38(15):e159-e159.

Travis PB, Goodman KJ, O'Rourke KM, Groves FD, Sinha D, Nicholas JS, VanDerslice J, Lackland D, Mena KD. 2009. The association of drinking water quality and sewage disposal with *Helicobacter pylori* incidence in infants: the potential role of water-borne transmission. J Water Health. 8(1):192-203.

Trieber CA, Taylor DE. 2002. Mutations in the 16S rRNA genes of *Helicobacter pylori* mediate resistance to tetracycline. J Bacteriol. 184(8):2131–2140.

Truong-Bolduc QC, Dunman PM, Strahilevitz J, Projan SJ, Hooper DC. 2005. MgrA is a multiple regulator of two new efflux pumps in *Staphylococcus aureus*. J Bacteriol. 187(7):2395-2405.

Tsang AKL, Lee HH, Yiu S-M, Lau SKP, Woo PCY. 2017. Failure of phylogeny inferred from multilocus sequence typing to represent bacterial phylogeny. Sci Rep. 7(1):4536.

Tseng Y-S, Wu D-C, Chang C-Y, Kuo C-H, Yang Y-C, Jan C-M, Su Y-C, Kuo F-C, Chang L-L. 2009. Amoxicillin resistance with β-lactamase production in *Helicobacter pylori*. Eur J Clin Investig. 39(9):807–812.

Tsugawa H, Suzuki H, Muraoka H, Ikeda F, Hirata K, Matsuzaki J, Saito Y, Hibi T. 2011a. Enhanced bacterial efflux system is the first step to the development of metronidazole resistance in *Helicobacter pylori*. Biochem Biophys Res Comm. 404(2):656–660.

Tsugawa H, Suzuki H, Satoh K, Hirata K, Matsuzaki J, Saito Y, Suematsu M, Hibi T. 2011b. Two amino acids mutation of ferric uptake regulator determines *Helicobacter pylori* resistance to metronidazole. Antioxid Redox Sign. 14(1):15-23.

Tursi A, Di Mario F, Franceschi M, De Bastiani R, Elisei W, Baldassarre G, Ferronato A, Grillo S, Landi S, Zamparella M et al. 2017. New bismuth-containing quadruple therapy in patients infected with *Helicobacter pylori*: a first Italian experience in clinical practice. Helicobacter. 22(3):e12371.

Tyson J, O'Neil N, Jain M, Olsen H, Hieter P, Snutch T. 2017. Whole genome sequencing and assembly of a *Caenorhabditis elegans* genome with complex genomic rearrangements using the MinION sequencing device. bioRxiv.099143.

Umeda M, Kobayashi H, Takeuchi Y, Hayashi J, Morotome-Hayashi Y, Yano K, Aoki A, Ohkusa T, Ishikawa I. 2003. High prevalence of *Helicobacter pylori* detected by PCR in the oral cavities of periodontitis patients. J Periodontol. 74(1):129–134.

Umehara S, Higashi H, Ohnishi N, Asaka M, Hatakeyama M. 2003. Effects of *Helicobacter pylori* CagA protein on the growth and survival of B lymphocytes, the origin of MALT lymphoma. Oncogene. 22(51):8337–8342.

Uotani T, Miftahussurur M, Yamaoka Y. 2015. Effect of bacterial and host factors on *Helicobacter pylori* eradication therapy. Expert Opin Ther Targets. 19(12):1637–1650.

Ureta-Vidal A, Ettwiller L, Birney E. 2003. Comparative genomics: genome-wide analysis in metazoan eukaryotes. Nat Rev Genet. 4(4):251–262.

Valdivieso-García A, Imgrund R, Deckert A, Varughese BM, Harris K, Bunimov N, Reid-Smith R, McEwen S. 2009. Cost analysis and antimicrobial susceptibility testing comparing the E test and the agar dilution method in *Campylobacter jejuni* and *Campylobacter coli*. Diagn Microbiol Infect Dis. 65(2):168–174.

Vale FF, Nunes A, Oleastro M, Gomes JP, Sampaio DA, Rocha R, Vítor JMB, Engstrand L, Pascoe B, Berthenet E et al. 2017. Genomic structure and insertion sites of *Helicobacter pylori* prophages from various geographical origins. Sci Rep. 7:42471–42471.

Valle J, Kekki M, Sipponen P, Ihamäki T, Siurala M. 1996. Long-term course and consequences of *Helicobacter pylori* gastritis results of a 32-year follow-up study. Scand J Gastroenterol. 31(6):546–550.

van Amsterdam K, Bart A, van der Ende A. 2005. A *Helicobacter pylori* TolC efflux pump confers resistance to metronidazole. Antimicrob Agents Chemother. 49(4):1477–1482.

Van Der Poorten D, Katelaris PH. 2007. The effectiveness of rifabutin triple therapy for patients with difficult-to-eradicate *Helicobacter pylori* in clinical practice. Aliment Pharmacol Therapeut. 26(11-12):1537–1542.

van Doorn L-J, Glupczynski Y, Kusters JG, Mégraud F, Midolo P, Maggi-Solcà N, Queiroz DMM, Nouhan N, Stet E, Quint WGV. 2001. Accurate prediction of macrolide resistance in *Helicobacter pylori* by a PCR line probe assay for detection of mutations in the 23S rRNA gene: multicenter validation study. Antimicrob Agents Chemother. 45(5):1500-1504.

Van Dyk TK, Templeton LJ, Cantera KA, Sharpe PL, Sariaslani FS. 2004. Characterization of the *Escherichia coli* AaeAB efflux pump: a metabolic relief valve? J Bacteriol. 186(21):7196-7204.

Vandamme P, Harrington CS, Jalava K, On SLW. 2000. Misidentifying Helicobacters: the *Helicobacter cinaedi* example. J Clin Microbiol. 38(6):2261-2266.

Veldhuyzen van Zanten SJ, Sherman PM. 1994. *Helicobacter pylori* infection as a cause of gastritis, duodenal ulcer, gastric cancer and nonulcer dyspepsia: a systematic overview. CMAJ. 150(2):177–185.

Venerito M, Krieger T, Ecker T, Leandro G, Malfertheiner P. 2013. Meta-analysis of bismuth quadruple therapy versus clarithromycin triple therapy for empiric primary treatment of *Helicobacter pylori* infection. Digestion. 88(1):33–45.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA et al. 2001. The sequence of the human genome. Science. 291(5507):1304–1351.

Vergara M, Catalán M, Gisbert JP, Calvet X. 2005. Meta-analysis: role of *Helicobacter pylori* eradication in the prevention of peptic ulcer in NSAID users. Aliment Pharmacol Therapeut. 21(12):1411–1418.

Vergara M, Vallve M, Gisbert JP, Calvet X. 2003. Meta-analysis: comparative efficacy of different proton-pump inhibitors in triple therapy for *Helicobacter pylori* eradication. Aliment Pharmacol Therapeut. 18(6):647–654.

Versalovic J, Osato MS, Spakovsky K, Dore MP, Reddy R, Stone GG, Shortridge D, Flamm RK, Tanaka SK, Graham DY. 1997. Point mutations in the 23S rRNA gene of *Helicobacter pylori* associated with different levels of clarithromycin resistance. J Antimicrob Chemother. 40(2):283–286.

Versalovic J, Shortridge D, Kibler K, Griffy MV, Beyer J, Flamm RK, Tanaka SK, Graham DY, Go MF. 1996. Mutations in 23S rRNA are associated with clarithromycin resistance in *Helicobacter pylori*. Antimicrob Agents Chemother. 40(2):477–480.

Veses-Garcia M, Antypas H, Löffler S, Brauner A, Andersson-Svahn H, Richter-Dahlfors A. 2018. Rapid phenotypic antibiotic susceptibility testing of uropathogens using optical signal analysis on the nanowell slide. Front Microbiol. 9(1530).

Vester B, Douthwaite S. 2001. Macrolide resistance conferred by base substitutions in 23S rRNA. Antimicrob Agents Chemother. 45(1):1–12.

Walker AS, Eyre DW, Wyllie DH, Dingle KE, Harding RM, O'Connor L, Griffiths D, Vaughan A, Finney J, Wilcox MH et al. 2012. Characterisation of *Clostridium difficile* hospital ward-based transmission using extensive epidemiological data and molecular typing. PLoS Med. 9(2):e1001172-e1001172.

Wang D, Guo Q, Lv Z, Yuan Y, Gong Y. 2018. Molecular detection of *H.pylori* antibiotic-resistant genes and bioinformatics predictive analysis. bioRxiv.325654.

Wang D, Guo Q, Yuan Y, Gong Y. 2019. The antibiotic resistance of *Helicobacter pylori* to five antibiotics and influencing factors in an area of China with a high risk of gastric cancer. BMC Microbiol. 19(1):152.

Wang G, Taylor DE. 1998. Site-specific mutations in the 23S rRNA gene of *Helicobacter pylori* confer two types of resistance to macrolide-lincosamide-streptogramin B antibiotics. Antimicrob Agents Chemother. 42(8):1952-1958.

Wang G, Wilson TJ, Jiang Q, Taylor DE. 2001. Spontaneous mutations that confer antibiotic resistance in *Helicobacter pylori*. Antimicrob Agents Chemother. 45(3):727–733.

Wang H-C, Cheng F-C, Wu M-S, Shu H-Y, Sun HS, Wang Y-C, Su I-J, Wu C-J. 2015a. Genome sequences of three *Helicobacter pylori* strains from patients with gastric mucosa-associated lymphoid tissue lymphoma. Genome Announc. 3(2):e00229–00215.

Wang L, Lin Z, Chen S, Li J, Chen C, Huang Z, Ye B, Ding J, Li W, Wu L et al. 2017. Ten-day bismuth-containing quadruple therapy is effective as first-line therapy for *Helicobacter pylori*–related chronic gastritis: a prospective randomized study in China. Clin Microbiol Infect. 23(6):391–395.

Wang M-Y, Chen C, Shao C, Wang S-B, Wang A-C, Yang Y-C, Yuan X-Y, Shao S-H. 2015b. Intact long-type DupA protein in *Helicobacter pylori* is an ATPase involved in multifunctional biological activities. Microb Pathog. 81:53–59.

Wang X, Jordan IK, Mayer LW. 2015c. A phylogenetic perspective on molecular epidemiology. In: Tang Y-W, Sussman M, Liu D, Poxton I, Schwartzman J, editors. Molecular Medical Microbiology (Second Edition). Boston: Academic Press. p. 517–536.

Wang X, Sturegard E, Rupar R, Nilsson H-O, Aleljung PA, Carlen B, Willen R, Wadstrom T. 1997. Infection of BALB/c A mice by spiral and coccoid forms of *Helicobacter pylori*. J Med Microbiol. 46(8):657–663.

Wang Y, Coleman-Derr D, Chen G, Gu YQ. 2015d. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 43(W1):W78-W84.

Wang Y, Qian P-Y. 2009. Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. PLoS One. 4(10):e7401-e7401.

Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Maechler M, Magnusson A, Moeller S et al. 2015. gplots: Various R Programming Tools for Plotting Data.

Warren JR, Marshall B. 1983. Unidentified curved bacilli on gastric epithelium in actve chronic gastritis. Lancet. 321(8336):1273–1275.

Watari J, Chen N, Amenta PS, Fukui H, Oshima T, Tomita T, Miwa H, Lim K-J, Das KM. 2014. *Helicobacter pylori* associated chronic gastritis, clinical syndromes, precancerous lesions, and pathogenesis of gastric cancer development. World J Gastroenterol. 20(18):5461–5473.

Watson CL, Owen RJ, Said B, Lai S, Lee JV, Surman-Lee S, Nichols G. 2004. Detection of *Helicobacter pylori* by PCR but not culture in water and biofilm samples from drinking water distribution systems in England. J Appl Microbiol. 97(4):690-698.

Watts GS, Youens-Clark K, Slepian MJ, Wolk DM, Oshiro MM, Metzger GS, Dhingra D, Cranmer LD, Hurwitz BL. 2017. 16S rRNA gene sequencing on a benchtop sequencer: accuracy for identification of clinically important bacteria. J Appl Microbiol. 123(6):1584-1596.

Weeks DL, Eskandari S, Scott DR, Sachs G. 2000. A H+-gated rrea channel: the link between *Helicobacter pylori* urease and gastric colonization. Science. 287(5452):482-485.

Wei L, Liu Y, Dubchak I, Shon J, Park J. 2002. Comparative genomics approaches to study organism similarities and differences. J Biomed Informat. 35(2):142–150.

Wermeille J, Cunningham M, Dederding J-P, Girard L, Baumann R, Zelger G, Buri P, Metry J-M, Sitavanc R, Gallaz L et al. 2002. Failure of *Helicobacter pylori* eradication: is poor compliance the main cause? Gastroenterol Clin Biol. 26(3):216–219.

Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 13(6):e1005595.

Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 31(20):3350–3352.

Willén R, Carlén B, Wang X, Papadogiannakis N, Odselius R, Wadström T. 2000. Morphologic conversion of *Helicobacter pylori* from spiral to coccoid form: scanning (SEM) and transmission electron microscopy (TEM) suggest viability. Ups J Med Sci. 105(1):31–40.

Wirth T, Wang X, Linz B, Novick RP, Lum JK, Blaser M, Morelli G, Falush D, Achtman M. 2004. Distinguishing human ethnic groups by means of sequences from *Helicobacter pylori*: lessons from Ladakh. PNAS. 101(14):4746-4751.

Woese CR. 1987. Bacterial evolution. Microbiol Rev. 51(2):221–271.

Woese CR, Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. PNAS. 74(11):5088–5090.

Woese CR, Fox GE, Zablen L, Uchida T, Bonen L, Pechman K, Lewis BJ, Stahl D. 1975. Conservation of primary structure in 16S ribosomal RNA. Nature. 254(5495):83–86.

Wold B, Myers RM. 2008. Sequence census methods for functional genomics. Nat Methods. 5(1):19–21.

Woo PCY, Lau SKP, Teng JLL, Tse H, Yuen KY. 2008. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. Clin Microbiol Infect. 14(10):908-934.

Woo PCY, Ng KHL, Lau SKP, Yip K-t, Fung AMY, Leung K-w, Tam DMW, Que T-l, Yuen K-y. 2003. Usefulness of the microSeq 500 16S rRibosomal DNA-based bacterial identification system for identification of clinically significant bacterial isolates with ambiguous biochemical profiles. J Clin Microbiol. 41(5):1996–2001.

Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 15(3):R46.

Worku ML, Sidebotham RL, Walker MM, Keshavarz T, Karim QN. 1999. The relationship between *Helicobacter pylori* motility, morphology and phase of growth: implications for gastric colonization and pathology. Microbiology. 145(10):2803–2811.

World Health Organization. Antimicrobial resistance. 2020. [accessed August 17 2020]. https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance.

Wu DC, Hsu PI, Wu JY, Opekun AR, Kuo CH, Wu IC, Wang SSW, Chen A, Hung WC, Graham DY. 2010. Sequential and concomitant therapy with four drugs is equally effective for eradication of *H. pylori* infection. Clin Gastroenterol Hepatol. 8(1):36–41.e31.

Wu H, Shi XD, Wang HT, Liu JX. 2000. Resistance of *Helicobacter pylori* to metronidazole, tetracycline and amoxycillin. J Antimicrob Chemother. 46(1):121–123.

Wu JY, Kim JJ, Reddy R, Wang WM, Graham DY, Kwon DH. 2005. Tetracycline-resistant clinical *Helicobacter pylori* isolates with and without mutations in 16S rRNA-encoding genes. Antimicrob Agents Chemother. 49(2):578–583.

Wueppenhorst N, Stueger H-P, Kist M, Glocker E. 2009. Identification and molecular characterization of triple- and quadruple-resistant *Helicobacter pylori* clinical isolates in Germany. J Antimicrob Chemother. 63(4):648–653.

Wunder C, Churin Y, Winau F, Warnecke D, Vieth M, Lindner B, Zähringer U, Mollenkopf H-J, Heinz E, Meyer TF. 2006. Cholesterol glucosylation promotes immune evasion by *Helicobacter pylori*. Nat Med. 12(9):1030–1038.

Xia X. 2013. What is comparative genomics? In: Xia X, editor. Comparative Genomics. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 1–20.

Xiang Z, Censini S, Bayeli PF, Telford JL, Figura N, Rappuoli R, Covacci A. 1995. Analysis of expression of CagA and VacA virulence factors in 43 strains of *Helicobacter pylori* reveals that clinical isolates can be divided into two major types and that CagA is not necessary for expression of the vacuolating cytotoxin. Infect Immun. 63(1):94–98.

Xie Z, Tang H. 2017. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. Bioinformatics. 33(21):3340-3347.

Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, Zhang G, Gu YQ, Coleman-Derr D, Xia Q et al. 2019. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 47(W1):W52-W58.

Yahaghi E, Khamesipour F, Mashayekhi F, Safarpoor Dehkordi F, Sakhaei MH, Masoudimanesh M, Khameneie MK. 2014. *Helicobacter pylori* in vegetables and salads: genotyping and antimicrobial resistance properties. Biomed Res Int. 2014:757941-757941.

Yahiro K, Akazawa Y, Nakano M, Suzuki H, Hisatune J, Isomoto H, Sap J, Noda M, Moss J, Hirayama T. 2015. *Helicobacter pylori* VacA induces apoptosis by accumulation of connexin 43 in autophagic vesicles via a Rac1/ERK-dependent pathway. Cell Death Discov. 1(1):15035.

Yakoob J, Fan X, Hu G, Liu L, Zhang Z. 2001. Antibiotic susceptibility of *Helicobacter pylori* in the Chinese population. J Gastroenterol Hepatol. 16(9):981–985.

Yamaoka Y, Kwon DH, Graham DY. 2000. A M(r) 34,000 proinflammatory outer membrane protein (*oipA*) of *Helicobacter pylori*. PNAS. 97(13):7533–7538.

Yang B, Wang Y, Qian P-Y. 2016. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. BMC Bioinformatics. 17(1):135.

Yang J-C, Lin C-J, Wang H-L, Chen J-D, Kao JY, Shun C-T, Lu C-W, Lin B-R, Shieh M-J, Chang M-C et al. 2015. High-dose dual therapy is superior to standard first-line or rescue therapy for *Helicobacter pylori* infection. Clin Gastroenterol Hepatol. 13(5):895–905.e895.

Yang J, Zhang Y, Fan L, Zhu YJ, Wang TY, Wang XW, Chen DF, Lan CH. 2019. Eradication efficacy of modified dual therapy compared with bismuth-containing quadruple therapy as a first-line treatment of *Helicobacter pylori*. Am J Gastroenterol Suppl. 114(3):437–445.

Yang Y-J, Wu J-J, Sheu B-S, Kao A-W, Huang A-H. 2004. The *rdxA* gene plays a more major role than *frxA* gene mutation in high-level metronidazole resistance of *Helicobacter pylori* in Taiwan. Helicobacter. 9(5):400–407.

Yari F, Abiri R, Aryan E, Ahmadi Jouybari T, Navabi J, Alvandi A. 2016. Loop-mediated isothermal amplification as a fast noninvasive method of *Helicobacter pylori* diagnosis. J Clin Lab Anal. 30(5):464-470.

Ye X, Li J, Lu M, Deng G, Jiang X, Tian Y, Quan Y, Jian Q. 2011. Identification and molecular typing of *Streptococcus agalactiae* isolated from pond-cultured tilapia in China. Fish Res. 77(4):623-632.

Yohe S, Thyagarajan B. 2017. Review of clinical next-generation sequencing. Arch Pathol Lab Med. 141(11):1544–1557.

Yuan X-Y, Yan J-J, Yang Y-C, Wu C-M, Hu Y, Geng J-L. 2017. *Helicobacter pylori* with East Asian-type cagPAI genes is more virulent than strains with Western-type in some cagPAI genes. Braz J Microbiol. 48(2):218–224.

Yücel O, Sayan A, Yildiz M. 2009. The factors associated with asymptomatic carriage of *Helicobacter pylori* in children and their mothers living in three socio-economic settings. Jpn J Infect Dis. 62(2):120-124.

Yuki M, Sakamoto M, Kuwahara H, Hongoh Y, Ohkuma M. 2017. Draft genome sequence of *Lactococcus* sp. strain Rs-Y01, isolated from the gut of the lower termite *Reticulitermes speratus*. Genome Announc. 5(38):e00999–00917.

Zaman S, Fitzpatrick M, Lindahl L, Zengel J. 2007. Novel mutations in ribosomal proteins L4 and L22 that confer erythromycin resistance in *Escherichia coli*. Mol Microbiol. 66(4):1039-1050.

Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup F M, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 67(11):2640-2644.

Zarzecka U, Modrak-Wójcik A, Figaj D, Apanowicz M, Lesner A, Bzowska A, Lipinska B, Zawilak-Pawlik A, Backert S, Skorko-Glonek J. 2019. Properties of the HtrA protease From bacterium *Helicobacter pylori* whose activity is indispensable for growth under stress conditions. Front Microbiol. 10(961).

Zhang L, Mulrooney SB, Leung AFK, Zeng Y, Ko BBC, Hausinger RP, Sun H. 2006. Inhibition of urease by bismuth(III): implications for the mechanism of action of bismuth drugs. Biometals. 19(5):503–511.

Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. Nucleic Acids Res. 39(Web Server issue):W347–W352.

Zhou Z, White KA, Polissi A, Georgopoulos C, Raetz CRH. 1998. Function of *Escherichia coli* MsbA, an essential ABC family transporter, in Lipid A and phospholipid biosynthesis. J Biol Chem. 273(20):12466-12475.

Zullo A, Scaccianoce G, De Francesco V, Ruggiero V, D'Ambrosio P, Castorani L, Bonfrate L, Vannella L, Hassan C, Portincasa P. 2013. Concomitant, sequential, and hybrid therapy for *H. pylori* eradication: a pilot study. Clin Res Hepatol Gastroenterol. 37(6):647–650.

Appendices

Appendix A: Whole Genome Sequence Data Set

The whole genome dataset used in Sections 3.2.4, 3.2.7, and 4.2.7. Please view the dataset in Supplementary File Appendix A.

Appendix B: Quality of Raw Reads

The quality of the raw reads for the four *H. pylori* isolates analysed can be viewed below. For each isolate, GeneWiz (Suzhou, China) generated base content and distribution analysis (Figure S1), Phred quality score along reads (Figure S2) and a pie chart of the classification of raw reads (Figure S3).

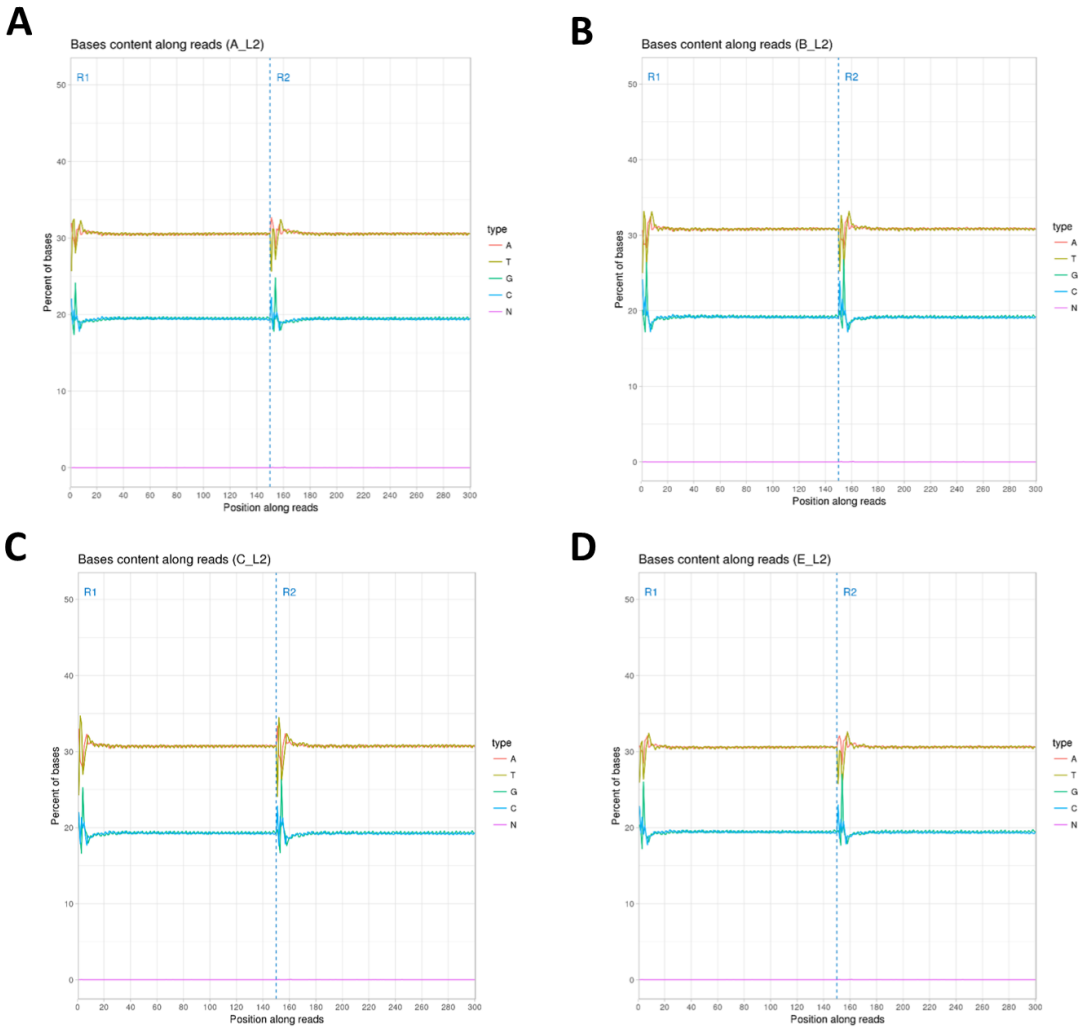

Figure S1: Base content along reads for isolates A (A), B (B), C (C), and E (D).

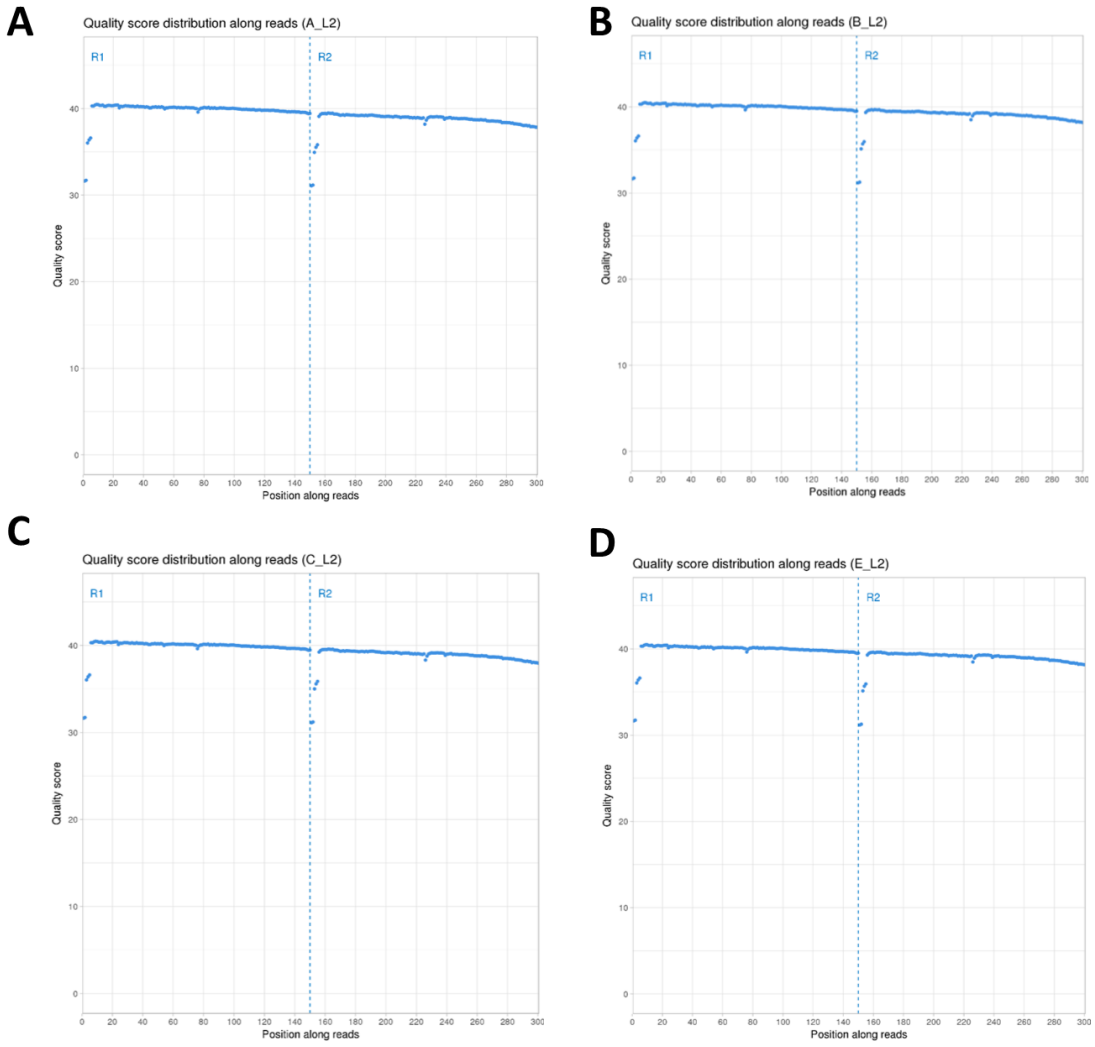

Figure S2: Phred quality score for isolates A (A), B (B), C (C), and E (D).

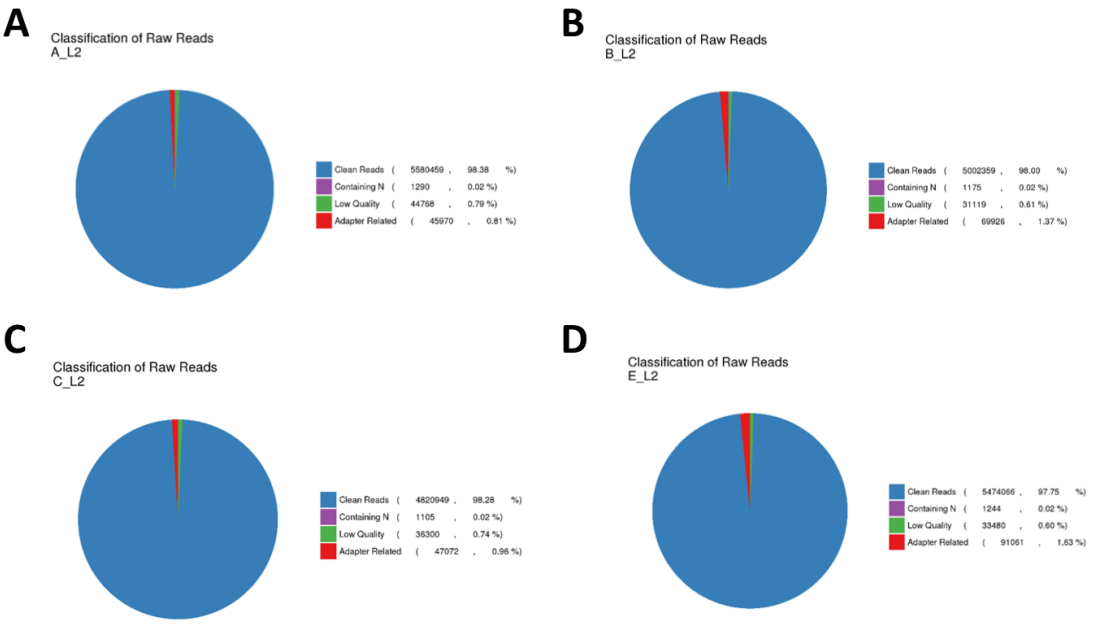

Figure S3: Raw read classification for isolates A (A), B (B), C (C), and E (D).

Appendix C: PhyloPhlAn Trees

The expanded nucleotide and protein sequence trees produced by PhyloPhlAn2 may be viewed in Figure S4 and Figure S5, respectively.

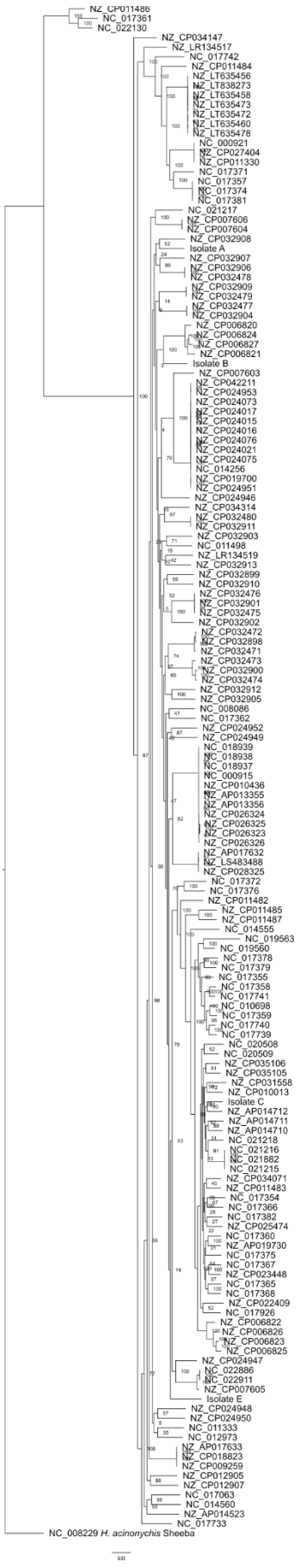

Figure S4: Full PhyloPhlAn2 tree of nucleotide sequences of 155 *H. pylori* complete genomes and four new *H. pylori* isolates (A, B, C, and E). Scale indicates number of substitutions per site.

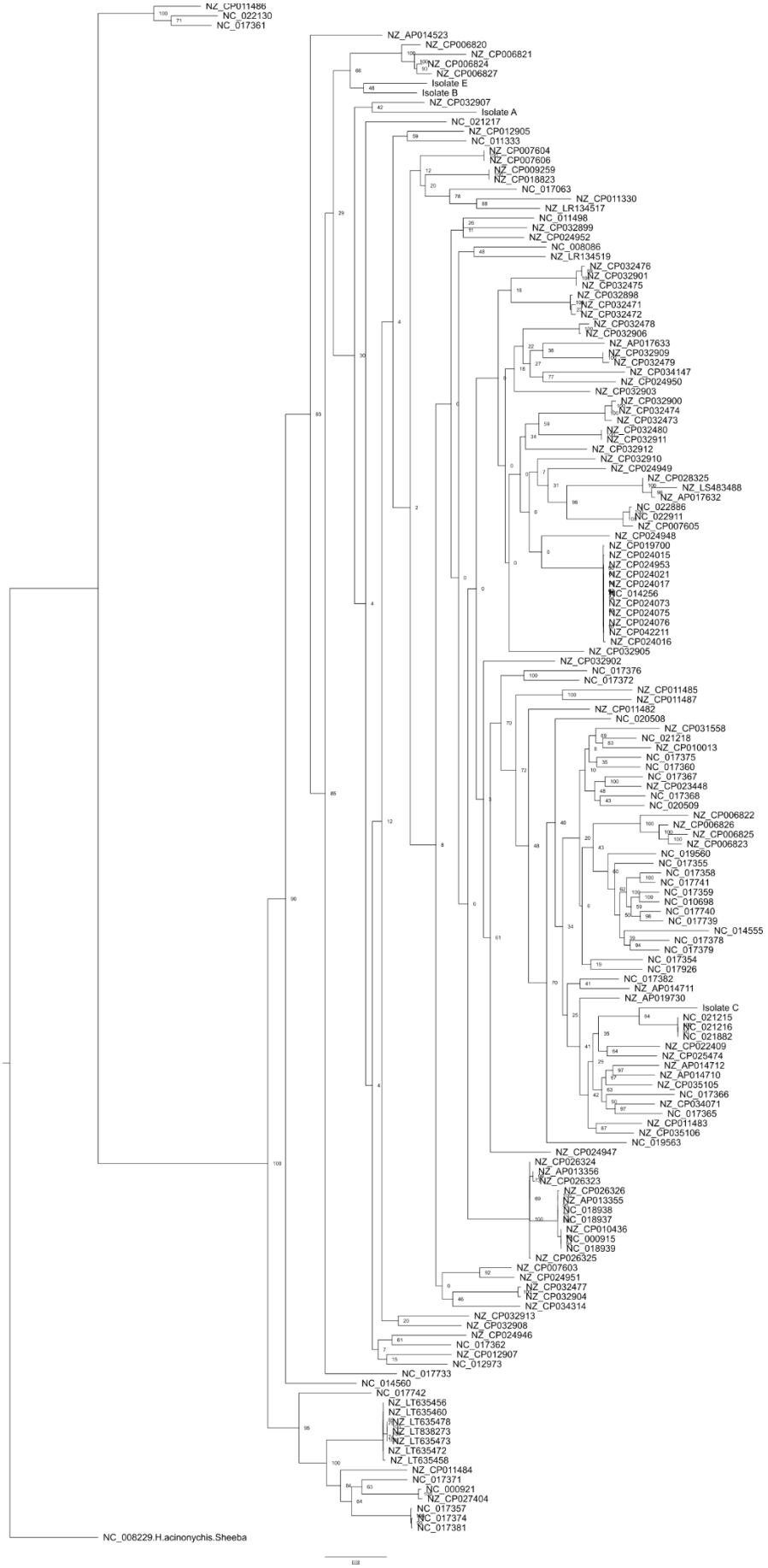

Figure S5: Full PhyloPhlAn2 tree of protein sequences of 155 *H. pylori* complete genomes and four new *H. pylori* isolates (A, B, C, and E). Scale indicates number of substitutions per site.

Appendix D: Nullarbor Report

The complete Nullarbor report may be viewed in Supplementary File Appendix D.

Appendix E: VFDB Output

The complete VFDB output may be viewed in Supplementary File Appendix E.

Appendix F: Efflux SNP Analysis

Evaluation of the SNPs identified in efflux pumps HefABC, HP1181, and HP1184 may be viewed in Supplementary File Appendix F.

Appendix G: RGI Output

The complete RGI output obtained from CARD may be viewed in Supplementary File Appendix G.

Appendix H: CARD Hit Analysis

Analysis of the proteins identified in isolates A, B, C, and E using CARD may be viewed in Supplementary File Appendix H.

Appendix I: OrthoVenn2 Protein Analysis

Analysis of the proteins identified in isolates B, C, and E using OrthoVenn2 may be viewed in Supplementary File Appendix I.

Appendix J: Presence/Absence of A2147G Mutation over Nucleotide PhyloPhlAn2 tree.

The expanded nucleotide PhyloPhlAn2 tree, overlaid with the presence of absence of the A2147G mutation can be viewed in Figure S6.

Figure S6: Full nucleotide sequence PhyloPhlAn2 tree overlaid with presence of A2147G mutation. Red circles next to accession numbers indicate that the A2147G mutation is present in all copies of the 23S rRNA gene. Blue circle indicates that the A2147G mutation was only present in one copy of the 23S rRNA gene. Scale indicates number of substitutions per site.