Examination of the phenotypic impacts of *Helicobacter pylori* diversification within a single host

Laura K. Jackson

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Nina Salama, Chair

Andrew Oberst

Kevin Cheung

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Examination of the phenotypic impacts of *Helicobacter pylori* diversification within a single host

Laura K. Jackson

Chair of the Supervisory Committee:
Nina R. Salama, Professor
Department of Microbiology, University of Washington

Helicobacter pylori, one of the most common bacterial pathogens colonizing humans, is the main agent responsible for stomach ulcers and cancer. Certain strain types are associated with increased risk of disease, however many factors contributing to disease outcome remain unknown. Prior work has documented genetic diversity among bacterial populations within single individuals, but the impact of this diversity for continued bacterial infection or disease progression remains understudied. In our analysis we examined both genetic and functional features of many stomach isolates from a single individual infected over six years. During these six years the subject shifted from having excess acid production and a duodenal ulcer to lower acid production from gastric atrophy. The 39 isolates form sub-populations based on gene sequence changes that accumulated in the different isolates. In addition to having

distinguishing genetic features, these sub-populations also have differences in several bacterial properties, including cell shape, motility, ability to activate immune responses, and colonization in healthy mice and a mouse model of metaplasia. This apparent functional specialization suggests that the bacterial sub-populations may have adapted to distinct niches within the stomach during chronic infection. However, in this patient we find subgroups with distinct genotypic and phenotypic properties cluster independently from the stomach region from which they were isolated.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Completion of this dissertation work would not have been possible without the many mentors, peers, collaborators, family, and friends who have supported me in many different ways throughout this experience.

To all the Salama lab members, I could not think of a more talented and enjoyable group of people to have experienced this journey with me. I will miss scrounging for free snacks and beer, commiserating over failed cloning attempts, celebrating grant and paper successes, and creating secret collaborative projects. These made my PhD memorable and fun. I would especially like to thank Valerie O'Brien and former members Kris Blair, Sean Schneider, Lauren Brumage and Hajirah Farah for their valuable scientific contributions to this dissertation work.

This project also would not have been possible without all the collaborators who pushed this project forward. I'd like to thank Rick Peek Jr. and Uma Krishna, at Vanderbilt for kindly donating the J99 samples, our collaborators at the Hutch, Trevor Bedford, who provided guidance on the computational analysis and use of Nextstrain, and Barney Potter who helped implement it. I owe a world of gratitude to the entire FHCRC Shared Resources especially: Matthew Fitzgibbon, Steve MacFarlane, Dave McDonald, and Luna Yu, the MCB program administrators, Maia Low and Andrea Brocato, and our HB admin, Pam Lindberg. I also appreciate all the members of my supervisory committee--Maitreya Dunham, Pradeep Singh, Kevin Cheung, and Andrew Oberst for guidance with the project.

I thank my parents, Evan and Louise, who almost never failed to pick up the phone when I called and who gave me the courage to take this huge leap in the first place, and my always loving siblings, Brian and Julia. Thank you to my husband, George, I hope you know how much your friendship, love, and support has meant to me throughout this experience. To all the additional “family”, thank you for cheering me on along every step of way. I am especially grateful for our local friends Hayley, Rachel, Natalie, and Nate who made Seattle a place that will always feel like a home.

Finally, I owe my sincerest gratitude to my wonderful mentor, Nina Salama. It is difficult to sum up all the ways your mentorship has positively shaped my PhD experience and development as a scientist. Thank you for leading by example, teaching me patience and diligence, providing resources to pursue exciting and interesting scientific questions, and allowing me the freedom to fail and strategies to rebuild. It has been a privilege to have you as a mentor.

DEDICATION

To my family,

your love and encouragement means the world to me.

To my husband George,

you have made this entire journey and lifelong dream possible.

Chapter 1. INTRODUCTION

1.1 OVERVIEW OF *HELICOBACTER PYLORI* PATHOGENESIS AND DISEASE

Helicobacter pylori, first discovered by Barry Marshall and Robin Warren in 1982 [1], is a gram negative bacterial pathogen whose only known reservoir is the human stomach. Estimations of colonization are around 50% globally [2]. The association of *H. pylori* with humans began early in our evolutionary history, before migrations out of Africa, dating back 50-60,000 years [3]. Infections are chronic, persisting throughout one's lifetime, and typically asymptomatic, inducing only mild gastritis in the majority of individuals. However in a fraction of cases, prolonged infection causes serious clinical disease including ulcers, MALT-lymphoma and gastric cancer [4]. Only about 15-20% of individuals develop ulcers and 1-2% progress to gastric cancer [5]. Yet, almost 90% of gastric cancer is associated with *H. pylori* infection, comprising 5.5% of the global burden of cancer [6,7]. Therefore strategies for predicting disease risk and intervention are imperative.

Risk of disease associated with infection is unevenly distributed across human populations in distinct geographic regions. Many populations in Asia have high prevalence of *H. pylori* associated disease relative to infection rates, whereas this pattern is the opposite in African populations [8,9]. Properties of the specific bacterial strain, host, and environment that modulate risk have been identified. Individuals carrying *H. pylori* strains that encode bacterial cytotoxin, CagA, or have polymorphisms in host innate immune pathways including IL-1β, have increased risk of gastric cancer [10,11]. Additionally, individuals with high sodium diets also have higher chances of

developing ulcers [12]. However, these factors alone are insufficient predictors of disease risk.

1.2 PATHOGENESIS OF *H. PYLORI*

The human stomach is the only known niche of *H. pylori.* The stomach is comprised of the main cavity or lumen where digestive enzymes and acid is secreted and the gastric epithelium which is protected by a thick layer of mucus. The epithelium has distinct glandular compartments--the corpus and antrum, which have unique gland architecture and cell type composition summarized in Figure 1.1. *H. pylori* has adapted several specialized properties that allow for its efficient colonization of the human stomach Figure 1.2. Entry into the gastric cavity likely occurs via oral-oral or oral-fecal transmission routes [13]. It cannot survive for long in the gastric lumen and therefore increases the local pH by secreting urease which converts urea to ammonia [14] and alters rheological properties of gastric mucins facilitating penetration of the gastric mucus [15]. Several sensors and receptors comprise an acid and nutrient sensing chemotaxis system which directs bacterial movement toward the gastric epithelium, its preferred niche [16].

H. pylori propels itself through the thick gastric mucus protecting the epithelium with unipolar sheathed flagella [17]. Non-motile mutants or mutants lacking specific chemotaxis proteins cannot establish colonization of animal models [18,19]. Extensive work in our lab has also identified helical cell shape as a bacterial factor enhancing stomach colonization [20]. Non-helical (rod) isolates can colonize mice but are outcompeted and elicit less gastric pathology compared to helical isolates [21]. Parameters of helical cell shape vary between clinical isolates and therefore it remains

an open question which morphologies are optimal for survival at different stages of disease.

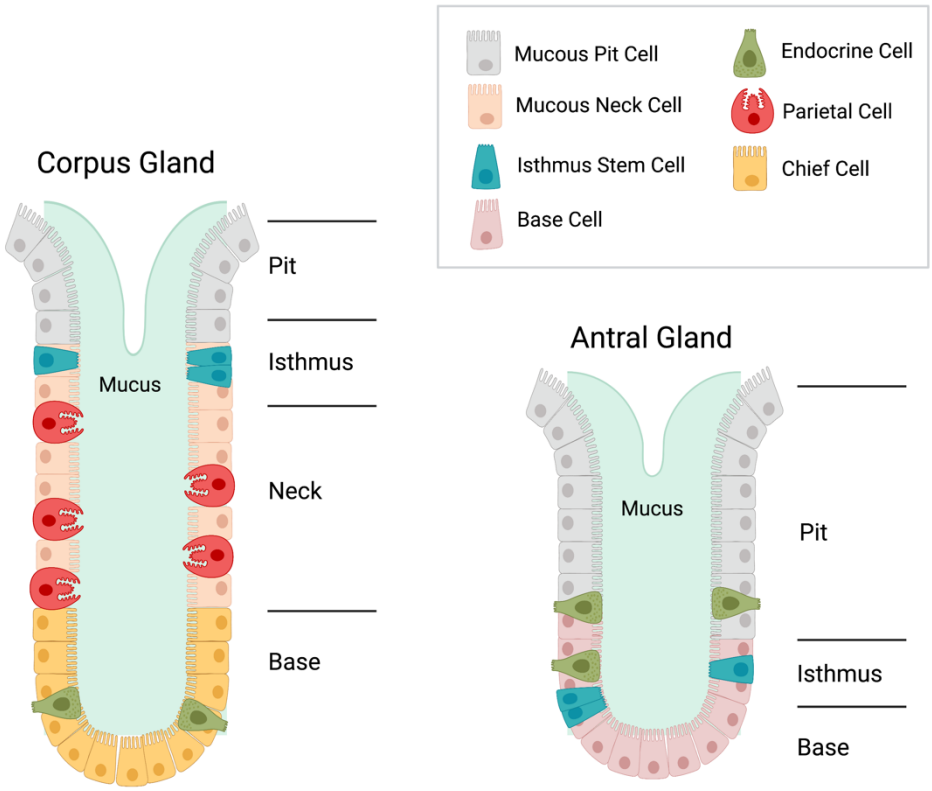


Figure 1.1. Cell type composition of corpus and antral glands.

Corpus gland lineages differentiate from the stem cells in the isthmus into endocrine cells, mucous cells, acid-secreting parietal cells, and zymogen-secreting chief cells. Antral glands lineages differentiate from stem cells closer to the base of the glands into base, endocrine, and pit lineages as described in the key.

Once *H. pylori* reaches the epithelium, it can bind to host receptors using specialized adhesins and disrupt host signaling via injected and secreted toxins, resulting in loss of polarity and breakdown of the epithelial barrier [22]. *H. pylori* has a specialized Type IV secretion system (T4SS) that injects the CagPAI encoded toxin CagA, and other

bacterial effectors including peptidoglycan components of the cell wall and polysaccharide components of the cell envelope into the host [23,24]. This both disrupts signaling within the host and modulates innate immune pathways through NF-κβ, increasing local cytokine secretion and resulting in immune cell recruitment [25,26]. *H. pylori* also has secreted bacterial toxins such as VacA, which induces vacuolation of host cells, Ggt which can initiate apoptotic cell death, and HtrA a serine protease which can cleave cell to cell junctions [27–30]. Reactive oxygen species generated by infiltrating neutrophils and macrophages can induce double stranded breaks in surrounding epithelial cells [31].

Manifestations of disease caused by infection is associated with location of colonization and inflammation and the nature of disruptions to acid homeostasis. *H. pylori* typically colonizes the antrum in human infections to avoid acid-secreting parietal cells in the corpus glands. However, colonizing populations can expand to the corpus during the course of infection [32]. Infection generates a robust immune response that results in persistent chronic inflammation of the stomach lining (gastritis) in all cases. Ulcers are associated with hyper acid secretion and colonization in the antrum. In contrast, gastric cancer is associated with acid hyposecretion and colonization in the corpus [33]. Gastric cancer develops in a series well-established histological steps progressing from loss of acid secretion due to atrophy of the corpus glands, SPEM (spasmolytic polypeptide expressing metaplasia), intestinal metaplasia (IM), dysplasia, and eventually gastric cancer [34].

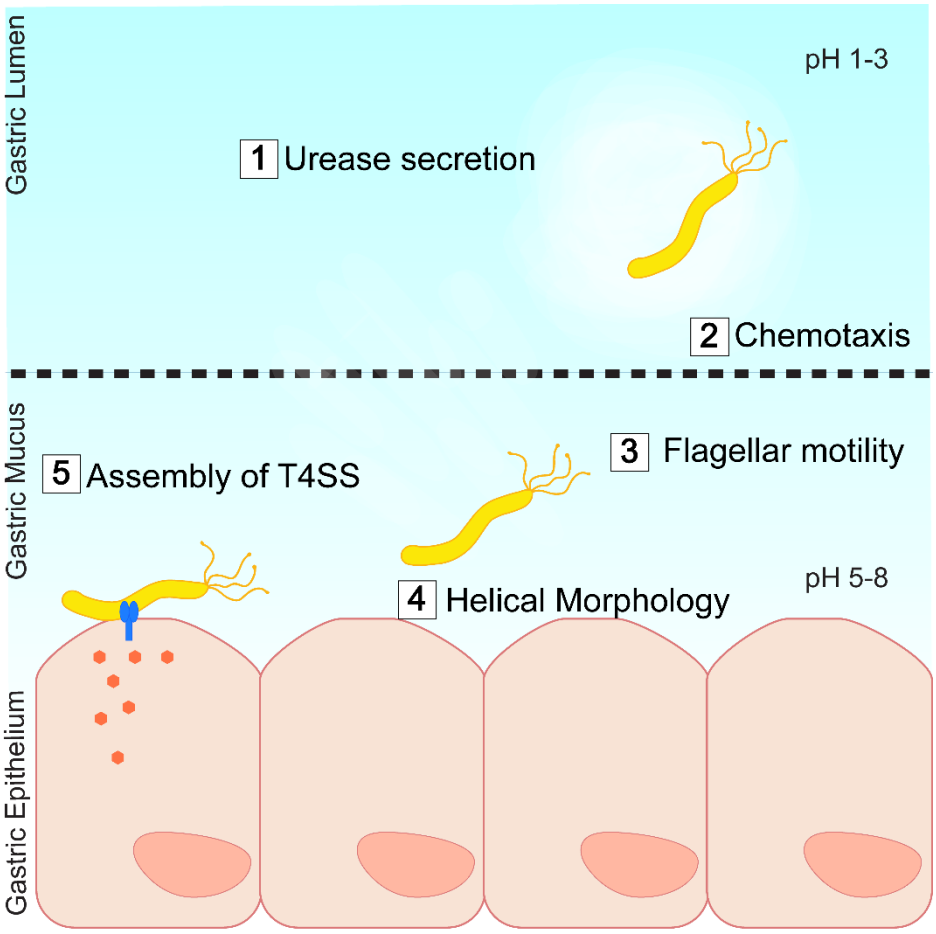


Figure 1.2. Mechanisms promoting *H. pylori* colonization of the gastric niche and pathogenesis

(1) Urease is secreted to increase local pH around the bacterium and altering the rheologic properties of gastric mucin. (2) Chemotaxis driven (3) flagellar motility enhanced by (4) helical cell shape allows *H. pylori* to penetrate the gastric mucus to colonize the gastric epithelium. Upon host cell attachment, the T4SS is assembled and can inject bacterial effectors into the host cell cytoplasm (5).

1.3 GENETIC DIVERSITY IMPACTS DISEASE OUTCOMES AND PROGRESSION

H. pylori is one of the most genetically plastic bacterial pathogens [35]. Genetic diversity among strains from different geographic regions is reflective of a panmictic population structure with variably present genes transferred horizontally via recombination [36].

Pairwise sequence comparisons of strains isolated from different geographic regions found between 6% and 7% of coding regions are variably present between strains [37], showing individual strains have many unique genetic features. The genome encodes a large set of highly variable outer membrane proteins that can be divided into five paralogous gene families comprising approximately 4% of the genome [38]. Other variably present genes are found within a large transposon (Tn) associated genomic island region called the plasticity zone (TnPz) [39]. Strains are defined into phylogeographic clusters, where *H. pylori* diversity mirrors evolution and distribution of human populations [40].

Specific strain factors impact disease severity. Individuals infected with strains carrying the Cag pathogenicity island (CagPAI), encoding the type IV secretion system (T4SS) and effector toxin CagA have two-fold increased risk for developing gastric cancer; and individuals with certain alleles of vacuolating cytotoxin (VacA) have over ten-fold increased risk [41–44]. CagA positive strains can be further stratified based on the specific tyrosine phosphorylation sites (EPIYA sites) present in the C-terminal region of the protein [45]. Many strains isolated from Asian populations carry a more virulent CagA EPIYA type (EPIYA-D), which partially accounts for higher rates of gastric cancer in these populations [46]. Strains expressing certain adhesins including *oipA*, s*abA*, *babA, hopQ*, also increase risk of disease [47].

1.4 MECHANISMS OF GENETIC DIVERSIFICATION

Much of gene regulation occurs through genomic alterations [48]. *H. pylori* has several phase variable genes whose expression is altered due to slipped strand mispairing in homo-polymeric tracts [49,50]. In addition, *H. pylori* has a somewhat

elevated baseline mutation rate compared to other bacteria; 10^6-10^8 substitutions/site/generation compared to the 10^9 substitutions/site/generation reported for *Escherichia coli* [51–53]. This is due to absent mismatch repair genes and deficiencies in the exonuclease domain of Pol1 [54]. However, base-excision repair is robust, preventing hypermutator phenotypes [55]. Variation is largely driven by high rates of intra and inter-genomic recombination. Intragenomic recombination can alter protein expression via gene conversion among paralogous families of outer membrane proteins [50]. Additionally, as a naturally competent bacterium, *H. pylori* incorporates DNA from genetically distinct strains into its chromosome, further varying gene content and sequence [56,57].

1.5 WITHIN HOST EVOLUTION AND SELECTIVE PRESSURES WITHIN THE STOMACH

Until recently, the extent of *H. pylori* genetic diversity within host was unknown and therefore ability to adapt in response to changing selective pressures associated with tissue remodeling and immune pressure was predicted to be limited. However, genome sequencing has allowed appreciation of the breadth of genetic diversity within the bacterial population infecting single hosts. This diversity likely reflects adaptation to specific niches within the stomach. *H. pylori* can colonize the inner gastric mucus layer, form epithelial cell adherent microcolonies, and penetrate into the gastric glands in both the antrum and corpus [58,59]. Furthermore, sequence divergence between paired isolates reveals evolution with molecular clock rate of 1.5×10^5 and recombination rate of 5.5×10^5 substitution per site per year [60,61], suggesting *H. pylori* can also adapt to disease associated tissue changes and fluctuations in immune responses that occur during prolonged infection.

1.6 J99 COLLECTION BACKGROUND AND SUMMARY

One such collection of isolates used to examine within host diversity was collected at Vanderbilt University from a single patient at two time points. At the time of the original biopsy (yr 1994), the source individual had a duodenal ulcer, indicative of *H. pylori* infection localized to the antrum and consistent with recovery of 13 single colonies from the single antral biopsy processed for culture. Six years later (yr 2000), after refusing antibiotic therapy, 30 additional single colony isolates were collected from distinct biopsy sites. At this time, this individual had corpus predominant gastritis and signs of gastric atrophy, including decreased production of stomach acid, indicating the spread of infection to the main body of the stomach [62,63]. One of the isolates, J99, was one of the first clinical isolates to be fully sequenced and a reference sequence was made publicly available in 1999 [64]. Israel *et al.* used microarrays to characterize a subset of these isolates in a 2001 study that showed individual colonies could be differentiated genetically [65].

1.7 A CASE STUDY OF WITHIN HOST EVOLUTION

Like Israel et al., a few other studies that have explored within host *H. pylori* isolates, uncovering substantial genetic diversity within a single host during the course of infection [66–68]. Most studies examining isolate populations from a single individual are limited to comparative sequence analysis on a few isolates or a single time point and have mostly focused on quantifying rather than characterizing diversity present. The sequence diversity detected by these studies have created greater interest in understanding the functional consequences of *H. pylori* diversity within host on pathogenesis. For example, Furuta et al., noted isolates collected within host and

between family members had sequence diversity in several known *H. pylori* virulence genes, including polymorphisms in *cagA* and *vacA* [69]. Falush et al. noted similar alterations in *vacA* and also in specific pathogenesis associated genes involved in urease secretion and flagellar motility in paired isolates collected from single hosts [70]. However, many open questions remain pertaining to the functional significance of adaptation particularly in the absence of antibiotic treatment, opening the door for more functional phenotypic studies. From the bacterial perspective, how do genetic adaptations observed promote survival, growth, and transmission. From the host perspective, how do changes to the infecting strain population impact interactions with the host and potentially modify disease outcome or progression. In the following chapters my collaborators and I have utilized the J99 collection of isolates to investigate within host genetic diversity and its impacts on phenotypes related to bacterial colonization and interaction with the human host. This exploration identified bacterial population subgroups with respect to both genotype and phenotype that cluster independently from the stomach region from which they were isolated. We were able to discover genetic determinants of phenotypic alterations in induction of the immune response, cell shape, (Chapter 2) and motility (Chapter 3). This genetic diversity also promotes differential colonization in a healthy mouse model and a mouse model of metaplasia (Chapter 2, 4).

Chapter 2. *HELICOBACTER PYLORI* DIVERSIFICATION DURING CHRONIC INFECTION DURING CHRONIC INFECTION WITHIN A SINGLE HOST GENERATES SUB-POPULATIONS WITH DISTINCT PHENOTYPES

2.1 PREFACE

This chapter is based off of a manuscript published with the following citation:

Jackson LK, Potter B, Schneider S, Fitzgibbon M, Blair K, Farah H, Krishna U, Bedford T, Peek RM Jr, Salama NR. *Helicobacter pylori* diversification during chronic infection within a single host generates sub-populations with distinct phenotypes. PLoS Pathog. 2020 Dec 28;16(12):e1008686. doi: 10.1371/journal.ppat.1008686. PMID: 33370399.

2.2 INTRODUCTION

Helicobacter pylori is a bacterial pathogen that colonizes the human stomach of approximately half of the world's population [71]. Infections persist throughout life without intervention and can lead to gastric and duodenal ulcers, MALT lymphoma, and gastric cancer in a subset of individuals [5,72]. *H. pylori* exhibits marked genetic diversity compared to other bacterial pathogens, which can impact treatment efficacy and disease severity [60,73–75]. Typically, antibiotics and proton pump inhibitors are employed for treatment, but variable prevalence of antibiotic resistance across populations make implementing a single treatment regimen difficult [76,77]. Strain-specific genotypes also contribute to increased disease risk within distinct ethno-geographic populations. Individuals with strains carrying the Cag pathogenicity island (Cag PAI), encoding a type IV secretion system (T4SS) and effector toxin CagA, have an increased risk of gastric cancer [10,78]. Cag PAI encoded genes, *cagA* and *cagY*,

exhibit significant allelic variation between individuals and have been identified as targets of positive selection within the global population [79]. Both CagA and CagY have been shown to modulate the host inflammatory response [80,81]. Recombination events within *cagY*, which encodes a structural component of the Cag T4SS with homology to the VirB10 component of other T4S systems, modifies secretion of inflammatory cytokines from epithelial cells [82,83]. CagA alters host responses through its interaction with intracellular kinases leading to the activation of the NFκB pathway [84,85]. In addition to Cag PAI genes, certain alleles of vacuolating cytotoxin, *vacA*, and frequent phase variation as well as recombination mediated gain and loss of outer membrane protein (OMP) adhesins BabA, SabA, and HopQ, have been linked to strain differences in pathogenesis [27,86–88].

Although *H. pylori* does encode several transcriptional regulators, much of gene regulation occurs through genomic alterations [48]. *H. pylori* has several phase variable genes whose expression is altered due to slipped strand mispairing in homo-polymeric tracts [49,50]. In addition, *H. pylori* has a somewhat elevated baseline mutation rate compared to other bacteria; 10^{6}-10^{8} substitutions/site/generation compared to the 10^{9} substitutions/site/generation reported for *Escherichia coli* [51–53]. This is due to absent mismatch repair genes and deficiencies in the exonuclease domain of Pol1 [54]. However, base-excision repair is robust, preventing hypermutator phenotypes [55]. Variation is largely driven by high rates of intra and inter-genomic recombination. Intragenomic recombination can alter protein expression via gene conversion among paralogous families of outer membrane proteins [50]. Additionally, as a naturally

competent bacterium, *H. pylori* incorporates DNA from genetically distinct strains into its chromosome, further varying gene content and sequence [56,57].

The human stomach is the only known niche for *H. pylori;* therefore, the breadth of genomic diversity across global populations likely reflects adaptation to individual host stomach environments [3]. More recently, genetic diversity within a single host has also become appreciated, suggesting the existence of structured niches within the stomach with distinct selective pressures [65,66,89,90]. *H. pylori* can colonize the epithelial surface of the inner gastric mucus layer, form cell adherent microcolonies, and penetrate into the gastric glands in both the antrum and corpus (Figure 2.1) [58,59]. The antrum and corpus have distinct gland architecture and cell type composition, providing unique challenges to bacterial survival. Gastric environments also change during lifelong infection. While most acute infections start in the antrum where the pH is closer to neutral, *H. pylori* can expand into the corpus [16,91]. Changes in bacterial localization are associated with histologic changes, including loss of the parietal cells (gastric atrophy), a risk factor for the development of gastric cancer [34]. These changes are accompanied by fluctuations in immune responses and alteration of glycosylation patterns affecting OMP-receptor binding to cell surface and mucus [49,86,92].

Whole genome sequencing of sequential isolates from individual infections detected accumulation of point mutations and frequent recombination among co-colonizing isolates. Within host diversity is acquired in a clock-like manner with average estimated mutation rate of 2.5×10^{-5} substitutions per site per year. However, in the majority of individuals recombination is the main driver of within-host diversification, which introduces changes at a median rate of 5.5×10^{-5} substitutions per initiation site

per year [61,70]. The rate of polymorphisms introduced via recombination compared to mutation (r/m rates) vary widely between individuals and likely depends on the presence of multiple co-infecting strains [93,94]. Phylogenetic studies of isolates within a single individual have observed signatures of diversifying selection in support of selective pressures during chronic stomach colonization [60,70,90]. However, phenotypic variation in infecting populations over time has not been well studied. To define both genetic changes that occur during infection and their functional consequences, we leveraged *H. pylori* isolates from a single individual collected at two time points spanning 6 years (1994-2000). One of multiple isolates obtained from culture of a single antral biopsy in 1994, J99, has previously been sequenced and a complete reference genome is available [64]. This same individual, who had not been successfully treated for *H. pylori*, underwent a repeat endoscopy performed in 2000 from which single biopsies from the corpus, antrum, and gastric metaplasia in the duodenum were cultured and additional *H. pylori* isolates recovered. A subset of isolates from the second time point (yr 2000) were analyzed by PCR microarray as part of a study highlighting diversity of isolates from a single individual [65].

Here we combined whole genome sequence analysis of multiple isolates from this subject with extensive phenotypic characterization to explore the rates and extent of genetic and phenotypic diversification within a single host. In this individual, during six years of chronic colonization, isolates adapted to occupy at least two distinct niches within the stomach reflected by differential ability to colonize a mouse model. We identified the genetic basis for modulation of Cag-dependent inflammatory cytokine induction and morphologic diversification. Neither of these phenotypes fully account for

the differences in colonization of the mouse model, highlighting the multifactorial selection pressures operant during chronic stomach colonization.

2.3 RESULTS

2.3.1 *Whole genome sequencing detects within-host genetic diversification of bacterial populations.*

For this study we analyzed 39 isolates from two distinct sampling time points. At the time of the original biopsy (yr 1994), the source individual had a duodenal ulcer, indicative of *H. pylori* infection localized to the antrum and consistent with recovery of multiple single colonies from the single antral biopsy processed for culture. Six years later (yr 2000), after refusing antibiotic therapy, additional single colony isolates were collected from distinct biopsy sites. At this time, this individual had corpus predominant gastritis and signs of gastric atrophy, including decreased production of stomach acid, indicating the spread of infection to the main body of the stomach (Figure 2.1) [62,63]. Twelve ancestral isolates, including *H. pylori* strain J99, were all collected in 1994 from the antral biopsy. From the second time point (recent, yr 2000), we analyzed 27 isolates from the antrum (n=12), corpus (n=12), and duodenum (n=3).

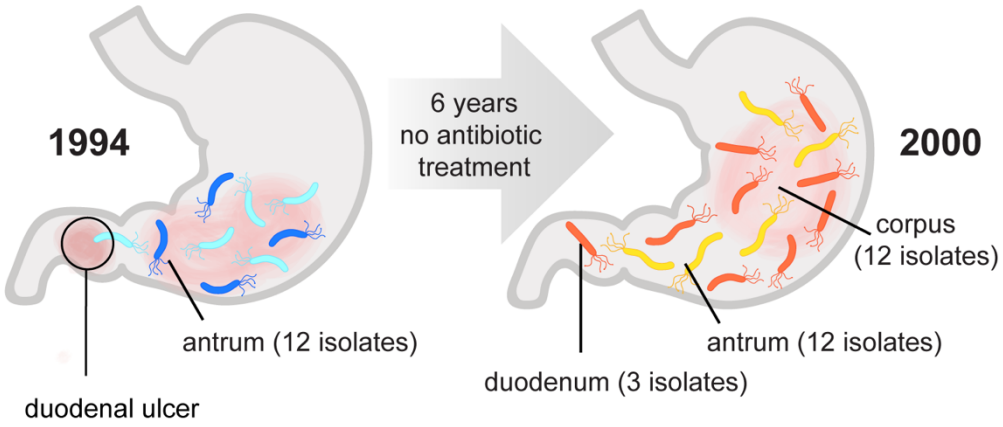


Figure 2.1. Sampling utilized to characterize genetic and phenotypic diversification of infecting population over six years.

This study leverages a collection of *H. pylori* isolates obtained from a treatment-naïve subject initially presenting with a duodenal ulcer at two different time points over a 6-year period of infection. A total of 12 isolates were analyzed from a single antral biopsy in 1994, and a total of 27 isolates were analyzed from single corpus, antrum, and duodenum biopsies collected in 2000 as indicated. In 2000 the subject displayed corpus atrophic gastritis and elevated stomach pH.

In order to measure the genetic diversity of *H. pylori* populations both within each time point and between time points, we performed whole genome sequencing using Illumina MiSeq. Short reads were aligned using the published sequence of J99 as the reference (AE001439). Our J99 isolate differed from the reference J99 (AE001439) at 755 polymorphic sites (625 SNPs, 130 indels); of these 553 polymorphisms were common to all isolates in the collection and were excluded from downstream analysis. All 39 isolates shared 99.99% average nucleotide sequence identity (ANI) to our J99 strain at shared sites mapping to the reference. By comparison, Alm et al. found J99

shares 92% ANI with strain 26695, originating from a distinct individual and geographic region [64]. High ANI among the isolates in the collection is consistent with either a single diversifying strain population or a mixed infection with highly related strains. Unique SNPs, and insertion and deletion (indel) events detected in the collection are reported in Table 2.1. Excluding polymorphisms common to all isolates, a total of 1,767 SNPs and 485 indels were identified (Table 2.1). This sequence variation represents changes introduced by both de-novo mutation and recombination**.** SNPs were distributed proportionally across coding and intergenic regions. By contrast, indels were biased towards intergenic regions (chi-squared, p-value<0.0001). Depletion of indels within coding regions likely reflects purifying selection due to high potential of indels to disrupt gene function by introducing frameshifts. Of the total unique SNPs and indels detected (n= 2,252), 238 were shared between ancestral and recent populations, while 263 and 1,751 were exclusively found within the ancestral and recent populations, respectively. The high number of polymorphisms unique to recent isolates demonstrates the substantial population divergence that occurred in this patient over time.

Table 2.1 Summary of unique SNPs, Indels detected by WGS among all the isolates (n=39) and in the subset of recent isolates (n=27, yr 2000)

	Total[a]	Coding[a]	nS[b]	S[b]	Intergenic[a]
Total SNPs	1,767	1,632	704	928	135
Recent SNPs	1,379	1,270	536	734	109
	Total[a]	**Coding[a]**			**Intergenic[a]**
Total Indels	485	298			187
Recent Indels	372	231			141

[a]Unique events are labeled as either coding or intergenic.
[b]Events within coding regions are further subdivided into nonsynonymous (nS) or synonymous (S) categories.

To assess extent of genetic diversity within the collection, we calculated the average pairwise genetic distance (π) for unique pairwise comparisons of isolates from the same time point (within time point). For between time point comparisons, only antral isolates (n=24) were used to reduce potential confounding effects introduced from comparing isolates from different anatomical locations. The average genetic distances (π, nucleotide differences/site) of within time point pairs was 6.75×10^{-5}, while π of between time point pairs was 8.23×10^{-4}, indicating within host evolution with an average molecular clock rate of 1.3×10^{-4} substitutions/site/year (Figure 2.2A). Overall, recent antral isolates have increased diversity ($\pi=9.9 \times 10^{-5}$) compared to the ancestral isolates ($\pi=3.6 \times 10^{-5}$), demonstrating accumulation of genetic diversity during chronic infection (Figure 2.2B).

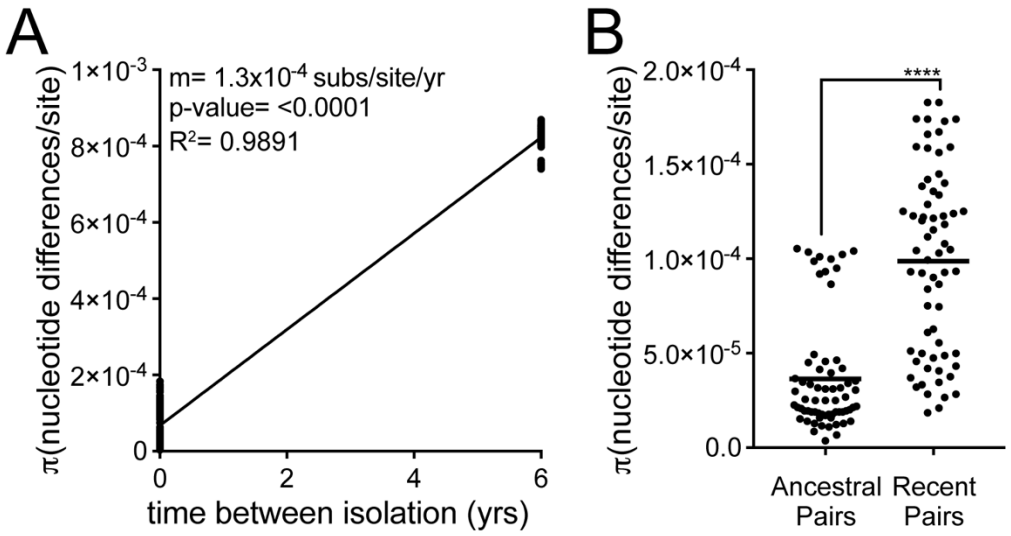


Figure 2.2. Pairwise comparisons show within host diversification with increasing diversity of infecting populations over time.

(**A**) Plot shows genetic distance for all pairwise comparisons of antral samples isolated within the same time point (n=132, time between isolation=0 yrs) and samples isolated

from different time points (n=144, time between isolation=6 yrs). Each point is a unique pairwise comparison (n=276). Linear regression with the slope (m) as the estimation of the molecular clock rate, p-value derived from F-test, and correlation coefficient (R^2) shown. (**B**) Each point represents a pairwise comparison between antral isolates within the ancestral population (yr 1994, n=66) or recent population (yr 2000, n=66). The average values between all pairwise comparisons in the population (π statistic) is shown with a black bar. Significance was determined using a Student's t-test (****, $p<0.0001$).

2.3.2 *Identification of genomic regions enriched for within-host genetic variation*

To identify regions of the genome that accumulate within host variation, we examined enrichment of nonsynonymous SNPs (nSNPs) in specific genes and functional classes assigned by the microbial genome database (MGDB) [95]. Out of the 1,495 genes in the reference sequence, 931(62.2%) are annotated with a functional class (Figure 2.3A). Enrichment or depletion was determined by comparing the distribution of nSNPs among MGDB classes to expected values based on a normal distribution (Figure 2.3). We observed cell envelope genes, including OMPs, accumulated a disproportionate number of nSNPs in both the total dataset of unique nSNPs and the subset of nSNPs unique to recent group of isolates (Figure 2.3B-C). This was expected since other studies that have looked at within host populations have shown OMPs accumulate diversity due to frequent inter and intragenomic recombination [60,61,88]. Accordingly, cell envelope diversification may serve a selective advantage in both the acute phase of infection as a mechanism of adaptation to a specific host and in the chronic phase of infection as a mechanism to persist in changing host environments.

The MGDB does not specifically analyze antibiotic resistance genes. While this subject had no known history of antibiotic treatment, the presence of antibiotic

resistance to metronizidole, ampicillin, clarithromycin, was previously tested. Four isolates, three antral and one duodenal, were resistant to clarithromycin due to a mutation in the 23S rRNA gene, but all the other isolates were sensitive to all three [65]. To validate, we queried our sequence data for polymorphisms known to confer antibiotic resistance listed in the Pathosystems Resource Integration Center (PATRIC) database. Using these methods, no additional polymorphisms indicative of antibiotic resistance were discovered.

Individual genes acquiring the most genetic variation over the six year period were identified. The number of unique nSNPs among the recent isolates detected for each gene were counted and weighted according to the gene length. The genes most highly enriched for unique nSNPs are listed in Table 2.2. Genes with excess accumulation of nSNPs during chronic infection.

Many of the genes identified encode OMPs (*babA*, *sabA*, *sabB*, and *hopQ*) that play roles in adhesion and exhibit variation between and within hosts [96,97]. Accumulation of nSNPs contributed to allelic diversification. For 7 of the top 15 genes listed, multiple unique alleles were detected within the recent population. For 4 of these genes, jhp1300 and jhp1409, and two outer membrane proteins *sabA* and *babA,* allelic variation in the ancestral population was also detected.

In order to differentiate polymorphism introduced by recombination from repeated mutation, we bioinformatically identified importation sites with ClonalFrameML [98]. A total of 228 importation events across 52 genes were detected in total. All of the genes accumulating an excess of nSNPs listed in Table 2.2 had at least one detectable recombination event, except jhp0929, a gene of unknown function located within a large

highly variable TnPZ associated genomic region called the plasticity zone [39]. The average import size (δ) was 136bp and divergence of import (ν) was 3.9 x 10^{-2} subs/site, suggesting imports were from a genetically distinct strain either not sampled or not present at the time of sampling.

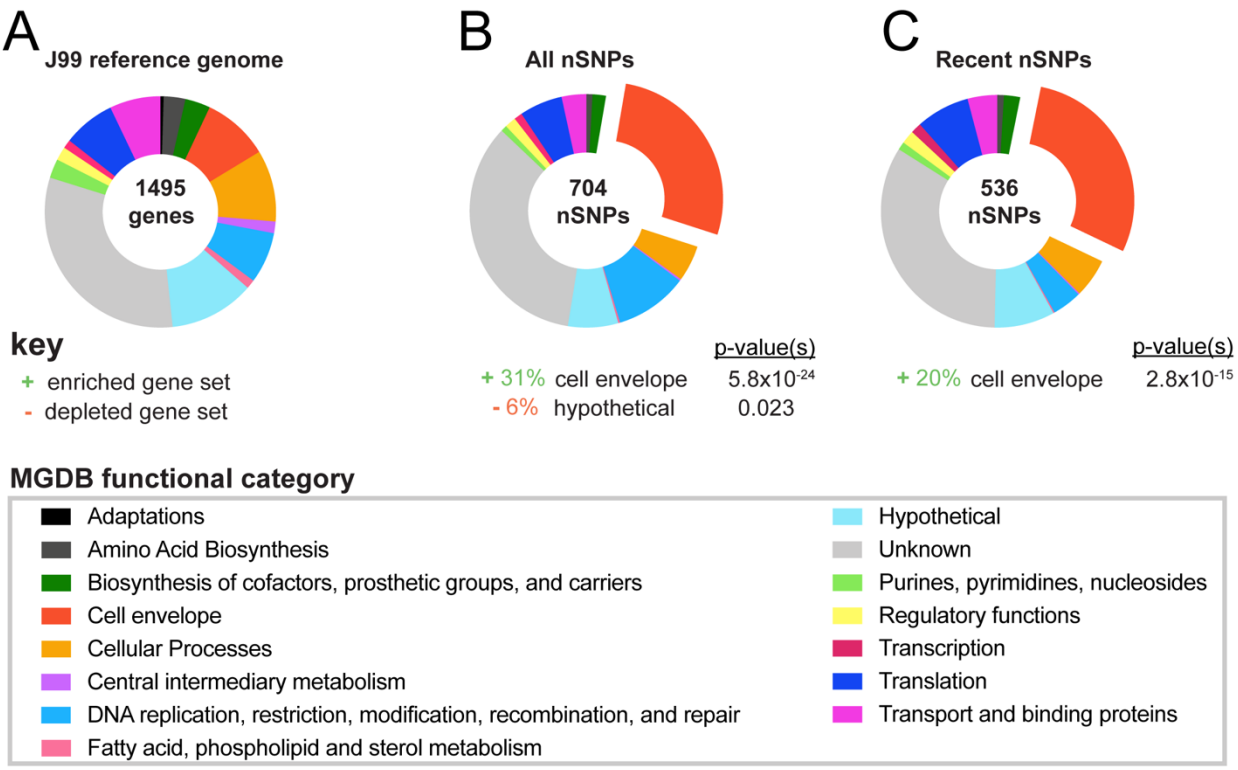


Figure 2.3. Cell envelope genes accumulate genetic variation during chronic infection.

(**A**) Proportion of genes within the reference genome (J99) comprising each of 14 functional classes identified in the Microbial Genome Database and color-coded according to key [95]. Percentage of genes with unknown function are labeled in gray. (**B-C**) Proportion of nSNPs from (**B**) the entire dataset (all nSNPs) and (**C**) from the subset unique to recent isolates (recent nSNPs) that fall within each functional class. Categories with statistically significant enrichment or depletion are listed below each chart with associated percentages and p-values (<0.03). Fisher's exact tests were used

to determine significance and corrected for multiple testing using Benjamini and Hochberg false discovery rate methods.

Table 2.2. Genes with excess accumulation of nSNPs during chronic infection

Gene ID[a]	Annotation[a]	MGDB function[a]	Unique nSNPs	Z-scores[b]	Estimated no. of alleles[c]	
					Ancestral	Recent
jhp1300		Unknown	18	21.52	2	19
jhp1103	*hopQ*	Outer membrane protein	45	14.89	1	11
jhp1068	*birA*	Biotin protein ligase	11	10.96	1	1
jhp0659	*sabB*	Outer membrane protein	27	8.94	1	13
jhp0303		Translocase subunit	3	7.9	1	1
jhp1096	*glnP_1*	Transport and binding proteins	8	7.73	1	1
jhp1409		Unknown	44	7.39	2	19
jhp0302	*argS*	Arginine-tRNA ligase	18	6.98	1	1
jhp0336		Unknown	19	5.07	1	1
jhp1097	*glnP_2*	ABC transporter permease	5	4.63	1	1
jhp0634		Unknown	7	4.29	1	1
jhp1102		Guanine permease	9	4.26	1	1
jhp0833	*babA*	Outer membrane protein	15	4.16	2	4
jhp0662	*sabA*	Outer membrane protein	13	4.12	4	5
jhp0929		Unknown	3	4.1	1	3

[a]The top 15 annotated genes with Z-scores > 4 for number of nSNPs uniquely acquired in the yr 2000 group of isolates within single genes are shown. [b]Z-scores were calculated using number of counts per gene normalized according to gene length. [c]The estimated number of alleles among 27 isolates in the recent population (1994) and 12 isolates in the ancestral population are shown.

2.3.3 *Genomic diversity within the recent population is not driven by stomach region specific adaptation*

To display the genetic relatedness of all the isolates in the collection, a maximum likelihood tree was generated with consensus sequences from the SNP calls using the Nextstrain platform (Figure 2.4A, Figure 2.5A [99]). Isolates collected from the two separate time points (1994 and 2000) cluster into distinct groups on the tree with a long branch representing an average divergence of 7.8×10^{-4} substitutions/site between the two populations. This depiction does not represent a clonal genealogy however, as

SNPs included in these analyses can arise from recombination and therefore are not necessarily clonally derived [100].

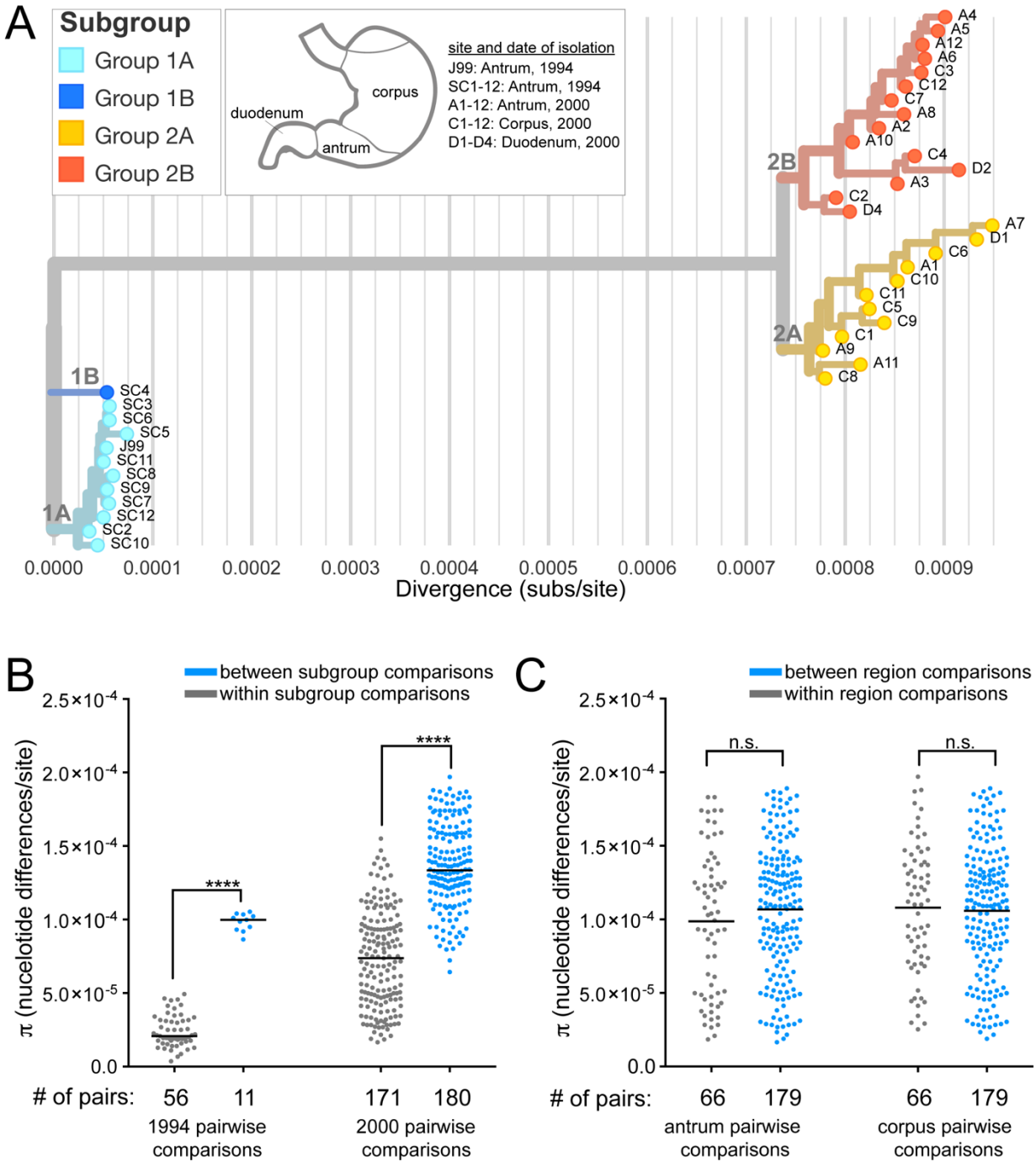


Figure 2.4. Clustering of strains by genetic similarity suggests distinct subgroups that do not correlate with biopsy site.

(**A**) A maximum likelihood tree was generated from all SNPs in the collection with Nextstrain [99]. Isolates were named according to the anatomic region of their source biopsy as annotated in the key and branch coloring was added to distinguish genetically related subgroups within the collection. Light and dark blue isolates are from 1994 (1A, 1B, respectively). Yellow (2A) and red isolates (2B) are from 2000. The X-axis shows the divergence from the ancestral root (number of substitutions (subs)/site). (**B-C**) Point on the plot represents the pairwise genetic distances calculated for groups of isolates described with a black bar representing the mean (π). (**B**) Pairwise comparisons of isolates within subgroups displayed on the maximum likelihood tree (gray) have smaller genetic distances on average than pairwise comparisons of isolates from different subgroups (blue) from the same time point. Significance was determined using a Student's t-test (****, $p<0.0001$). (**C**) Pairwise comparisons of recent isolates within the same stomach regions (gray) have the same average genetic distance as pairwise comparisons of isolates from different stomach regions (blue) in both the antrum and corpus.

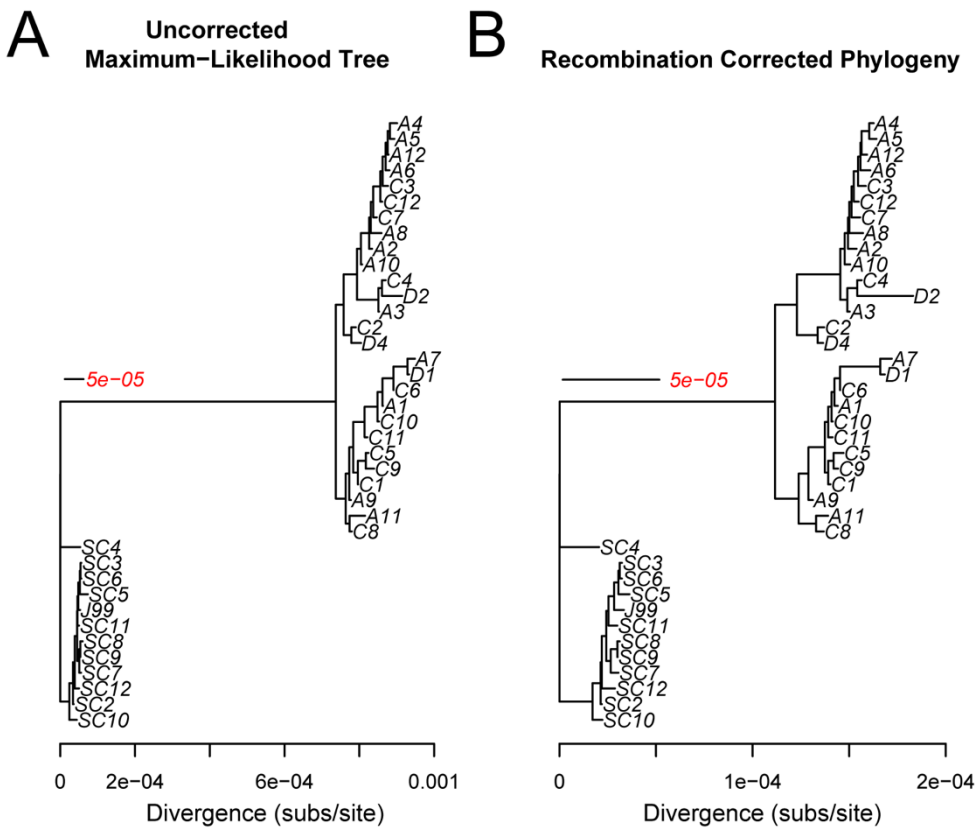


Figure 2.5. Comparison of maximum-likelihood tree and recombination corrected phylogeny have altered branch scale but the same overall topology.

X-axis represents divergence from common ancestor (substitutions/site) and scale bar with length indicated in red text is shown. Scale of branches is indicated in red. (**A**) The maximum likelihood tree constructed in Nextstrain (Figure 2.4A) and (**B**) the recombination corrected phylogeny created in ClonalFrameML with rescaled branch lengths are shown. Tree construction is described in detail in the methods.

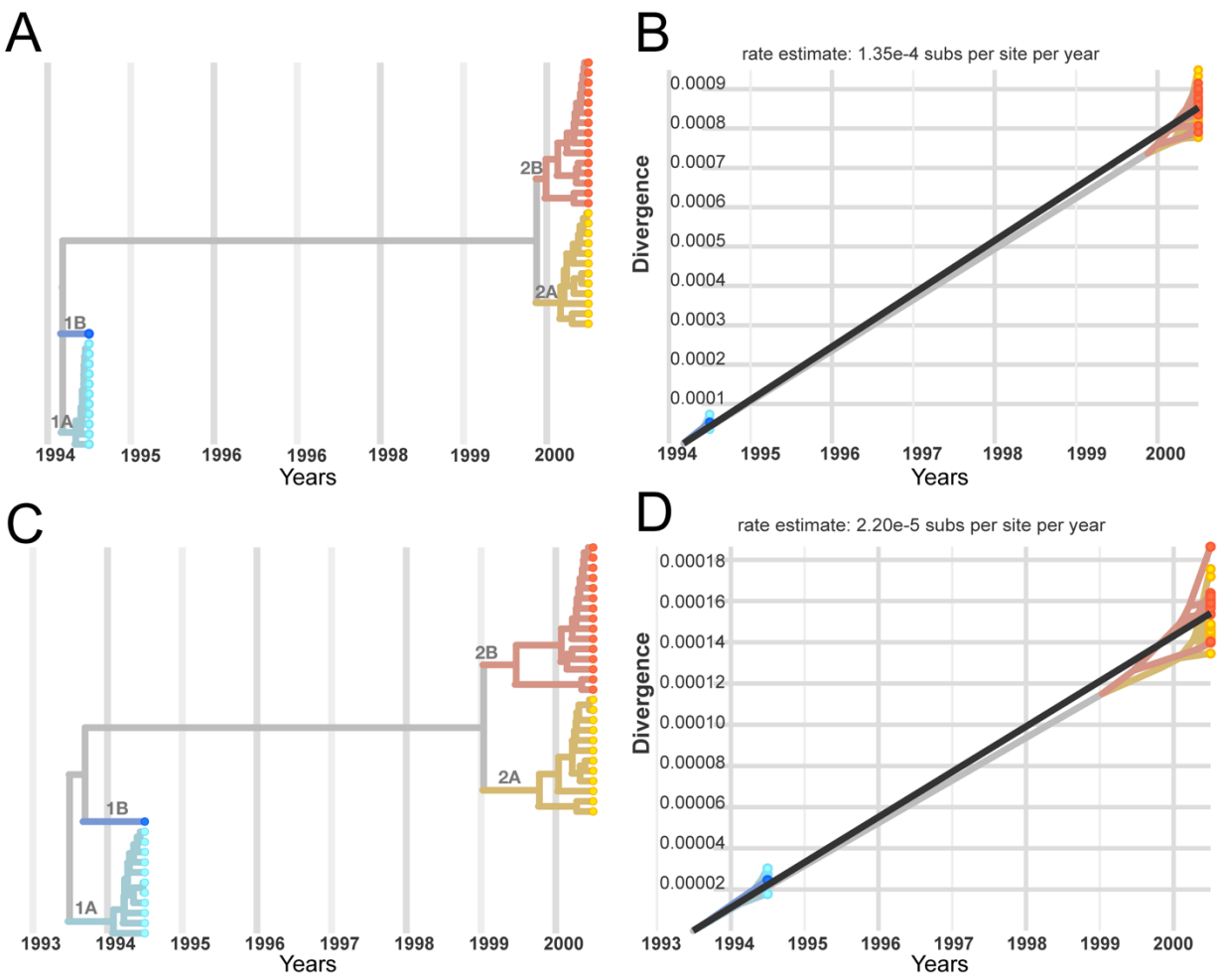


Figure 2.6. Molecular clock rates based on divergence from the ancestral root differ between time-scaled maximum likelihood tree and recombination corrected phylogeny.

X-axis indicates the date in years and Y-axis **(B, D**) indicates divergence (substitutions/site) from the ancestral root. (**A**) A time-scaled version of the maximum likelihood tree with (**B**) molecular clock rates inferred from the root-to-tip distance using TreeTime [101]. (**C-D**) Shows time-scaled tree (**C**) and inferred molecular clock for the recombination corrected tree (**D**).

The inferred molecular clock rate calculated from a time-scaled tree using the root-to-tip distance (1.35×10^{-4} subs/site/year) is similar to the molecular clock rate calculated from pairwise comparisons (Figure 2.2A, Figure 2.6A-B). Expectedly, this rate is much higher than other within-host estimates that exclude predicted

recombination events [61]. To explore the impact of recombination on the tree branch scale and topology, we generated a recombination corrected phylogeny using ClonalFrameML (Figure 2.5B). We found the branch lengths of the maximum likelihood tree (Figure 2.4A, Figure 2.5A) were on average 5 times greater than that of the recombination corrected tree, but general topology remained the same (Figure 2.5B). The inferred molecular clock rate from the recombination corrected time-scaled tree was 2.2×10^{-5} subs/site/year, which is consistent with published estimates (Figure 2.6D) [61,94,98]. We estimated a time to most recent common ancestor (TMRCA) of 1 year for the ancestral population and 1.5 years for the recent population (Figure 2.6C). Low TMRCA are typically observed in phylogenetic studies of within host evolution of *H. pylori*; and in the absence of antibiotic treatment, this is thought to occur as a result of immune pressure [90,94].

We defined four distinct subgroups within the collection based on shared genetic characteristics reflected in the pairwise comparisons, the maximum likelihood tree, and recombination corrected phylogeny (Figure 2.4A-B, Figure 2.5). While the majority of isolates within the ancestral group are highly related, one isolate, SC4, is more divergent and clusters separately on the tree. The average pairwise genetic distance between SC4 and each ancestral isolate is 173 nucleotide differences whereas the average pairwise genetic distance among all other unique pairs of ancestral isolates is 39 nucleotide differences. Thus, we named two subgroups of the ancestral isolates according to this divergence (1A and 1B). Within the recent group, there is additional clustering of the isolates into two subgroups, named 2A and 2B. Group 2A, is comprised

of 12 total isolates with 147 unique polymorphisms (SNPs and indels) and group 2B is comprised of 15 total isolates with 216 unique polymorphisms.

Isolates collected from the most recent time point originated from biopsy samples from distinct stomach regions, allowing us to examine if region specific adaptation drives subgroup formation. To assess this, π for all the pairs isolated from the same source biopsy (within region) was compared to π from all the pairs from different source biopsies (between region). Although we hypothesized that isolates from the same source biopsy would be more similar, we instead found between region pairs have the same level of diversity as within region pairs (Figure 2.4C). We also did not find any specific SNPs or indels associated with isolates from either the antrum or corpus. This suggests subgroup differentiation within time points is not defined by stomach region adaptation in this patient. Consistent with this finding, both recent subgroups (yr 2000) contain isolates from all three biopsy locations and nearest neighbors on the tree often come from different biopsies (Figure 2.4A). The formation of distinct subgroups within a population of isolates collected from a single time point, suggests the possibility of adaptation to niches within host, but these distinct niches must be present in all regions of the stomach sampled.

2.3.4 *Recent H. pylori isolates have increased proinflammatory activity driven by cagY genetic variation*

Substantial genetic divergence of *H. pylori* populations over this six year period of infection, coupled with enrichment of polymorphisms in genes related to virulence, prompted exploration of pathogenic phenotypes. First, we tested ability of each strain to initiate an inflammatory response. Each of the 39 isolates was co-cultured with a gastric

epithelial cell line (AGS) for 24hrs (MOI=10) and the release of inflammatory cytokine interleukin-8 (IL-8) was measured in the supernatants. The J99 ancestral strain and J99 *ΔcagE*, a mutant that blocks assembly of the Cag T4SS, were used as controls in each independent experiment. All isolates were Cag PAI+ and induced IL-8 at levels above J99 *ΔcagE*. However, isolates from the most recent time point on average induced more IL-8 compared to ancestral isolates (Figure 2.7A-B). There was some heterogeneity in this phenotype with the least inflammatory isolates inducing 12% less and the most inflammatory isolates inducing 56% more IL-8 than J99 (Figure 2.7B). Isolates with similar IL-8 phenotypes clustered together on the maximum likelihood tree; and comparison of the genetic and phenotypic distances between pairs of isolates from both time points (n=276) showed genetic divergence correlates with phenotypic divergence in induction of IL-8 (Figure 2.7C). There is also a weaker correlation between genetic and phenotypic distances of within timepoint pairs, indicating genetic divergence within the population at a single time point may also contribute to the phenotypic heterogeneity observed (Figure 2.7D). These data show that isolates able to induce more inflammation persisted, indicating a possible adaptive advantage of pro-inflammatory activity during chronic infection in this patient.

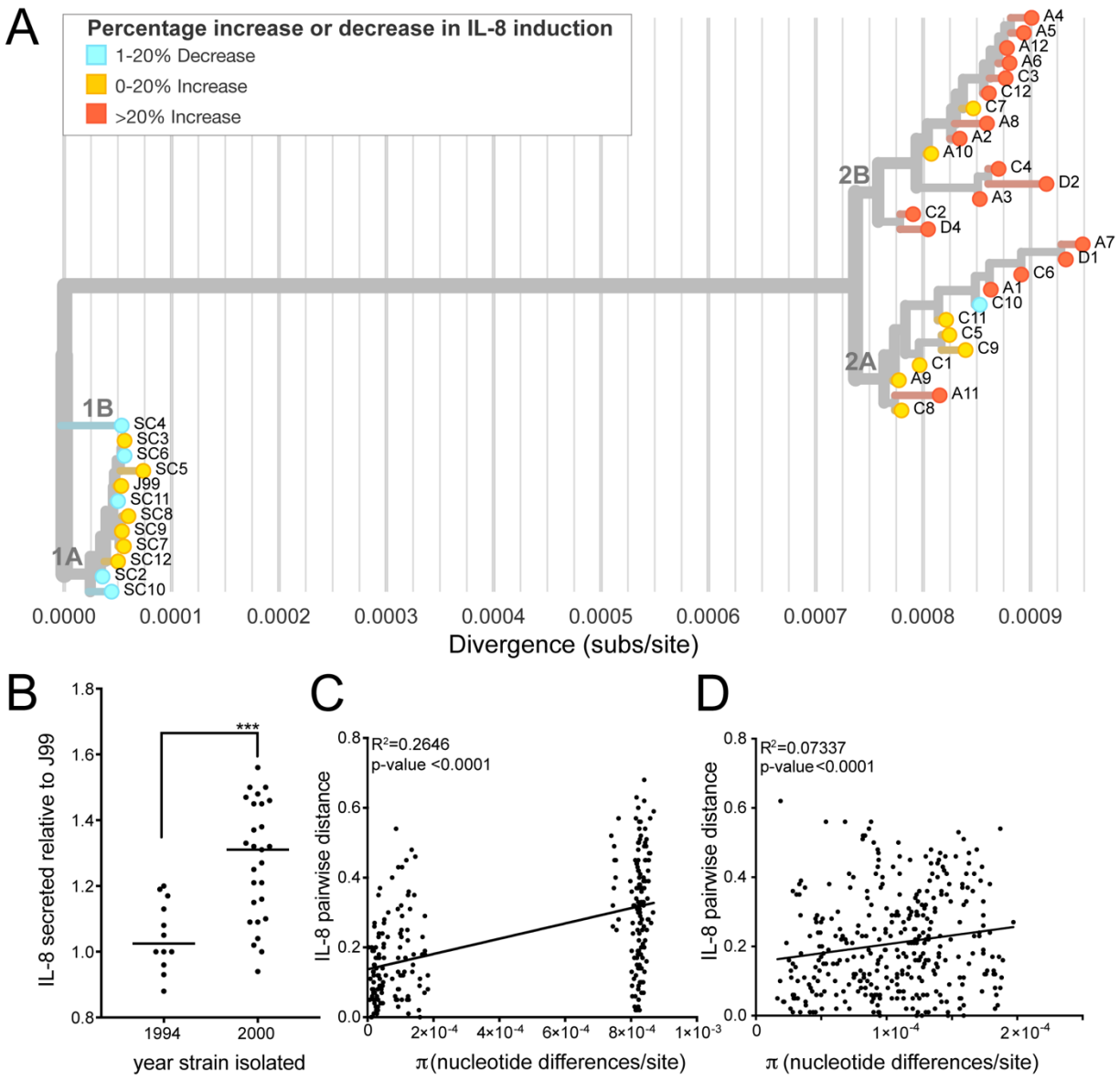


Figure 2.7. Isolates from recent time point have increased induction of IL-8 secretion during co-culture with gastric epithelial cells.

(**A**) The maximum likelihood tree from Figure 2.4A overlaid with IL-8 induction phenotype of each isolate after 24 hours of co-culture (MOI=10) with gastric epithelial cell line (AGS). Leaf colors represent percent increased or decreased IL-8 secretion relative to ancestral isolate J99. (**B**) Each point shows mean value of IL-8 detected in the supernatants of infected AGS cells relative to J99 for independent isolates. The mean value was calculated from at least two experiments with triplicate wells. Black line represents the mean values from each subset of isolates (yr 1994, yr 2000).

Significance was determined using a Student's t-test (****, p<0.0001). (**C-D**) Comparison of genetic (π) and phenotypic (relative IL-8 secreted) distances between unique pairs of isolates from and between time points (**C**) and within the same time point (**D**) is shown. Plot shows a linear regression with p-value derived from F-test and correlation coefficient (R^2) reported.

To investigate the genetic basis of shared IL-8 phenotypes, we focused on nSNPs that occurred within the Cag PAI. While there was no enrichment of nSNPs within the Cag PAI as a whole (chi-squared, p-value=0.3921), we did see enrichment in two Cag PAI genes, *cagY* and *cagA* (Figure 2.8A). Recent isolates had 7 unique nSNPs in *cagA*, however none were localized to known functional domains (Figure 2.8B) [102]. Several nSNPs were detected within the middle repeat region of *cagY*. This domain contains a series of long and short direct repeat sequences that can undergo recombination resulting in expansion or contraction of repeats. This can attenuate or enhance Cag T4SS-dependent IL-8 secretion. In animal models of infection, recombination events that diminish pro-inflammatory activity are dependent on adaptive immunity [82]. Due to the difficulties in precisely mapping these recombination events with short-read WGS data, we utilized restriction fragment length polymorphism (RFLP) together with Sanger sequencing to identify unique alleles of *cagY* within the collection (Figure 2.9A-B, Figure 2.10C). All of the recent isolates (Groups 2A and 2B) and ancestral isolate SC4 (Group 1B) share the same RFLP pattern, which is distinct from the RFLP pattern shared by the other ancestral isolates (Figure 2.9A, Figure 2.10A). Sanger sequencing revealed that all the isolates in group 1A, including J99, share the same sequence. However the allelic variant of *cagY* in SC4 is distinct from that in the recent isolates (Figure 2.10B). The SC4 *cagY* allele carries two nSNPs shared with

recent isolates, including one that introduced a restriction site seen by RFLP, however it also harbors two unique nSNPs not found in any other isolates in the collection. All recent isolates (Group 2A and 2B) have 9 nSNPs total compared to the J99 *cagY* allele including the two shared with SC4 (Figure 2.9C, Figure 2.9C). None of the alleles had expansion or contraction of the number of repeats, but likely arose from gene conversion from sequences within other repeats (Figure 2.9C).

In order to test for a functional link between the variation in *cagY* and the differences in IL-8 phenotype, we performed an allelic exchange experiment, replacing the *cagY* allele in J99 ancestral strain with the two other *cagY* allelic variants (Figure 2.9C). Co-culture of these engineered strains with AGS cells showed that the *cagY* allele shared by the recent isolates confers the increase in induction of IL-8 at 24hrs. The SC4 allele in the J99 genomic context induced similar levels of IL-8 secretion as J99. Therefore modulation of T4SS function can occur through the introduction of specific SNPs in the absence of expansion or contraction of the *cagY* repeats. The same experiment was performed with *cagA* variants, but IL-8 induction was not significantly different from J99 (Figure 2.8C).

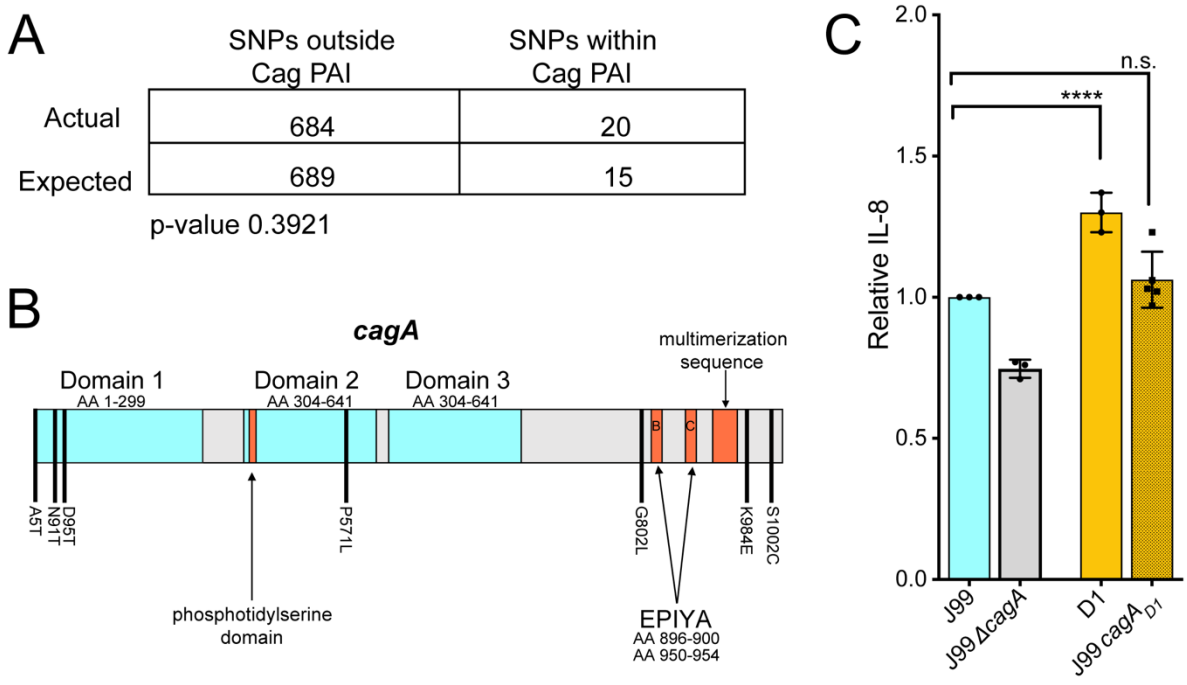


Figure 2.8. Genetic variation in *cagA* does not influence IL-8 induction during AGS cell co-culture.

(**A**) Contingency table of nSNPs falling within and outside Cag PAI compared to expected values based on a normal distribution. Significance was determined using a Fisher's exact test.(**B**) CagA gene schematic labeled with nonsynonymous amino acid changes shared by all recent isolates (black bars). The three protein domains identified in the published crystal structure (blue), including the flexible N-terminal region (Domain 1, amino acids 1-299), the anti-parallel beta sheet (Domain 2, amino acids 304-641), and Domain 3, amino acids 304-641 are labeled. Known host protein interaction motifs including the integrin binding phosphotidylserine domain, phosphotyrosine EPIYA sites, and multimerization sequence are also labeled in orange [103]. (**C**) Levels of IL-8 produced by *cagA* allelic exchange strains relative to J99 24 hrs post infection of AGS cells (MOI=10). Data points represent averaged values from triplicate wells from at least 3 independent biological replicates. Significance was determined with a one-way ANOVA with Dunnett's corrections (n.s., not significant; **** $p<0.0001$).

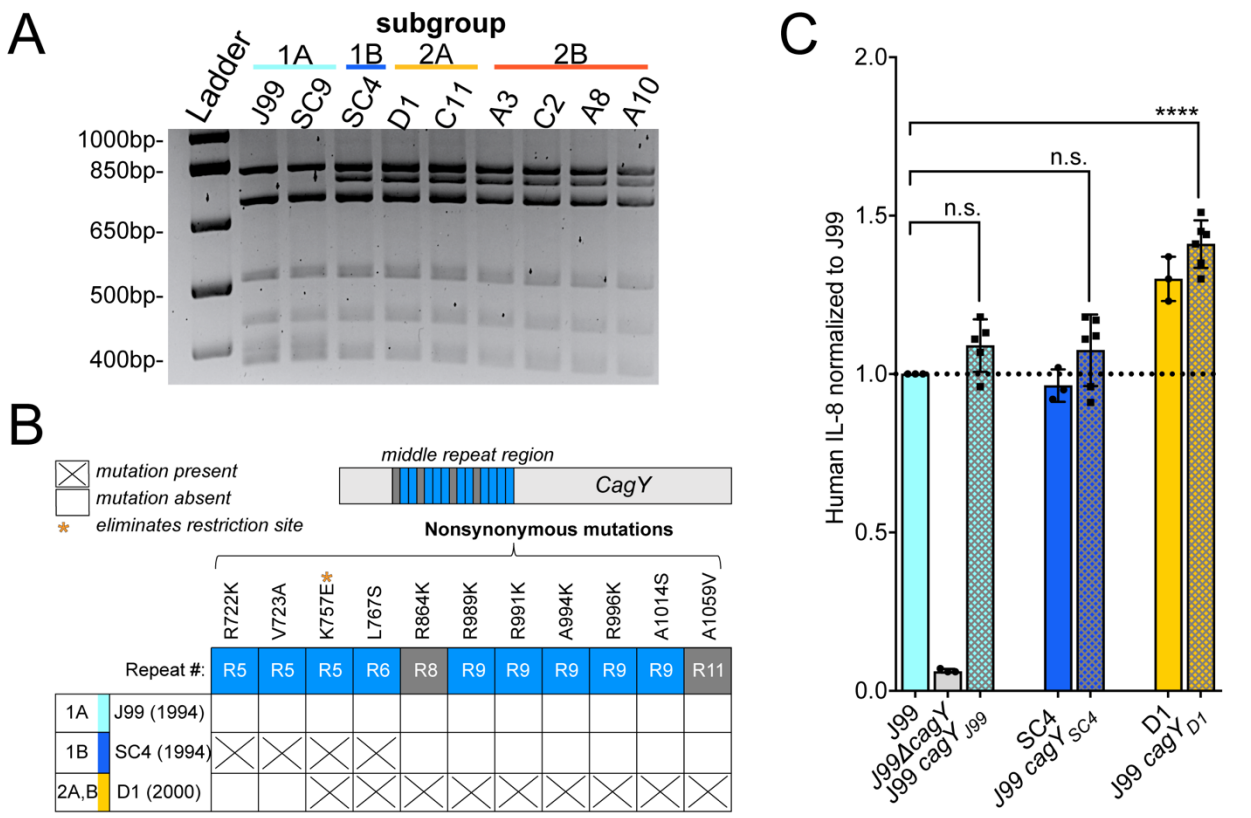
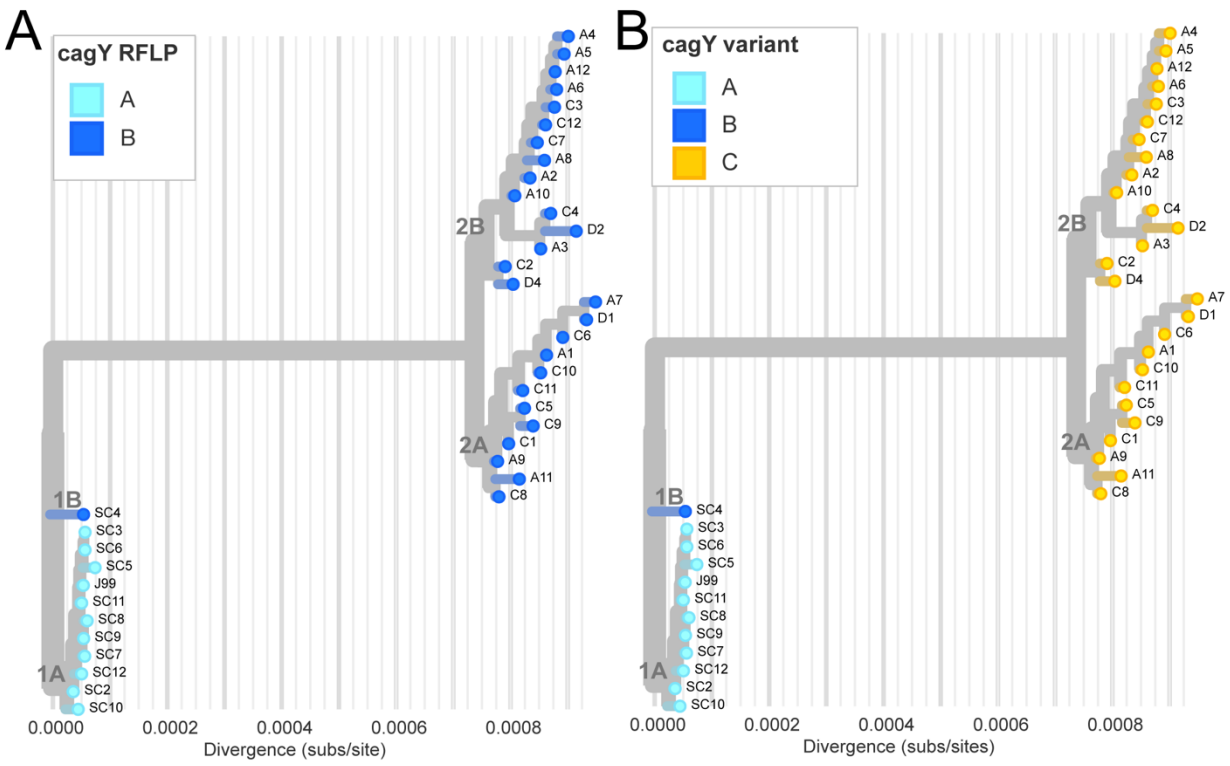


Figure 2.9. *cagY* genetic polymorphisms promote enhanced IL-8 secretion.

(A) RFLP analysis of amplified *cagY* repeat region from representative isolates digested with restriction enzyme (DdeI) reveals two distinct patterns within the populations. Isolates are colored by subgroup as in Figure 2.4A. (**B**) Amino acid polymorphisms for the three different allelic variants of *cagY* in our isolate collection detected by Sanger sequencing. Isolates listed represent the three alleles (J99, SC4, D1) with date of isolation and subgroup(s) indicated. All nSNPs detected map within the middle repeat region of *cagY* (15 total repeats; short in blue and long in gray). (**C**) Levels of IL-8 produced by *cagY* allelic exchange strains relative to J99 ancestral (y-axis line=1) 24hrs post infection of AGS cells (MOI=10). Data points represent averaged values from triplicate wells from at least 3 independent experiments. Significance was determined with a one-way ANOVA with Dunnett's corrections (n.s. not significant, **** $p<0.0001$).



C

```
J99  630 KKECEKLLTPEAKKKLEEAKKSVRAYLDCVSKAKNEAERKECEKLLTPEAKKLLENQALD
SC4  630 KKECEKLLTPEAKKKLEEAKKSVRAYLDCVSKAKNEAERKECEKLLTPEAKKLLENQALD
D1   630 KKECEKLLTPEAKKKLEEAKKSVRAYLDCVSKAKNEAERKECEKLLTPEAKKLLENQALD

J99  690 CLKNAKTDEERKECLKDLPKDLQKKVLAKESVRVYLDCVSKAKNEAERKECEKLLTPEAR
SC4  690 CLKNAKTDEERKECLKDLPKDLQKKVLAKESVKAYLDCVSKAKNEAERKECEKLLTPEAR
D1   690 CLKNAKTDEERKECLKDLPKDLQKKVLAKESVRVYLDCVSKAKNEAERKECEKLLTPEAR

J99  750 KLLEEAKKSVKAYKDCVLRARNEKEKQECEKLLTPEARKLLEESKKSVKAYLDCVSKAKN
SC4  750 KLLEEAKESVKAYKDCVSRARNEKEKQECEKLLTPEARKLLEESKKSVKAYLDCVSKAKN
D1   750 KLLEEAKESVKAYKDCVSRARNEKEKQECEKLLTPEARKLLEESKKSVKAYLDCVSKAKN

J99  810 EAERKECEKLLTPEARKLLEEAKESVKAYKDCVSRARNEKEKQECEKLLTPEARKLLENQ
SC4  810 EAERKECEKLLTPEARKLLEEAKESVKAYKDCVSRARNEKEKQECEKLLTPEAKKLLENQ
D1   810 EAERKECEKLLTPEARKLLEEAKESVKAYKDCVSRARNEKEKQECEKLLTPEAKKLLENQ

J99  870 ALDCLKNAKTEAEKKRCVKDLPKDLQKKVLAKESVRVYLDCVSKAKNEAERKECEKLLTP
SC4  870 ALDCLKNAKTEAEKKRCVKDLPKDLQKKVLAKESVRVYLDCVSKAKNEAERKECEKLLTP
D1   870 ALDCLKNAKTEAEKKRCVKDLPKDLQKKVLAKESVRVYLDCVSRARNEKEKECEKLLTP

J99  930 EARKLLEEAKESVKAYKDCVSRARNEKEKQECEKLLTPEARKLLEQEVKKSVKAYLDCVS
SC4  930 EARKLLEEAKESVKAYKDCVSRARNEKEKQECEKLLTPEARKLLEQEVKKSVKAYLDCVS
D1   930 EARKLLEESKESVKAYKDCVSRARNEKEKQECEKLLTPEARKLLEQEVKKSVKVYLDCVS

J99  990 RARNEKEKQECEKLLTPEARKLLENQALDCLKNAKTEAEKKRCVKDLPKDLQKKVLAKES
SC4  990 RARNEKEKQECEKLLTPEARKLLENQALDCLKNAKTEAEKKRCVKDLPKDLQKKVLAKES
D1   990 RARNEKEKQECEKLLTPEARKLLENQALDCLKNAKTEAEKKRCVKDLPKDLQKKVLAKES
```

Figure 2.10. Three different *cagY* alleles distinguish isolate groups and subgroups.

(**A**) Maximum likelihood tree overlaid with two different *cagY* RFLP subtypes detected with restriction enzyme DdeI. RFLP subtypes, named A and B according to the figure

legend, are shown in Figure 2.9A. (**B**) Maximum likelihood tree overlaid with unique *cagY* alleles detected with Sanger sequencing. Leaf colors correspond to each of the three unique alleles detected and reported in Figure 2.9B. Group 1A shares allele A, group 1B shares allele B, and groups 2A and 2B share allele C. (**C**) Amino acid alignment of multiple repeat regions of three representative *cagY* alleles detected in the collection. J99 represents the allele found in subgroup 1A (allele A), SC4 represents the allele found in subgroup 1B (allele B), and D1 represents the allele found in 2A and 2B (allele C). Polymorphic sites are highlighted with amino acids in blue representing the reference (J99, AE001439) and red indicating a nonsynonymous substitution.

2.3.5 *Sub-populations within the collection have distinct bacterial cell morphologies*

Cell morphology has also been linked to virulence in *H. pylori* [20]. In order to measure cell morphology we used CellTool, a program which takes 2-D phase contrast images and measures quantitative cell shape parameters from cell outlines [20]. Based on these measurements, isolates were divided into three phenotypic shape categories—short pitch, long pitch, and rod. Short pitch isolates have increased wavenumber per unit centerline axis length compared to the long pitch and rod isolates. Rod isolates have decreased side curvature per unit centerline axis length compared to the short and long pitch isolates (Figure 2.11B-C). Isolates with similar shape phenotypes cluster on the maximum likelihood tree (Figure 2.11A, Figure 2.12). Interestingly, all ten rod-shaped isolates from group 2B had loss of function mutations in *H. pylori* cell shape determining (*csd*) genes known to cause rod-shape morphology when deleted. We observed four unique mutations in *csd4*, one unique mutation in *csd5*, and three unique mutations in *csd6*, one of which was shared by three isolates. Thus, group 2B appears to have convergent evolution leading to straight-rod morphology (Figure 2.11D).

Pairwise comparisons show that recent isolates within the same cell shape phenotype category are more genetically similar than recent isolates from different cell shape categories (Figure 2.13A). Additionally, plots of the genetic and phenotypic distances between unique pairs of recent isolates from both time points (n=351) showed that genetic divergence positively correlates with phenotypic divergence both in wavenumber and size curvature per unit axis length (Figure 2.13B-C). This indicates a signature of selection for loss in helical shape within this sub-population of recent isolates.

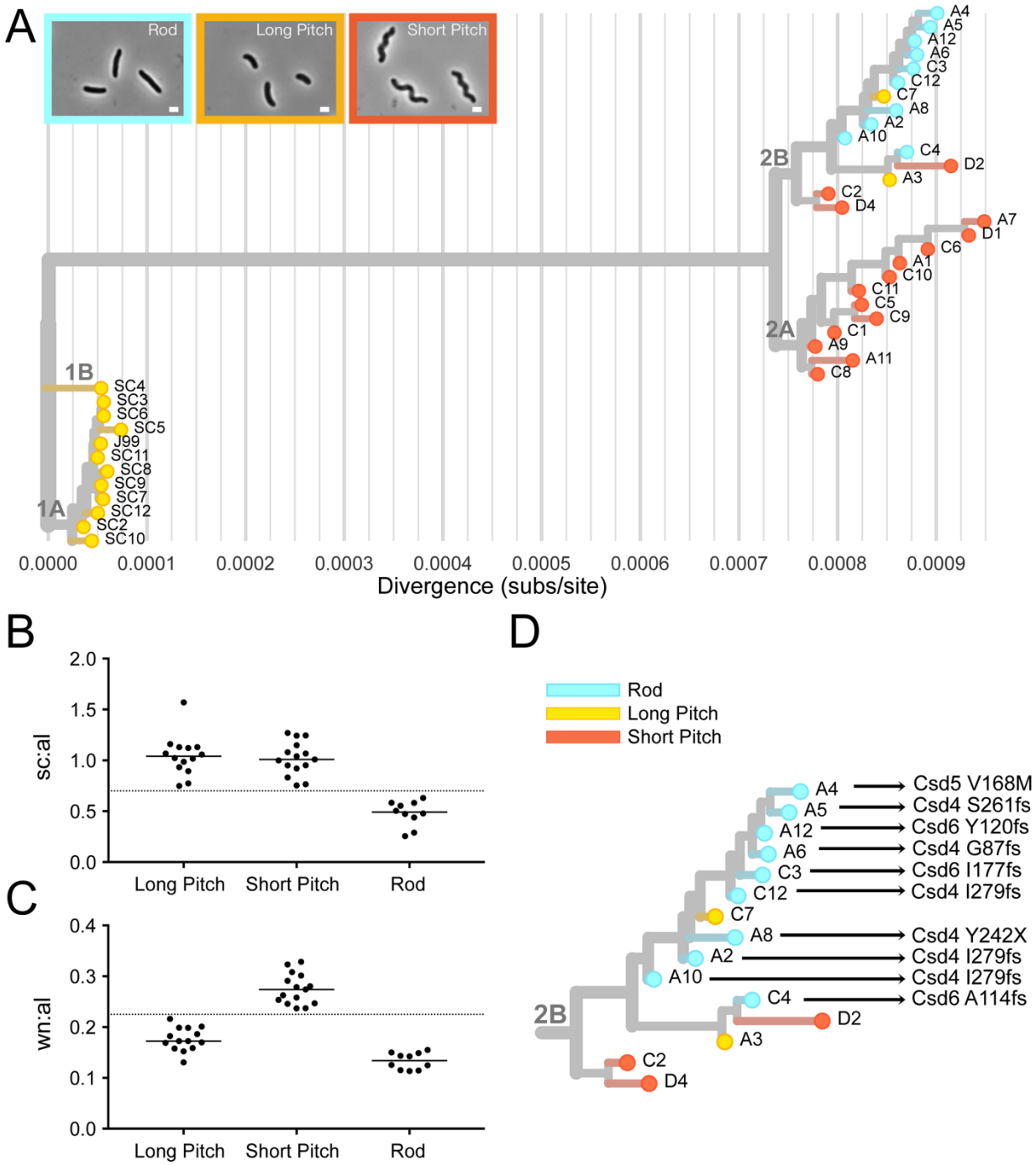


Figure 2.11. Cell morphology varies among genetically distinct subgroups.

(**A**) Cell shape phenotype clustering on maximum likelihood tree. Leaves colored to indicate cell morphology phenotype of each isolate with rods in blue, short pitched in yellow, and long pitched in orange. Representative phase contrast micrographs of

each morphologic class shown. (magnification = 100x, scale bar = 1 μm). (**B-C**) Cell shape parameters calculated from 2-D phase images with CellTool software for isolates with indicated cell morphologies. Individual points represent mean values for measurements taken from >100 cells/isolate. Side curvature and wavenumber values were normalized by cell centerline axis length. (**B**) Mean side curvature values normalized by centerline axis length (sc:al) is decreased in rod shaped cells (<0.7, as indicated by y-axis line) and (**C**) wavenumber normalized by centerline axis length (wn:al) is increased for cells that have increased wavenumber (>0.225, as indicated by y-axis line). **(D)** Subgroup 2B labeled with amino acid mutations in cell shape determining proteins (Csd4, Csd5, Csd6).

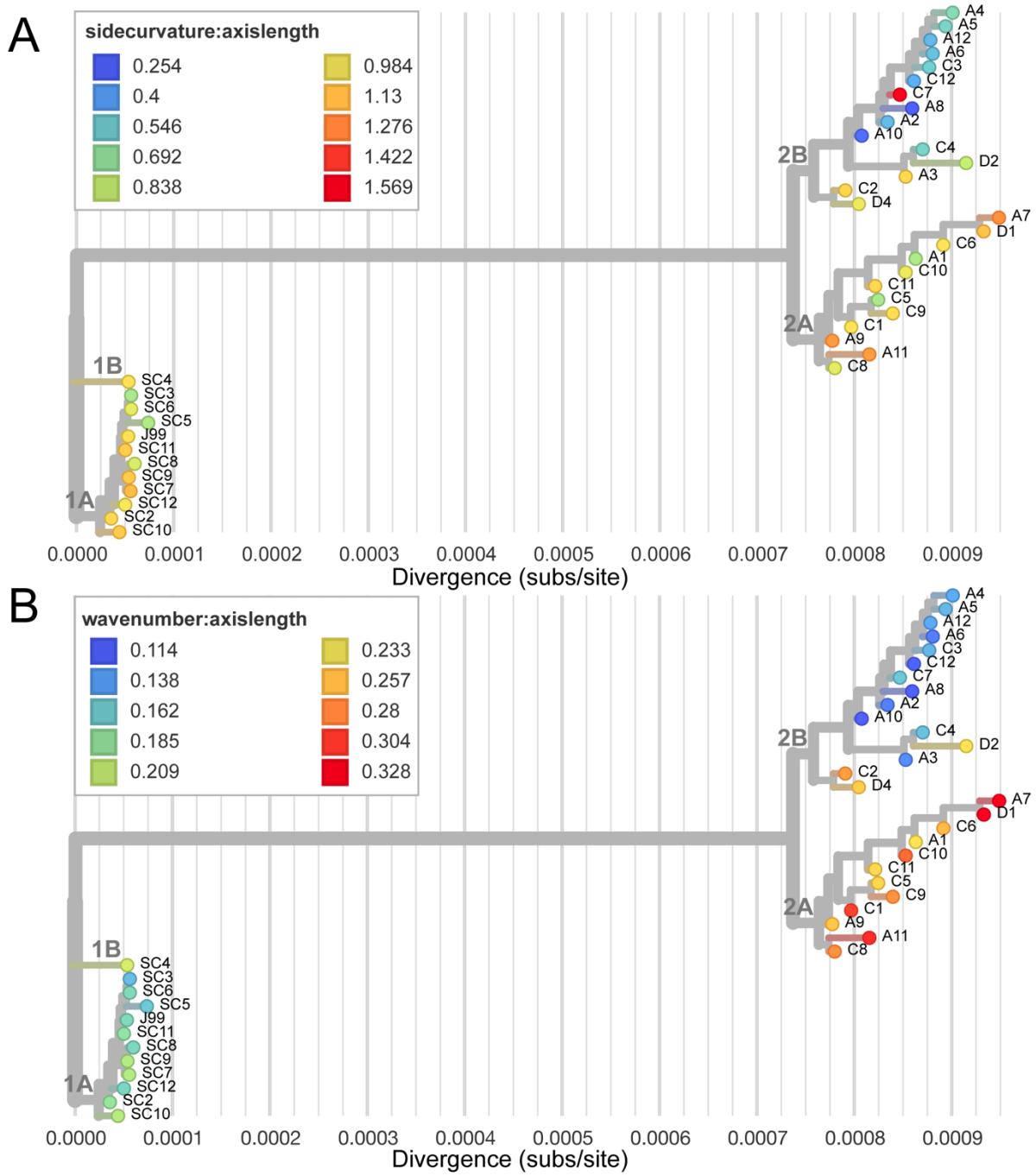


Figure 2.12. Cell shape parameters vary within and between isolate subgroups.

Maximum likelihood tree overlaid with cell shape measurements taken from 2-D phase contrast images using CellTool. Leaf colors represent side curvature normalized by centerline axis length (**A**) or wave number normalized by centerline axis length (**B**) as indicated in the figure legends.

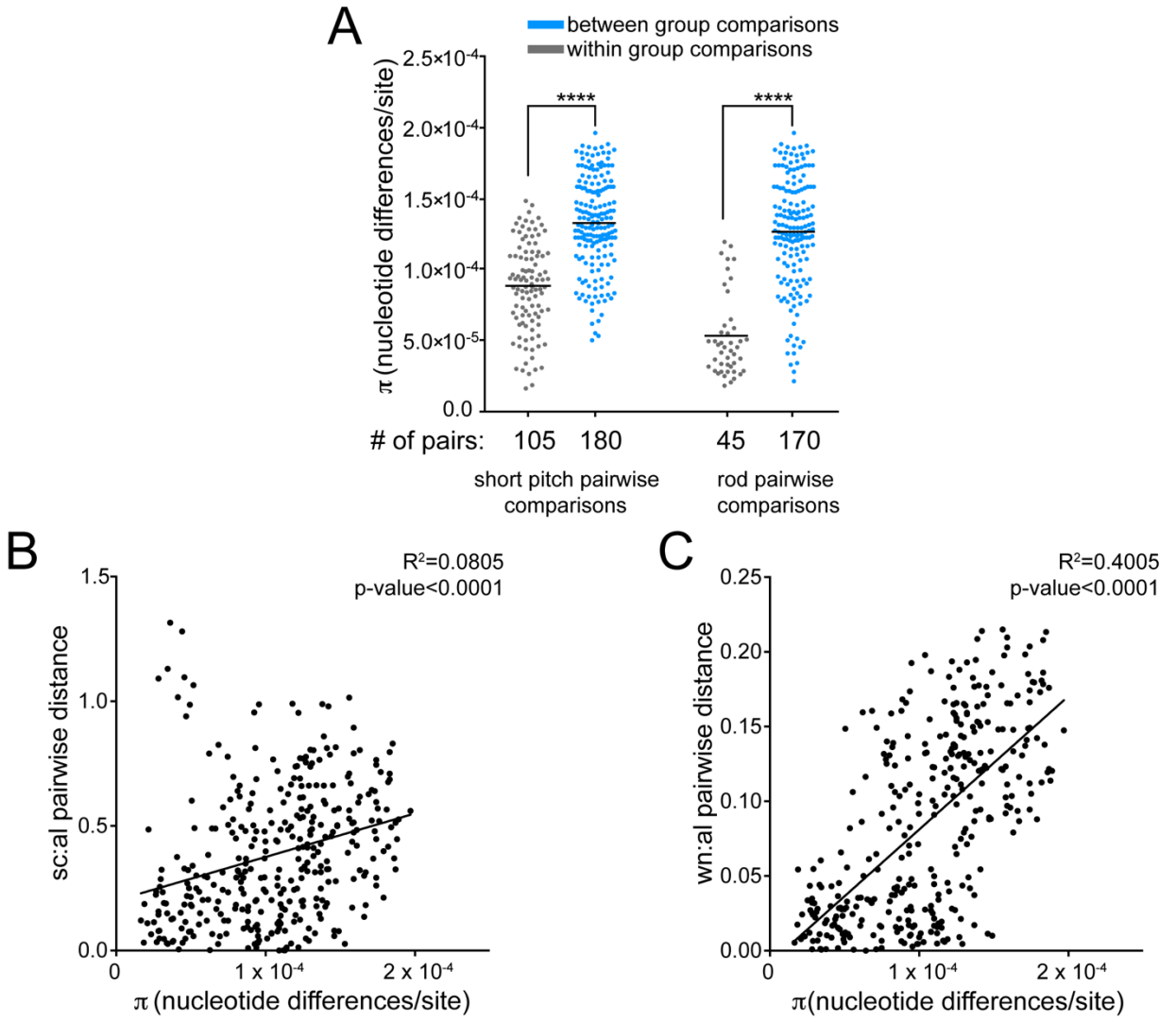


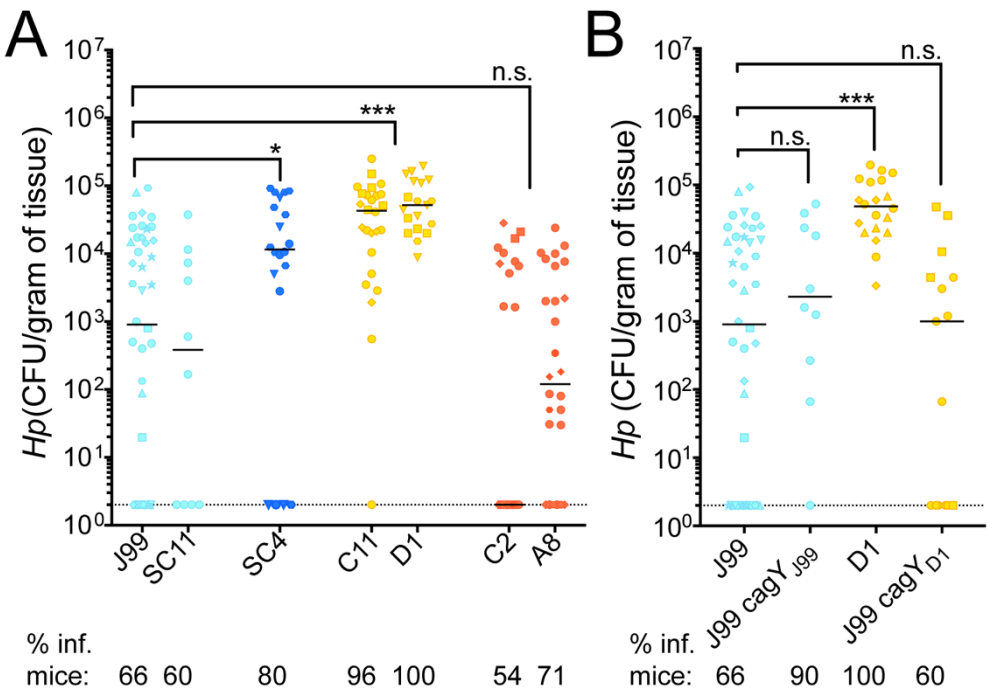
Figure 2.13. Cell morphology parameter divergences correlate with genetic distance during chronic infection.

A) Pairwise comparisons show that recent isolates within the same cell shape phenotype category (gray) are more genetically similar than recent isolates from different cell shape categories (blue). Each point represents a unique pairwise comparison between recent isolates. Midline represents the mean (π statistic). Significance was determined using Student's t-test (**** $p<0.0001$). (**B-C**). Correlation of genetic and phenotypic distances in (**B**) side curvature (sc) and (**C**) wavenumber (wn) per unit centerline axis length (al) between unique pairs of recent isolates (n=351) with each point displaying a unique pairwise comparison. Plot shows a linear regression with p-value derived from F-test and correlation coefficient (R^2) shown.

2.3.6 *Isolates differ in mouse colonization during acute infection*

Considering the observed phenotypic divergence among isolates between and within time points, we hypothesized that individual isolates may behave differently in a mouse stomach colonization model. C57BL/6 mice were infected with representative isolates from each time point and subgroup for 1 week. All isolates tested successfully colonized mice. However, the proportion of mice with detectable infection and loads (CFU/gram of stomach tissue) differed. Almost all the mice infected with the two isolates from group 2A (C11, D1) and the single isolate from the ancestral group 1B (SC4) had higher loads than representative isolates from the other groups and this increase coincided with a greater proportion of mice stably infected after one week compared to the others (Figure 2.14A). Isolates within different cell shape categories and IL-8 profiles were chosen when possible. Although loss of helical shape has been shown to decrease colonization, both helical and rod-shaped isolates from clade 2B infected at lower loads (Figure 2.14C). Since recombination events in *cagY* were detected in all isolates with increased loads, and these changes impacted the inflammatory response in-vitro, we tested to see if our ancestral strain (J99) with the recent variant of *cagY* (J99 $cagY_{D1}$) would also colonize at higher loads. However, J99 $cagY_{D1}$ colonized mice similarly to J99 ancestral (Figure 2.14B), indicating the increased mouse colonization phenotype is not conferred by *cagY* variation. Together these results suggest that there are additional, unknown properties of these isolates contributing to colonization (Figure 2.14C). However, the differences in colonization between groups 2A and 2B and 1A and

1B supports the assertion that subgroup differentiation has phenotypic consequences for infection.



C

Isolate	Subgroup	Shape Phenotype	IL-8 Phenotype	Median CFU/gram in Mouse
J99	1A	Long Pitch	0-20% increase	$9.0x10^2$
SC11	1A	Long Pitch	1-20% decrease	$3.8x10^2$
SC4	1B	Long Pitch	1-20% decrease	$1.2x10^4$
C11	2A	Short Pitch	0-20% increase	$4.3x10^4$
D1	2A	Short Pitch	>20% increase	$5.2x10^4$
C2	2B	Short Pitch	>20% increase	$2.0x10^0$
A8	2B	Rod	>20% increase	$1.2x10^2$

Figure 2.14. Mouse colonization among isolates from distinct subgroups differs and is not explained by *cagY* variation.

(**A-B**) Each point represents the colony forming units per gram of mouse tissue homogenate from a single mouse. Two CFU/gram is the limit of detection (dotted line). Biological replicates indicated by different symbol shapes with the median shown. P-values were calculated from pooled experimental replicates using a Mann-Whitney non-

parametric test (n.s. not significant, * $p<0.05$, *** $p<0.001$). Percentage of mice with bacterial colonization above the limit of detection (% inf. mice) is indicated below the isolate names. (**A)** Colonization in WT B6 mice 1 week post infection with representative isolates from each subgroup. Color indicates maximum likelihood tree subgroup 1A (light blue), 1B (dark blue), 2A (yellow), and 2B (red). (**B**) Alleles of *cagY* from indicated strains were engineered into strain J99 at the native locus and resultant isolates used for infection experiments. Data points in blue are mice infected with J99 variant of *cagY* and points colored in yellow were infected with D1 variant of *cagY*. (**C**) Summary of representative isolate phenotypic characteristics including the subgroup, bacterial shape, IL-8 phenotype (relative IL-8), and median mouse colonization (CFU/gram).

2.4 DISCUSSION

Within-host *H. pylori* isolates from a single individual, once thought to be homogenous, have since been shown to be genetically distinct. Next-generation sequencing provides tools to examine the breadth of diversity present, however little is known about how this diversity contributes to pathogenesis and disease progression. Our study characterized both genetic and phenotypic diversity of infecting populations from a single, chronically infected individual at two time points over a six-year period. We re-sequenced isolate J99, originally isolated from the patient in 1994 and detected 755 polymorphisms differing from the published reference (GenBank, AE001439). The majority (553 polymorphisms) were shared across all isolates, and were therefore excluded from the analysis. These polymorphisms reflect either improvements to sequencing methods since the original sequence was published in 1999, differences in

the J99 stock used in our experiment compared to that of Alm et al. (AE001439), or misalignments of short read data to the reference.

Previous studies have found within host evolution of *H. pylori* is shaped both by de-novo mutation and homologous recombination events, with recombination events generating the majority of the overall diversity [104]. Polymorphism introduced by both mechanisms can impact phenotype, therefore with our analysis we reported and explored sequence level diversity acquired by both mechanisms. Others have estimated the within-host mutation rate to be between 6.5-0.5 x 10^{-5} substitutions per site per year by excluding predicted recombination sites in order to make evolutionary inferences [61,89,94,98]. However, it remains debated whether diversity generated via recombination versus mutation can be accurately identified and filtered to reconstruct evolutionary relationships [105]. Since clustered nucleotide polymorphisms (CNPs), typically excluded from molecular clock rate calculations, were included in this analysis, our overall within-host molecular clock rate of 1.3 x 10^{-4} subs/site/yr is elevated compared with rates based on substitutions from mutation alone [61]. However, when we excluded polymorphism introduced via importation events using ClonalFrameML, we calculated a molecular clock rate of 2.2 x 10^{-5} subs/site/yr, consistent with the other published estimates [61,94]. This analysis also upholds a model of within host diversity arising predominantly from recombination.

We found that in this individual, with no known exposure to antibiotics, infecting populations increased diversity over time and clustered into genetically distinct subgroups, suggesting adaptation to specific host niches [62]. The high number of polymorphic sites within OMPs, including *hopQ*, *sabA*, *babA*, and *sabB*, were predicted

to have accumulated from recombination events and likely originated from a genetically related co-infecting strain, which was either not sampled or not present in the original population. High variation in these genes is unsurprising considering paralogous OMPs have been shown to diversify within-host via frequent intra and intergenomic recombination [88]. Accumulation of nSNPs in OMPs supports a model of adaptation driven by interactions with the host environment. In addition to OMPs, many genes that accumulated an excess of nSNPs had unknown function, providing additional targets to examine in future studies.

In other chronic infections, such as *Pseudomonas aeruginosa* infection in cystic fibrosis patients, the emergence of sub-populations is driven by region specific adaptation within distinct anatomical regions of the lung [106]. Evidence of anatomical stomach region specific adaptation in *H. pylori* infections is limited, but it appears to occur in only a small subset of patients [90].These signatures may be obfuscated by frequent population mixing and migration or deterioration of structured niches due to loss of acid production and other tissue changes [34].

While our data do not support subgroup divergence by anatomic region in this individual, the selective pressures at play appear to correspond to known pathogenicity phenotypes. Nonsynonymous polymorphisms detected only within the recent population (yr 2000), fell within known virulence genes, including OMPs involved in host adhesion and Cag PAI-associated genes and distinct alleles within Cag PAI were confirmed by Sanger sequencing. Differences in IL-8 secretion, bacterial cell morphology, and ability to colonize a mouse in an acute infection model were discovered among isolates,

suggesting subgroup divergence driven by tissue features that vary in the stomach across multiple anatomic locations.

Recombination within the middle repeat region of *cagY,* resulting in expansion or contraction of repeats, occurs frequently in short-term animal infections and in humans [82]. We observed modulation of IL-8 induction mediated by T4SS function through mutation and/or recombination without expansion or contraction of *cagY* repeats. The finding that isolates at later time points were more pro-inflammatory was surprising considering inflammation is thought to limit bacterial burden. However, *H. pylori* persists despite relatively high levels of inflammation, so it is possible this feature may be exploited during chronic infection in order to reduce competition for host resources by members of the microbiota [107].

Isolates within this collection also had differences in cell morphology. Morphology differences have been observed among strains from different individuals [108]; here we find that *H. pylori* morphologies differ among isolates from a single patient. In subgroup 2B, we discovered convergent loss of helical cell shape through multiple unique frameshift mutations in cell shape determining (*csd*) genes. Rod-shape isolates have previously been shown to have a colonization deficit manifest at early time points, but to recover during 1-3 months of chronic infection in mice [21]. Due to clustering of rod shapes in subgroup 2B, we suspect that helical shape, while important for early infection and transmission, may be detrimental at later stages of human infection or in particular stomach niches. Among isolates that retained helical shape, we detected more subtle differences in helical pitch, but it is unknown what genetic determinants are responsible or if these differences have direct impacts on colonization.

The observed differences in mouse colonization between isolates from each sub-population supports our initial hypothesis that there are functional consequences of sub-population divergence. Typically, clinical isolates infect mice poorly as mice are not natural hosts for *H. pylori*. However, a few clinical isolates have the intrinsic ability to colonize and can become more robust via serial passage in the mouse stomach [109]. Bacterial properties, including chemotaxis, cell shape, and activity of the Cag PAI, impact mouse colonization and are likely important in establishing human infections [110]. In our collection, there was heterogeneity in mouse colonization among isolates that corresponded to subgroup. Robust colonizers may be more likely to be involved in person-to-person transmission in humans, but it is also possible that these strains may behave differently in other animal models or human hosts.

Increases in colonization potential of representative isolates within 1B and 2B sub-populations does not correlate with differences in morphology or IL-8 phenotypes, indicating an additional unknown factor or combination of factors is responsible for conferring a colonization advantage. Further exploration of the genetic basis for mouse colonization advantage using the subgroup specific variants defined in this study may give new clues to the complex selective forces operant during chronic stomach colonization by *H. pylori*.

2.5 METHODS

2.5.1 *Growth and isolation of H. pylori*

In the initial sampling, a total of 43 *H. pylori* isolates (13 (antral, 1994), 5 (duodenum, 2000), 12 (corpus, 2000), 1 (cardia), and 12 (antral, 2000)) were collected from biopsy

samples from two separate upper gastrointestinal endoscopies performed in a single 48-yr old Caucasian male (1994) residing in Tennessee and treated at the Nashville VA Medical Center. A total of 39 isolates with sufficient sequence coverage (30x) were analyzed in this study (Figure 2.1). *H. pylori* isolates were grown on solid media, horse blood agar (HB agar) or shaking liquid cultures. HB agar plates contain 4% Columbia agar base (Oxoid, Hampshire, UK), 5% defibrinated horse blood (Hemostat Labs, Dixon, CA), 10 mg/ml vancomycin (Thermo Fisher Scientific, Waltham, MA), 2.5 U/ml polymyxin B (Sigma-Aldrich, St.Louis, MO), 8 mg/ml amphotericin B (Sigma-Aldrich), and 0.2% β-cylodextrin (Thermo Fisher). For HB agar plates used to grow *H. pylori* from homogenized mouse stomach, 5 mg/ml cefsulodin (Thermo Fisher), 5 mg/ml trimethoprim (Sigma) and 0.2mg/uL of Bacitracin (Acros Organics, Fisher) are added to prevent outgrowth of mouse microflora. Shaking liquid cultures were grown in brucella broth (Thermo Fisher Scientific, Waltham, MA) supplemented with 10% heat inactivated FBS (Gemini BioProducts, West Sacramento, CA). Plate and flasks were incubated at 37°C under micro-aerobic conditions in 10% CO2, 10% O2, 80% N2, as previously described [111]. For resistance marker selection, HB agar plates were supplemented with 15 μg/ml chloramphenicol, or 30 mg/ml sucrose, as appropriate.

2.5.2 *DNA extraction, genome sequencing*

Genomic DNA from each isolate to be sequenced was purified using the Wizard Genomic DNA Purification Kit (Promega, Fitchburg, WI) and libraries were constructed and indexed using NexteraRTM DNA Library Prep Kit (Illumina, San Diego, CA) and NexternaRTM Index Kit (Illumina). All cultured isolates (n=43) were sequenced on an Illumina MiSeq instrument in the Fred Hutchinson Genomics Shared Resource. Using

J99 ancestral as the reference strain (AE001439), variants were called from raw paired end reads using the Breseq v0.35.0 software with default parameters and SNPs were further validated using default Samtools software suite [112]. Four isolates with average coverage below 30x were dropped from the analysis. Of the remaining 39 isolates, 34 had coverage >100x. Additional sequencing metrics can be found in Table 2.3. Short read fastq sequence files from the remaining 39 isolates in this study are publicly available on NCBI SRA database (BioProject accession: PRJNA622860, <https://www.ncbi.nlm.nih.gov/sra/PRJNA622860>).

2.5.3 *Enrichment of nonsynonymous SNPs in genes and functional gene classes*

Gene annotations were made using the available Genbank file available for J99 (AE001439) with some manually added annotations of OMPs. All annotation files are available at <https://github.com/salama-lab/Hp-J99>. For identification of genes with excess accumulation of nSNPs and the number of alleles in each population, all unique nSNPs accumulated over the six years reported in Table 2.1 were used. Z-scores were calculated using number of counts per gene normalized according to gene length. Genes with nSNP accumulation greater or equal to four standard deviations from the mean are listed in Table 2.2. To identify enrichment of nSNPs within functional gene-sets, each of the 1,495 genes were annotated with designations in the Microbial Genome Database (MGDB, http://mbgd.genome.ad.jp/) [95]. A Fisher's exact test was used to identify MGDB gene class categories with enrichment or depletion of nSNPs. The number of nSNPs falling within certain MGDB categories were compared to expected values based on a normal distribution and p-values were corrected for multiple

testing using Benjamini and Hochberg false discovery rate methods [113]. Adjusted p-values <0.03 were considered statistically significant. To identify markers of antimicrobial resistance, polymorphisms detected in the collection were screened against a list of annotated antimicrobial resistance genes downloaded from PATRIC <https://www.patricbrc.org>.

2.5.4 *Pairwise distance calculations*

The number of nucleotide differences per site (genetic distance) between pairs of isolates was calculated using PopGenome (R, v 4.0.2) using the merged sample variant call file (VCF) file available on Github (nucleotide diversity, π, [114], PopGenome, [115]). All sites that did not align to the reference genome, J99, or had read depth <5 were excluded from the analysis, which removed missing or ambiguous sites resulting from structural variation or mis-mapping. The unique number of shared sites for each pair was calculated using the BEDtools intersect function [116]. The total number of nucleotide differences for each pair was derived from the merged sample VCF file with low quality sites filtered according to the read depth parameters above. All detected indels were excluded from this analysis to avoid inflation of genetic distance due to alignment errors within highly repetitive regions [117]. The statistical significance of differences between groups was assessed using Student's t-test as indicated in figure legends. The merged sample VCF file is available on the Github page <https://github.com/salama-lab/Hp-J99/blob/master/data/Hp_J99_032020.vcf.gz>.

2.5.5 *Construction of the maximum likelihood and recombination-corrected trees*

SNPs detected across each isolate reported in the merged sample VCF were imported into the reference J99 (AE001439) to create a consensus genome for each of the 39 isolates. The maximum-likelihood tree (Figure 2.5A, Figure 2.7A and Figure 2.11A) was built in IQtree implemented in Nextstrain v 1.8.1, using a general time reversible (GTR) substitution model. The ancestral root was inferred using joint maximum likelihood algorithm. All datasets, config files, documentation, and scripts used in this analysis or to generate figures are available publicly on our Github page <https://github.com/salama-lab/Hp-J99>. The interactive tree can be found at the Nextstrain community site <https://nextstrain.org/community/salama-lab/Hp-J99>. To create the recombination corrected tree with rescaled branch lengths, the reference-aligned consensus genomes were aligned using progressiveMauve assuming collinear genomes and a new rescaled tree (Figure 2.5B) was generated using ClonalFrameML using all sites in the reference (AE001439) [98,118]. Time scaled trees and molecular clock rates (Figure 2.5) were generated with TreeTime implemented within Nextstrain, which approximates divergence times and molecular clock rates using maximum likelihood methods [101]. Branch lengths of the maximum likelihood and recombination corrected trees were compared using the ape and adephylo packages in R, v 4.0.2.

2.5.6 *Sequencing and PCR-RFLP of cagY middle repeat region*

The *cagY* sequences were determined using Sanger sequencing using primers listed in Table 2.4 and PCR-RFLP as previously described [82]. Flanking primers were used to amplify the c*agY* repeat region from every isolate. Amplicons were purified with

QIAquick PCR purification kit according to the instructions from the manufacturer (Qiagen, MD) and digested with restriction enzyme DdeI (New England Biolabs, Ipswich, MA). Digested amplicons were run on a 3% agarose for visualization after ethidium bromide staining.

2.5.7 *Construction of H. pylori mutants*

Six J99 mutants were constructed (J99Δ*cagY*, J99 $cagY_{D1}$, J99 $cagY_{SC4}$, J99 Δ*cagA*, J99 $cagA_{D1}$, J99 Δ*cagE*) and are listed in Table 2.5. Isogenic knockout mutants, J99 Δ*cagY* and J99 Δ*cagA*, were constructed using a vector-free allelic replacement strategy. Upstream and downstream genomic regions flanking the gene were amplified and ligated to a *catsacB* cassette, which confers mutants both chloramphenicol resistant (*cat*) and sucrose sensitivity (*sacB*). Positive clones were selected with 15 μg/ml chloramphenicol, as previously described [119,120]. We integrated variant alleles of the deleted gene at the native locus using sucrose counter selection. All mutants were validated via diagnostic PCR and Sanger sequence. Primers used for generating *H. pylori* mutants are listed in Table 2.4.

2.5.8 *H. pylori co-culture experiment and IL-8 Detection*

AGS cells, from a human gastric adenocarcinoma cell line (ATCC CRL-1739), were grown in Dulbecco's modified Eagle's medium (DMEM) (Thermo-Fisher) supplemented with 10% heat-inactivated FBS (Gemini-Benchmark). For co-culture with *H. pylori*, AGS cells were seeded at 1×10^5 cells/well in 24-well plates 16h prior to infection. The day of infection, medium was removed from AGS cells and mid-log-phase (optical density at

600 nm (OD) 0.3-0.6) *H. pylori* resuspended in DMEM–10% FBS–20% Brucella broth was added at multiplicity of infection of 10:1. Supernatants from triplicate wells of each condition were collected at 24 hrs and assayed for the IL-8 concentration using a human IL-8 enzyme-linked immunosorbent assay (ELISA) kit according to the instructions of the manufacturer (BioLegend, San Diego, CA). IL-8 values were reported as normalized values defined as a proportion increased or decreased compared to values obtained for J99, which was included in each experimental replicate. P-values were calculated from at least two biological replicates where each isolate was assayed with triplicate wells using a one-way ANOVA with Dunnett's corrections.

2.5.9 *Analysis of cell morphology*

Phase contrast microscopy and quantitative analysis using CellTool software package was performed as previously described [20]. Bacterial cell masks were generated through thresholding function in ImageJ. Average side curvature, wavenumber, and centerline axis length were derived from thresholded images of bacteria (>100 cells/strain) using the CellTool software package. Average parameters were then used to calculate side curvature or wavenumber to centerline axis length ratios for each isolate.

2.5.10 *Mouse colonization*

Female C57BL/6 mice 24–28 days old were obtained from Jackson Laboratories and certified free of endogenous *Helicobacter* by the vendor. The mice were housed in sterilized microisolator cages with irradiated rodent chow, autoclaved corn cob bedding, and acidified, reverse-osmosis purified water. All mouse colonization experiments were

performed exactly as described [111]. The inoculum for each infection was 5 x 10^7 cells. After excision, the forestomach was removed and opened along the lesser curvature. Stomachs were divided in equal halves containing both antral and corpus regions and half stomachs were place in 0.5 mL of sterile BB10 media, weighed, and homogenized. Serial homogenate dilutions were plated on nonselective HB plates. After 5-9 days in tri-gas incubator, colony forming units (CFU) were enumerated and reported as CFU per gram of stomach tissue. P-values were calculated from pooled experimental replicates using a Mann-Whitney non-parametric test.

2.5.11 *Statistical analysis*

Statistical analyses were performed according to test specified above and in each figure legend using Prism v7 software (GraphPad) or R v4.0.2. P-values less than or equal to 0.05 were considered statistically significant and are marked with asterisks (*, $p<0.05$, **, $p<0.01$; ***, $p<0.001$; ****, $p<0.0001$; n.s., not significant).

Table 2.3. Short read sequencing metrics and BioSample IDs for the NCBI Sequence Read Archive (SRA).

Isolate	ref_accession	average_coverage[a]	seq_platform	db	BioProject_ID	BioSample_ID
A1	AE001439	167.0074	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531033
A2	AE001439	140.5104	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531034
A3	AE001439	188.954	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531035
A4	AE001439	105.9029	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531036
A5	AE001439	172.2794	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531037
A6	AE001439	152.7753	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531038
A7	AE001439	47.47044	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531039
A8	AE001439	106.284	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531040
A9	AE001439	178.5647	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531041
A10	AE001439	155.0792	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531042

A11	AE001439	95.71254	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531043
A12	AE001439	164.957	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531044
C1	AE001439	179.6071	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531045
C2	AE001439	175.8039	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531046
C3	AE001439	237.7504	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531047
C4	AE001439	119.2664	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531048
C5	AE001439	189.7908	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531049
C6	AE001439	115.9731	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531050
C7	AE001439	216.4329	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531051
C8	AE001439	187.6655	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531052
C9	AE001439	111.9965	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531053
C10	AE001439	249.5923	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531054
C11	AE001439	213.202	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531055
C12	AE001439	239.9674	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531056
J99	AE001439	172.0559	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531021
SC1	AE001439	3.367336	Illumina_MiSeq			
SC2	AE001439	91.46217	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531022
SC3	AE001439	254.4908	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531023
SC4	AE001439	119.908	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531024
SC5	AE001439	194.7596	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531025
SC6	AE001439	195.4864	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531026
SC7	AE001439	159.0469	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531027
SC8	AE001439	240.8298	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531028
SC9	AE001439	137.6455	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531029
SC10	AE001439	109.9515	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531030
SC11	AE001439	122.2275	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531031
SC12	AE001439	79.66712	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531032
Ca-1	AE001439	4.434502	Illumina_MiSeq			
D1	AE001439	86.96214	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531057
D2	AE001439	125.3843	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531058
D3	AE001439	7.874553	Illumina_MiSeq			
D4	AE001439	131.0103	Illumina_MiSeq	Genbank	PRJNA622860	SAMN14531059
D5	AE001439	14.22524	Illumina_MiSeq			

[a]Average read depth coverage for each genome is reported with isolates below <30 highlighted in red.

Table 2.4. Primers used in this study

Primer name	Sequence (5'-3')	Target/Use
CatsacB_F	GATATAGATTGAAAAGTGGAT	amplify catsacB cassette
CatsacB_R	GGATATCGGCATTTTCTTTTG	amplify catsacB cassette
CagA_F_UP	TGGTAAGCATGACACACACAAA	ampliy *cagA* upstream flanking region

CagA_R_UP_stitch	ATCCACTTTTCAATCTATATCTTGTTTCTCCTTCTATATC TAAATT	ampliy *cagA* upstream flanking region, homology to catsacB cassette
CagA_F_DN_stitch	CAAAAGAAAATGCCGATATCCGATTAAGGAATATCAAAA ACGCAA	ampliy *cagA* upstream flanking region, homology to catsacB cassette
CagA_R_DN	GGCATTATAAGGTGTGATAAT	ampliy cagA downstream flanking region
CagA_F	CTAGCCTTTAGACGCCTGCA	internal *cagA* sequencing primers
CagA_R	AGCGCGATCGATTTGCTCTA	internal *cagA* sequencing primers
CagY_F	AGCAAAGGCGCTAGAGACCCA	ampliy cagY upstream flanking region
CagY UP_R stitch	ATCCACTTTTCAATCTATATCGTCTTAACGCCTTTTTATT TAT	ampliy cagY upstream flanking region
CagY DN_F_stitch	GCAAAAGAAAATGCCGATATCCTCATGTTATTAAACTCG GTTCTT	ampliy cagY downstream flanking region
CagY DN_R	ATTCTTTTGCATATTGATTCTTAACTC	ampliy cagY downtream flanking region
CagY_MMR_F	GATATAGATTGAAAAGTGGAT	amplify cagY middle repeat region for RFLP
CagY_MMR_R	TGGTAAGCATGACACACACAAA	amplify cagY middle repeat region for RFLP
CagY_seq1	CATTATTGGCGGCATTGTAGTCGCT	internal cagY sequencing primers
CagY_seq2	ACTGAACCAACAAAAAGTTCAAGTGGC	internal cagY sequencing primers
CagY_seq3	GTTAGAAGAAGCCAAAAAGAGCGTTAGAGC	internal cagY sequencing primers
CagY_seq4	ATTTGAGCTTCTCTTCATCGCTCAAACCAT	internal cagY sequencing primers
CagY_seq5	ATGCGAGAAGTTGCTCACGCCT	internal cagY sequencing primers

Table 2.5. Strains used in this study.

Strains	Description	Reference
J99 ΔcagY	J99 with catsacB cassette at *cagY* native locus	This work
J99 cagY$_{D1}$	J99 with D1 *cagY* allele at native locus	This work
J99 cagY$_{SC4}$	J99 with SC4 *cagY* allele at native locus	This work
J99 ΔcagA	J99 with catsacB cassette at *cagA* native locus	This work
J99 cagA$_{D1}$	J99 with D1 *cagA* allele at native locus	This work
J99 ΔcagE	J99 with catsacB cassette at *cagE* native locus	This work

Chapter 3. *HELICOBACTER PYLORI* MOTILITY VARIES WITH CELL SHAPE DURING CHRONIC COLONIZATION IN A SINGLE INDIVIDUAL

3.1 PREFACE

The results presented in this chapter include contributions from UW Microbiology undergraduate student, Hajirah Farah, and UW MCB graduate student, Lauren Brumage.

3.2 INTRODUCTION

Helicobacter pylori colonizes the stomach of half the world's population. Helical cell shape, flagellar-mediated chemotactic motility, and urease secretion are features that allow *H. pylori* to escape the highly acidic environment of the gastric lumen, penetrate the protective mucus layer, and establish efficient colonization of the gastric epithelium [58].

Analogous to what has been observed for other gram-negative bacteria, helical cell shape in *H. pylori* is structured by the peptidoglycan (PG) cell wall sacculus [20,121]. Our group and others have identified enzymatic and structural components comprising the *H. pylori* 'shapesome,' a suite of proteins involved in patterning and maintaining helical shape [20,122,123]. Isogenic mutants that have lost helicity have reduced colonization and are outcompeted by wildtype in co-infections in animal models [20,124]. It has been proposed that the helical shape of *H. pylori* provides a colonization advantage by enhancing propulsion through viscous mucus in a corkscrew-like manner [125]. This hypothesis is strengthened by biophysical models that predict helical bodies to have increased swimming speed in viscous media compared to rods [126,127]. *In-*

vitro experimental evidence shows loss of cell shape decreases spreading motility by 10-20% in soft agar, but does not impact median flagellum number [124].

Motility is also dependent on the biosynthesis, assembly, and exportation of unipolar sheathed flagella. The flagellar machinery comprises three main components: the basal body, hook, and filament [17]. *H. pylori*'s chemotaxis system is also a major regulator of motility, directing movement towards optimal environments for growth [59,128]. External chemoattractants are sensed by four Tlp chemoreceptors—TlpA, TlpB, TlpC, and TlpD. TlpA senses arginine and bicarbonate, TlpB senses pH and the quorum sensing molecule autoinducer-2, TlpC senses lactate, and TlpD senses reactive oxygen species [129–131]. When metabolites are sensed by the cognate receptors, they transduce information to coupling protein CheW. This activates the histidine kinase CheA and response regulator CheY, controlling flagellar rotation and direction of movement [132,133]. Non-motile mutants without flagellar components or flagellar motor protein, have attenuated colonization in mice as compared to wildtype [19,134]. Several chemotactic mutants have defects in the establishment of acute colonization, while others can establish short-term infections, but do not expand to different niches during prolonged chronic infection [135].

H. pylori isolates have phenotypic variability in cell shape, motility, and chemotaxis [108]. Most characterized clinical isolates have retained helical shape, however the parameters of the helix are not uniform between different lab strains and altering growth conditions can initiate cell shape transformations [108,136–138]. The number of flagella can also vary between clinical isolates [108]. Individual isolates have been found to have anywhere between 0-8 flagella per cell [17,108]. Under certain

culture condition strains have been shown to lose flagellar motility via mutations in genes related to synthesis and assembly [139,140]. The significance of phenotype plasticity and heterogeneity is not completely understood. But, variation likely reflects adaptive responses to the selective forces in individual environments.

Morphology, motility, and chemotaxis are well defined mechanisms required for efficient colonization and establishment of acute infection by *H. pylori*. However, the role of these inter-related phenotypes during later stages of infection are not completely understood. In order to probe genetic and phenotypic changes that occur during infection, we are leveraging a collection of 39 isolates isolated from a single individual at different time points during infection that has been previously described [65]. In Chapter 2, we showed that sub-populations within the collection have distinct cell shapes. Here we expand on these findings leveraging the same collection to examine phenotypic variability in cell shape, motility, and chemotaxis.

3.3 RESULTS

3.3.1 *A rare MreB variant alters helical pitch of H. pylori*

Previous work examined genetic diversity of 39 *H. pylori* isolates from a single individual at two different time points during infection. In this patient, the populations diverged over 6 years of time and diversified to form two genetically distinct sub-populations at each time point (Figure 2.4A). Bacterial cell shape phenotypes varied across genetically defined sub-populations. As described in Chapter 2, all isolates collected from the first time point were helical. At the second time point, a subset of the population lost helical shape due to loss of function mutations in cell shape determining genes *csd4*, *csd5*, and *csd6* (Figure 2.11D). Another sub-population of recent isolates

retained helical shape, yet the helical pitch was altered compared to ancestral isolates and these groups can be differentiated by measuring quantitative parameters from 2-D images of the cells.

A genome-wide association test found six candidate nonsynonymous polymorphisms that correlated with pitch phenotypes across 4 unique genes (Figure 3.1A). Two of the four genes had unknown function. The adhesin *babA* and in bacterial actin homologue *mreB* each had a single non-synonymous mutation. We focused on the *mreB* polymorphism as our top candidate since MreB promotes and maintains rod shape in *E. coli* and contributes to normal cell shape in *H. pylori* [123,141]. All the recent isolates with altered helical pitch carried the variant allele with an alanine substituted for valine at position 284 (*mreB* V284A). MreB contains 4 domains that form an ATP binding pocket. The V284A mutation is located in domain IIA in a region shown to be involved in intra protofilament interaction (Figure 3.1B) [142].

We examined the contribution of this polymorphism to cell shape via an allelic replacement strategy where the variant allele *mreB* 284A amplified from isolate D1 was ligated to a chloramphenicol resistance cassette and introduced into the J99 background (*mreB* 284V) at the native locus. We found J99 strains with the *mreB* 284A variant had altered helical radius and pitch visible by 2-D phase contrast images. To better show reduction in number of helical turns per unit length, the bacteria were treated with the antibiotic aztreonam, which induces cell filamentation (Figure 3.1C). Together the images show that the MreB alanine mutation found in the short pitch isolates partially accounts for the observed cell shape phenotype. However the introduction of alanine mutation to the J99 background does not completely recapitulate

the shape phenotype observed in the recent strains. On average recent strains, including strain D1, had increased cell length (mean=3.97 μm) compared to cell length of ancestral isolates (mean=2.78 μm). Therefore there are other unknown contributions to observed changes in cell morphology between these populations.

Comparative sequence analysis shows that alanine in this position is conserved across other bacterial genera and *H. pylori* strains. The MreB proteins across seven different bacterial species and 11 *H. pylori* strains from different geographic regions were aligned and the only other strain to share a valine at that position was *Thermotoga martima (*Figure 3.2A-B*)*. To determine if the shape phenotype conferred by the valine variant is consistent across other strain backgrounds, we introduced the J99 *mreB* allele (284V) and D1 *mreB* allele (284A) into another strain background—LSH100, a derivative of G27. This isolate has a cell shape phenotype resembling that of the D1 isolate and has been used previously in our lab to identify proteins that contribute to and modify cell shape in *H. pylori*. The G27 variant differs from the J99 version at 3 polymorphic loci, including position 284 where G27 has an alanine and J99 has a valine. Isolate D1, the J99 recent isolate with MreB 284A allele, differs from G27 at only 2 polymorphic loci since it shares two polymorphic loci with J99 but differs at position 284. LSH100 $mreB_{J99}$ mutant strain had increased helical pitch compared to the control LSH100 strain engineered with the D1 *mreB* allele (LSH100 $mreB_{D1}$), which had a similar shape to WT LSH100. These changes in bacterial cell morphology can be differentiated with a scatterplot of centerline cell axis length vs side curvature measurements collected from 2-D image traces (Figure 3.2C). This indicates that the MreB mutation at the 284 position is responsible for this change in cell shape.

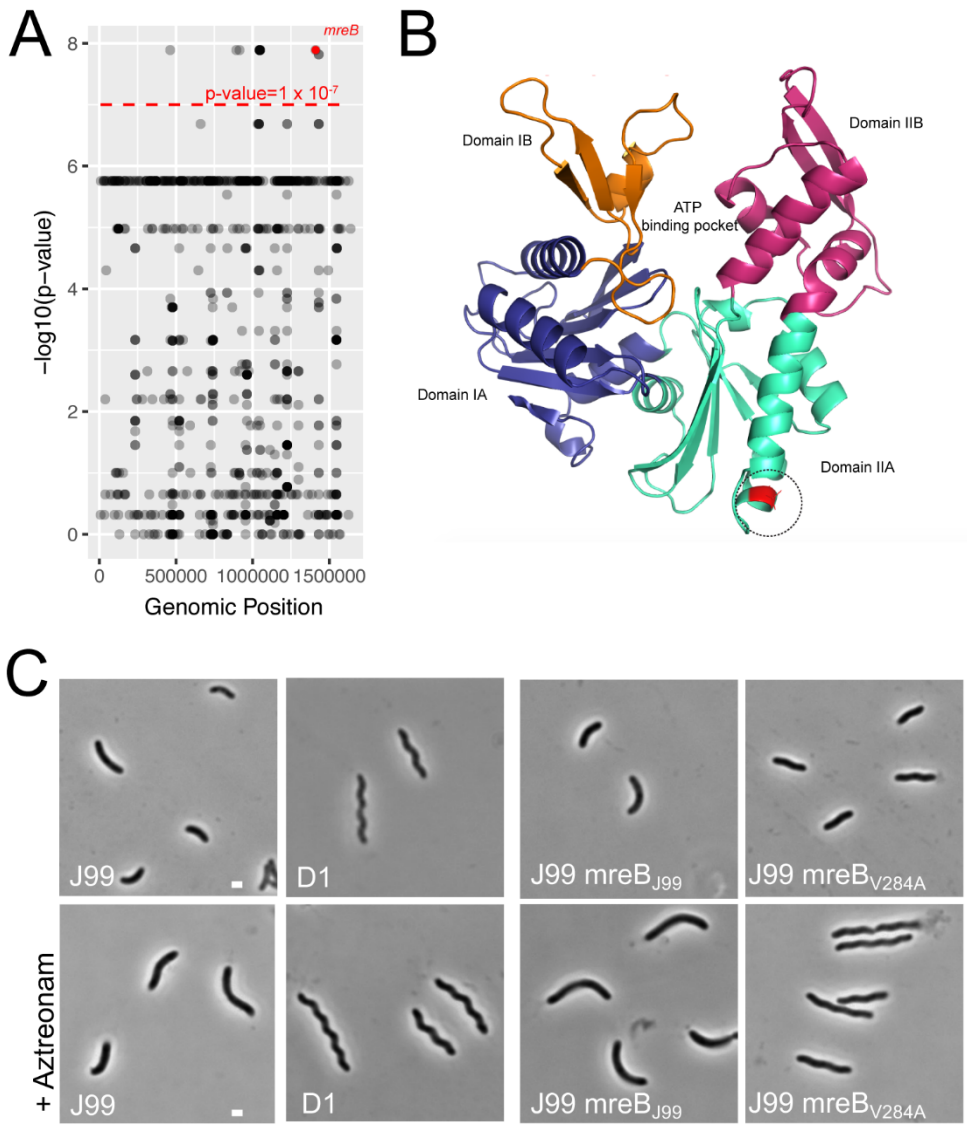
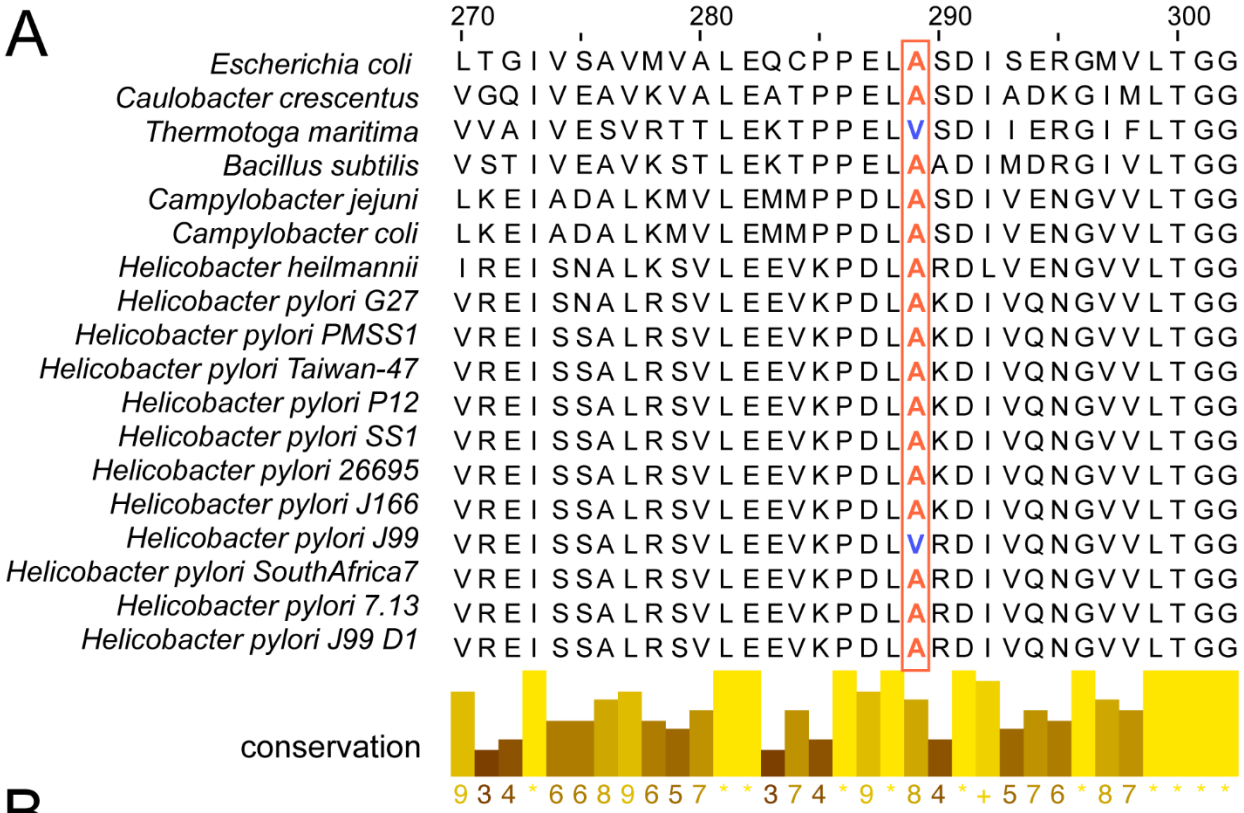


Figure 3.1. The short pitch phenotype maps to a single point mutation in *mreB.*

(**A**) Manhattan plot showing the association of 1,767 SNPs to short pitch (n=15) and long pitch (n=14) cell morphology phenotypes detected within the collection. The Y-axis represents the -log10 p-value and X axis shows the genomic position of the variant relative to the J99 reference genome (AE001439). The dashed red line indicates the threshold for genome-wide significance (p-value= $1x10^{-7}$). The *mreB* variant is highlighted in red and labeled. (**B**) J99 MreB amino acid sequence superimposed on the crystal structure of *C. crescentus* MreB (PDB 4CZJ) with domain IA in blue, domain IB in orange, domain IIA in green, and domain IIB in pink. Position 284 within domain IIA is colored in red and circled. (**C**) Phase contrast images of isolates J99, D1 and isogenic

mutants J99 *mreB*$_{J99}$ and J99 *mreB*$_{D1}$ (scale bar=1 μm) in log phase and after 3 hrs treatment with aztreonam to induce filamentation.

A


270
280
290
300
Escherichia coli LTGIVSAVMVALEQCPPELASDISERGMVLTGG
Caulobacter crescentus VGQIVEAVKVALEATPPELASDIADKGIMLTGG
Thermotoga maritima VVAIVESVRTTLEKTPPELVSDIIERGIFLTGG
Bacillus subtilis VSTIVEAVKSTLEKTPPELAADIMDRGIVLTGG
Campylobacter jejuni LKEIADALKMVLEMMPPDLASDIVENGVVLTGG
Campylobacter coli LKEIADALKMVLEMMPPDLASDIVENGVVLTGG
Helicobacter heilmannii IREISNALKSVLEEVKPDLARDLVENGVVLTGG
Helicobacter pylori G27 VREISNALRSVLEEVKPDLAKDIVQNGVVLTGG
Helicobacter pylori PMSS1 VREISSALRSVLEEVKPDLAKDIVQNGVVLTGG
Helicobacter pylori Taiwan-47 VREISSALRSVLEEVKPDLAKDIVQNGVVLTGG
Helicobacter pylori P12 VREISSALRSVLEEVKPDLAKDIVQNGVVLTGG
Helicobacter pylori SS1 VREISSALRSVLEEVKPDLAKDIVQNGVVLTGG
Helicobacter pylori 26695 VREISSALRSVLEEVKPDLAKDIVQNGVVLTGG
Helicobacter pylori J166 VREISSALRSVLEEVKPDLAKDIVQNGVVLTGG
Helicobacter pylori J99 VREISSALRSVLEEVKPDLVRDIVQNGVVLTGG
Helicobacter pylori SouthAfrica7 VREISSALRSVLEEVKPDLARDIVQNGVVLTGG
Helicobacter pylori 7.13 VREISSALRSVLEEVKPDLARDIVQNGVVLTGG
Helicobacter pylori J99 D1 VREISSALRSVLEEVKPDLARDIVQNGVVLTGG
conservation
9 3 4 * 6 6 8 9 6 5 7 * * 3 7 4 * 9 * 8 4 * + 5 7 6 * 8 7 * * * *

B

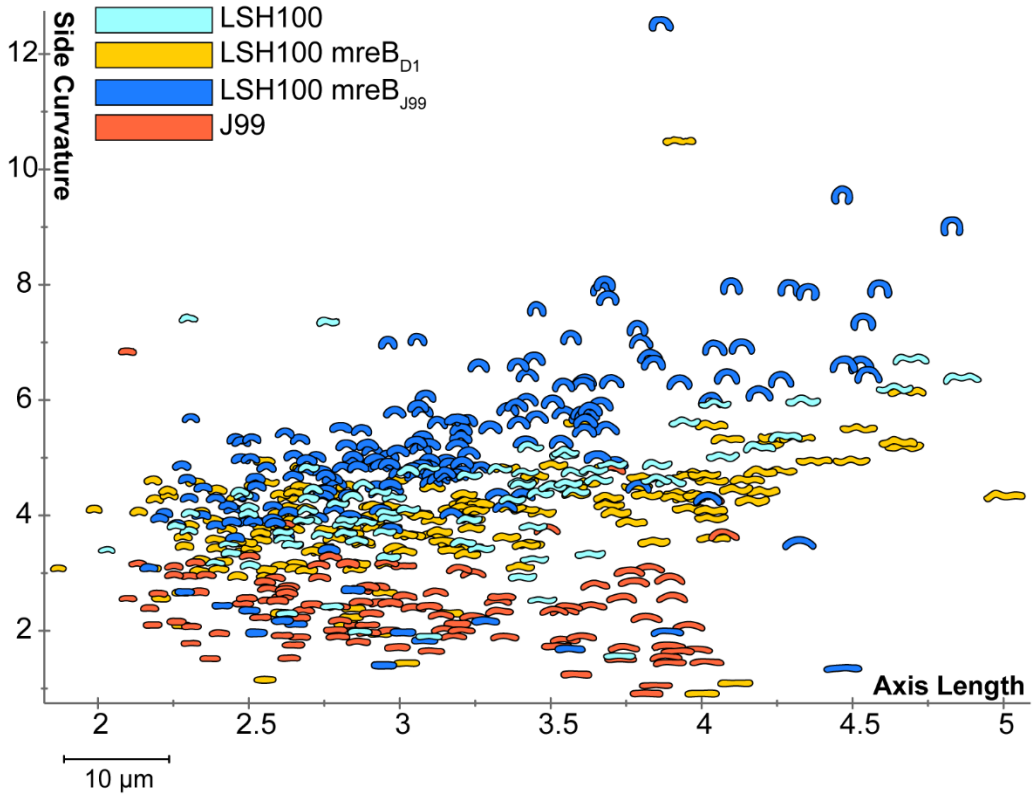

LSH100
LSH100 mreB$_{D1}$
LSH100 mreB$_{J99}$
J99
Side Curvature
12
10
8
6
4
2
Axis Length
2
2.5
3
3.5
4
4.5
5
10 µm

Figure 3.2. MreB 284V allele is a rare variant among *H. pylori* strains that alters helical cell shape when introduced to G27 background.

(A) Phylogenetic tree showing the evolutionary conservation and relatedness of MreB for 7 different bacterial species and 11 strains of *H. pylori* isolated from different geographic regions with Clustal sequence alignment of MreB amino acid positions 270-302 across other *H. pylori* strains and bacterial species. Yellow bars below sequence alignments show amino acid conservation at each position among the strains displayed in tree with asterisk (*) indicating complete conservation across all strains. The mutated position is highlighted with alanine in red and valine in blue. (**B**) CellTool scatter plot of population axis length vs. side curvature of LSH100 WT (light blue), LSH100 $mreB_{D1}$ (yellow) and LSH100 $mreB_{J99}$ (blue), and J99 (orange). Each point is a 2-D image trace of a single cell from within the population. N= 80-180 bacteria per strain.

3.3.2 *Recent isolates have attenuated motility*

Bacterial cell shape has been shown to alter swimming motility. Therefore soft agar swimming assays were used to measure motility for each isolate. Each isolate was inoculated into low percentage agar with a sterilized straight pin and incubated for 5 days. The diameter of the halo of bacterial growth normalized to the J99 ancestral isolate was used to quantify this phenotype. All of the isolates in the J99 collection were motile, traveling greater distances compared to a non-motile mutant with a flagellar motor protein subunit knocked out (J99 Δ*motB*). But motility varied across the J99 sub-populations. On average, motility was attenuated among recent isolates in the collection (0.62 ± 0.14 average distance migrated relative to J99) compared to the ancestral isolates (0.755 ± 0.17 average distance migrated relative to J99) (Figure 3.3). However,

there were outliers that deviated from these means. Isolate SC4 which also belongs to the subgroup 1B described in Chapter 2, had two-fold decreased motility compared to the average motility of the ancestral isolates. Similarly isolates D1, A3, and C10 were all outliers in the recent group that had increased motility relative to other isolates from that timepoint (Figure 3.3).

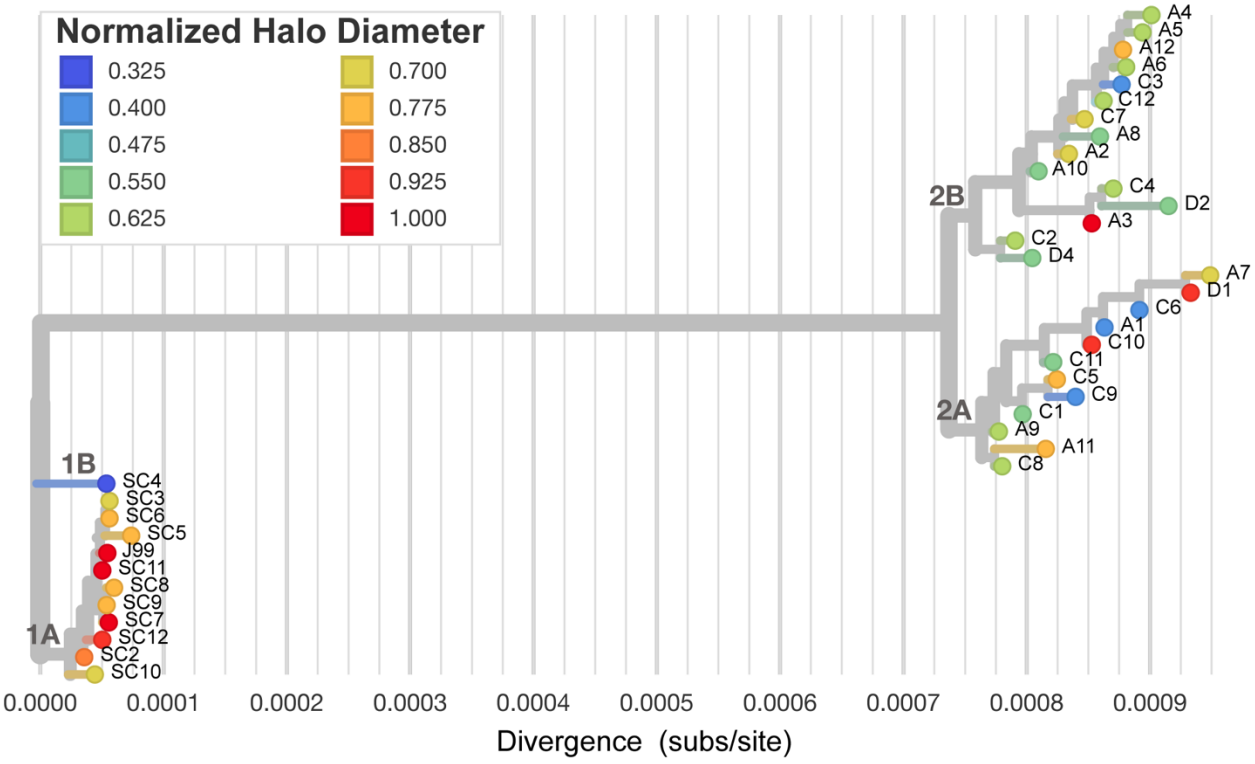


Figure 3.3. Recent isolates have attenuated swimming motility in soft agar.

The maximum likelihood isolate tree (Figure 2.4A) annotated with distance traveled in soft agar after 5 days (halo diameter) normalized to J99. Leaf colors represent the size of normalized halo diameters as indicated in the key.

3.3.3 *Altered cell shape, not flagella number, contributes to attenuated motility among recent isolates*

Previous work that has shown *H. pylori* isolates with helical shape travel further in motility agar than rod mutant isolates. Consistent with this, we found that the rod shaped isolates in the collection traveled similar distance to a J99 rod mutant (J99 Δcsd4), but decreased compared to J99 wildtype. Surprisingly, the J99 MreB^{V248A} mutant, which has retained helical shape but has altered pitch compared to J99 WT, also had attenuated motility. This mirrors the decreased motility observed among helical recent isolates compared to the ancestrals (Figure 3.4A). To determine if the differences in motility conferred by the residue 284 V to A mutation are a result of cell shape changes alone, we enumerated flagella of the mutant strains from transmission election micrographs. The MreB^{V284A} mutant does not have altered flagella number (data not shown), indicating that cell shape alterations are likely contributing to attenuated motility within this population.

However, these cell shape alterations only partially explain variability in motility observed in the dataset. Variation between strains within the long and short pitch helical isolates is higher ($stdev_{lpitch}$= 0.18, $stdev_{spitch}$=0.15) than rod shaped isolates ($stdev_{rod}$=0.09). Ancestral strain SC4 has 50% reduced motility compared to J99 despite having the same cell shape. (Figure 3.4B). This suggests that other underlying factors contribute to these phenotypes. Cell length is also varied within the entire population and increased cell length is associated with decreased motility in this dataset (p-value=0.03, R^2=0.1512). However, variation of cell length within time points is low (data not shown) and likely does explain the majority of the variation that is observed between

isolates. We next hypothesized that variation in flagella number between these strains may drive differences in motility, therefore we enumerated flagella from transmission election micrographs. There was significant variation in flagella number between strains in the collection. All the ancestral isolates where flagella were enumerated had a median flagellum number of 3, including SC4 which had diminished motility compared to the others. The flagella of the recent isolates varied significantly. Isolates C1, C10, and C11 had a median of 0 flagella. While, isolates A3, A8, D1, C5, had a median of 4 flagella (Figure 3.5A). Median flagella number did not correlate with halo diameter size, indicating other factors or a combination of factors contribute to the soft agar motility phenotype (Figure 3.5B).

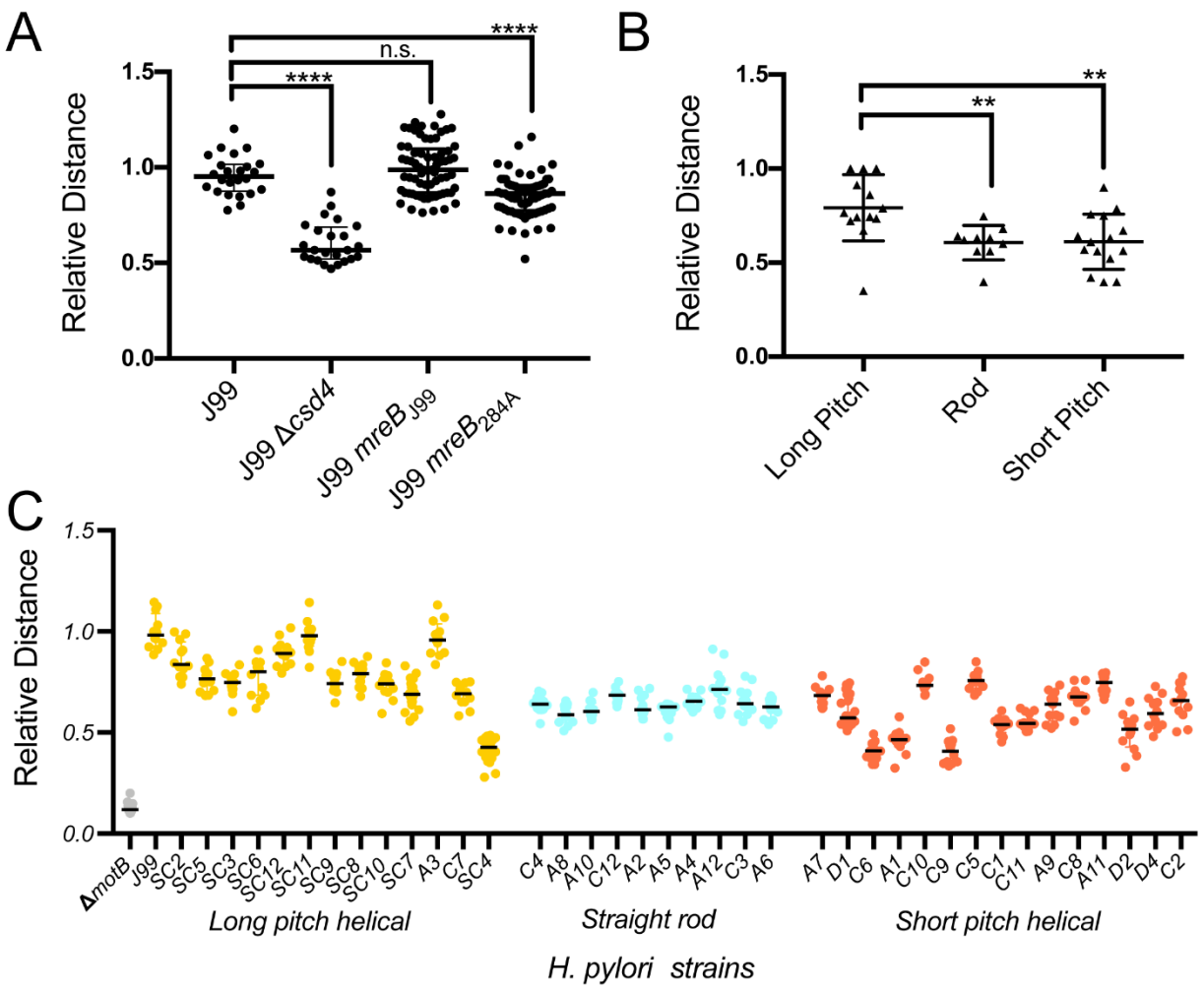


Figure 3.4. Distance traveled in soft agar differs according to cell morphology.

(**A**) Plot shows the distance traveled in motility agar over 5 days relative to J99 for J99 isogenic variants with straight rod (J99 Δ*csd4*) and short pitch helical cell shape (J99 *mreB*^{V284A}) as well as control strains J99, and J99 *mreB*$_{J99}$. Points represent pooled replicates from at least two independent experiments with bars representing the median and interquartile range. (**B**) Plot shows the average relative distance traveled in soft agar of the J99 patient isolates as reported in Figure 3.3 grouped by cell shape (long pitch, rod, short pitch). Bars represent the median value and interquartile range. (**C**) Spreading motility of all 39 isolates grouped by cell shape reported as distance relative to J99. Long pitch helical isolates are displayed in yellow, straight rods in blue, and short pitch isolates in red. Bar indicates the median value. Isolate J99 and isogenic mutant Δ*motB* served as positive and negative controls, respectively. P-values were

calculated from pooled experimental replicates using a Mann-Whitney non-parametric test (n.s. not significant, ** p<0.01, **** p<0.0001).

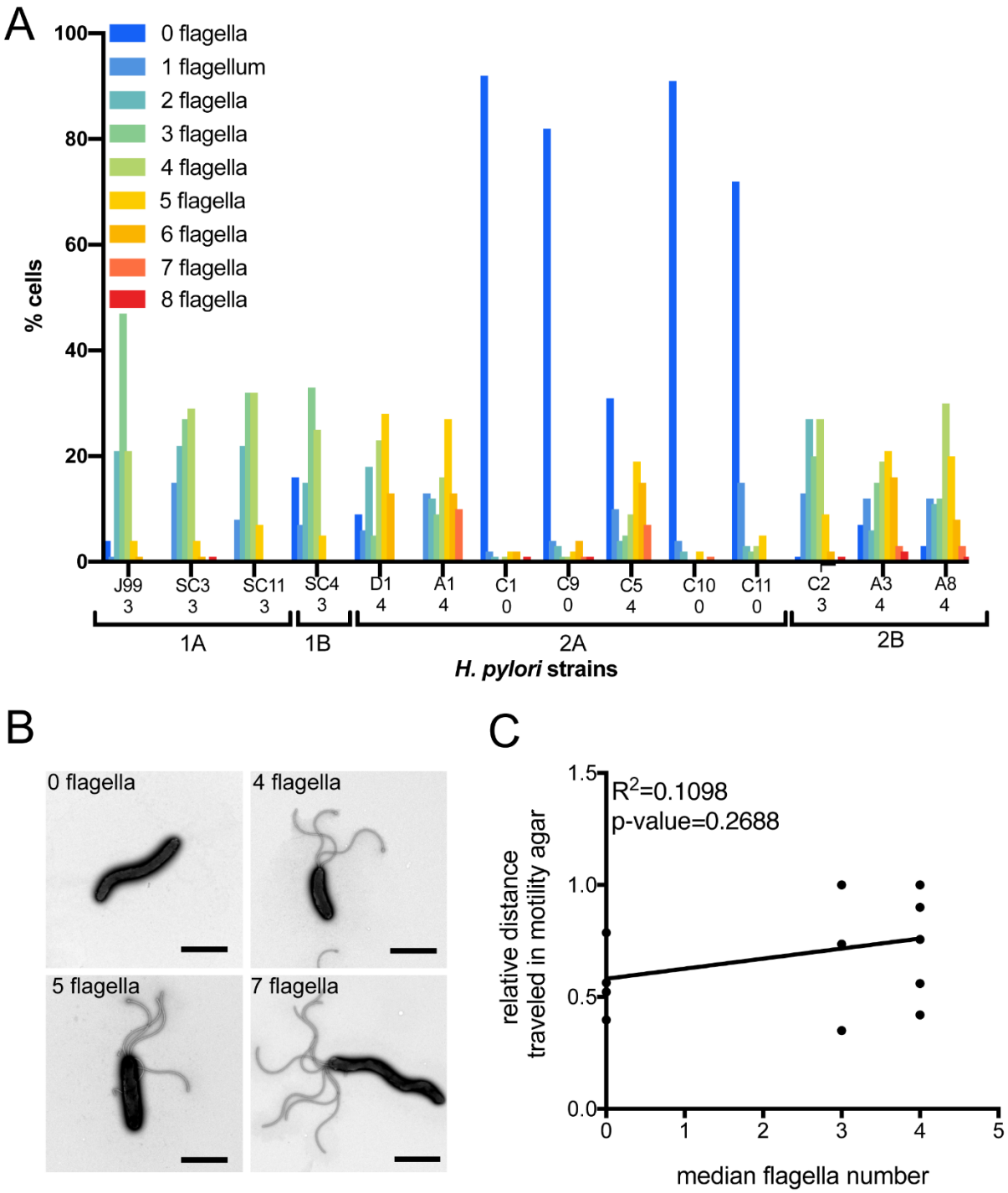


Figure 3.5. Flagellum number varies but is not correlated with motility.

(**A**) For each isolate, a histogram shows the distribution of flagellum number across individual cells within the clonal population. The Y-axis shows the percentage of cells within the total population with the indicated flagella count. Colors represent the flagellum count per cell observed as annotated in the key. Isolates are ordered and labeled based on the sub-population they belong to on the tree (1A, 1B, 2A, 2B). Below each isolate name is the median flagellum number rounded to the nearest integer (n=80-100 cells per isolate). (**B**) Representative micrographs showing individual cells with varying number of flagella. Scale bar, 2μm. (**C**) Correlation of median flagella number and average distance traveled in soft agar relative to J99 as reported in Figure 3.4C. Each point represents a single isolate in the J99 collection. Plot shows a linear regression with p-value derived from the F-test and correlation coefficient (R^2) shown.

3.3.4 *Chemotactic motility varied between isolates*

In addition to differences in the motility of these strains, isolates differed in chemotactic migration behavior in soft agar. These unique patterns, consisting of cell dense rings and zones of clearance, result from bacterial movement across nutrient gradients. Isolates were assigned to three different groups based on the chemotactic ring pattern observed. Several recent isolates were non-chemotactic and formed uniform halos emanating from the point of inoculation. Chemotactic isolates fell into two groups that can be differentiated by the size of the zones of lower density observed from the point of inoculation. Group A had a larger zone of lower density and well defined ring borders, while group B had a less defined ring with small zone of lower density (Figure 3.6A). Most isolates that lost chemotactic motility or have chemotactic type B, were located into subgroup 2A on the tree (Figure 3.6B). However, loss of chemotaxis patterns varied independently from motility measured by distances traveled in soft agar.

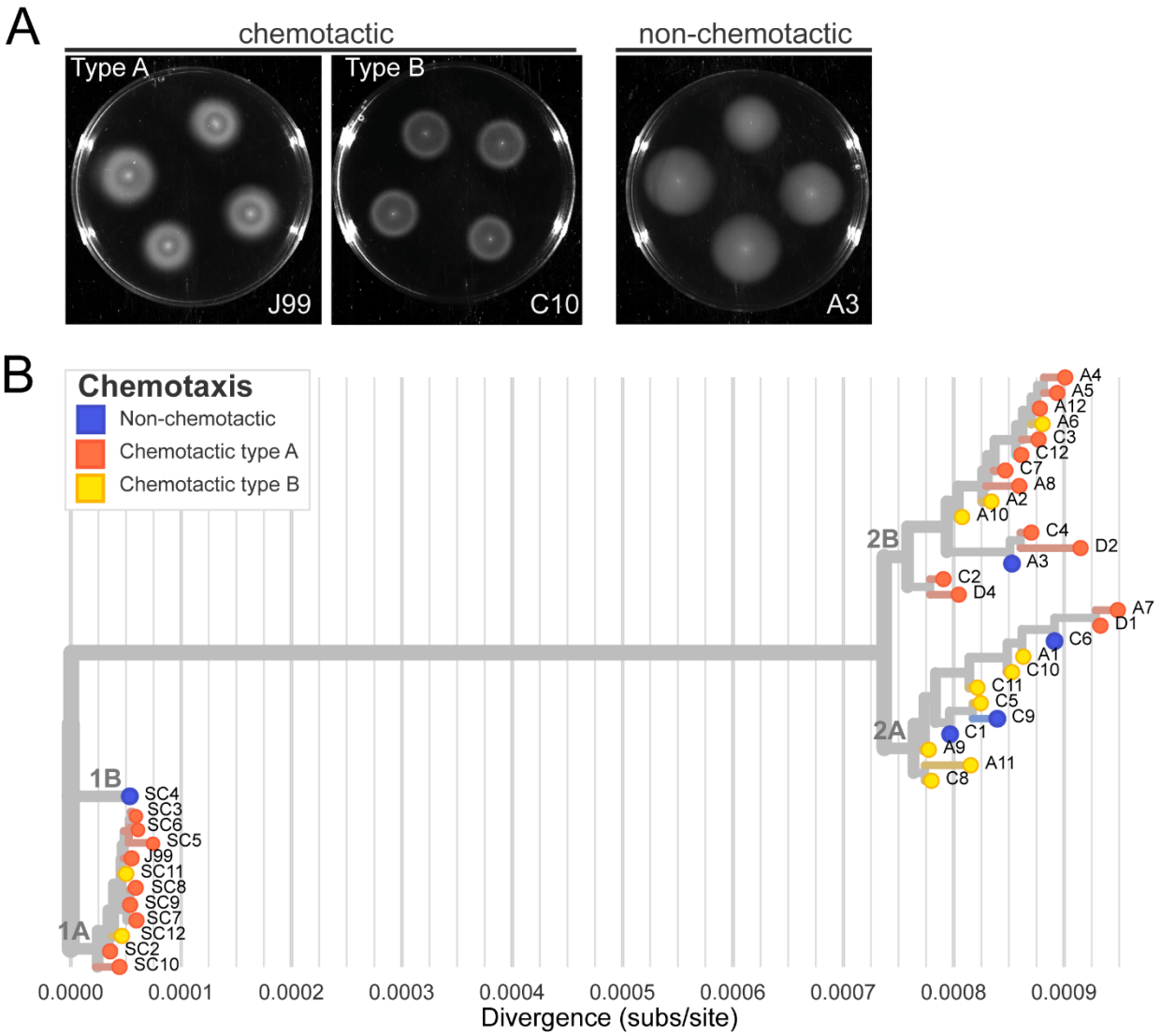


Figure 3.6. Isolates display variable chemotactic patterns in soft agar.

(A) Representative images of soft agar plates 5 days post inoculation showcasing phenotypic variation in chemotactic ring formation. Isolates J99 and C10 show chemotactic strains with the A and B phenotypes respectively. Isolate A3 is a representative of the non-chemotactic phenotype. **(B)** The maximum likelihood isolate tree (Figure 2.4A) overlaid with chemotactic phenotype groups. Non-chemotactic isolates are shown in blue, type A in red, and type B in yellow. The X-axis displays the divergence in number of substitutions per site (subs/site).

3.3.5 *SNPs located within genes associated with cell shape, chemotaxis, and motility*

A genome-wide screen was performed to identify SNPs associated with both chemotaxis and flagellar number phenotypes, however, no single polymorphism was significantly associated. This likely indicates that multiple genetic factors are involved or that there is evolutionary convergence on the same phenotype. To further probe the genetic underpinnings of these phenotypes, we compiled a list of 56 motility and chemotaxis genes listed in public databases, PATRIC and MDGB. We then queried against our list of 1,767 polymorphisms and 483 indels. A total of 10 nSNPs were found in these genes and are listed in Table 3.1. Three of the four chemoreceptors—*tlpA*, *tlpB*, *tlpC* and the histidine kinase *cheA*, involved in regulating chemotaxis, have nonsynonymous mutations. Additional mutations in flagellar proteins, *flgE, fliF, flhF*, were also detected. Interestingly, most mutations in these genes were confined to specific subgroups defined in Chapter 2 with little overlap.

Table 3.1 Nonsynonymous polymorphisms detected across 57 motility and chemotaxis genes

Gene ID	Gene Annotation	AA change	Function	Group 1A (n=11)	Group 1B (n=1)	Group 2A (n=12)	Group 2B (n=15)
jhp0075	*tlpC*	G254R	methyl-accepting chemotaxis transducer	0	0	12	15
jhp0091	*tlpA*	E77K	methyl-accepting chemotaxis protein	0	0	12	15
jhp0095	*tlpB*	K72E	methyl-accepting chemotaxis protein	0	0	12	15
jhp0095	*tlpB*	I245T	methyl-accepting chemotaxis protein	11	1	0	0
jhp0844	*flgE'*	L308F	flagellar hook protein homolog	0	0	12	15
jhp0325	*fliF*	E379K	flagellar basal-body M ring protein	0	0	0	15
jhp0325	*fliF*	Q75K	flagellar basal-body M ring protein	0	0	12	15
jhp0389	*flhF*	I239M	flagellar biosynthetic protein	0	0	12	15
jhp0989	*cheA*	R781H	histidine kinase	0	0	0	12
jhp0989	*cheA*	R806H	histidine kinase	11	0	0	0

Indels were detected in only two genes on this list. The indels were in TlpA, a methyl-accepting chemotaxis chemoreceptor protein that senses arginine, bicarbonate and is repelled by acid, and ArsS, the sensor component of a two-component transcriptional regulator. The majority of the isolates with mutations in this region were located in subgroups 1B and 2A. TlpA was predicted to be out of frame for isolates within subgroup 1A but shifted in frame for all isolates in 1B, 2A and a few in 2B (Figure 3.7A). These predictions are consistent with expression levels showing loss of TlpA expression in isolates with predicted out-of-frame variants (Figure 3.7C). ArsS was predicted to be in-frame for all isolates, but there was high variability in a cytosine tract in the C-terminal end of the protein that generates transcripts of varying lengths. J99 ancestral has a tract length of 10 cytosines. The isoforms have been described previously and are named for the last three amino acids in the protein. The predicted isoform for J99 is LWG since the last amino acids are leucine (L), tryptophan (W), and glycine (G). Other isoforms that have been documented among clinical isolates are SND, PKI, and EKQ. However only SND, LWG and PKI isoforms were detected in this collection (Figure 3.7B) [143]. To test functional significance of these C-terminal repeats, the entire *arsS* locus was deleted from strain J99 and replaced with another variant present in isolate D1 (LWG 10→3 cytosines). Motility in the J99 $arsS_{D1}$ was compared to J99 $arsS_{J99}$ with the J99 WT *arsS* variant (LWG 10) at the native locus. The J99 $arsS_{D1}$ strain was found to have decreased motility compared to the J99 *arsS* strain with the J99 WT *arsS* variant but did not alter chemotaxis phenotype (Figure 3.7D). Sanger sequencing revealed that many isoforms exist in a single isolate clone

and likely the tract length varies during a single *in-vitro* passage. This suggests that *H. pylori* may be dynamically regulating motility in response to environmental signals.

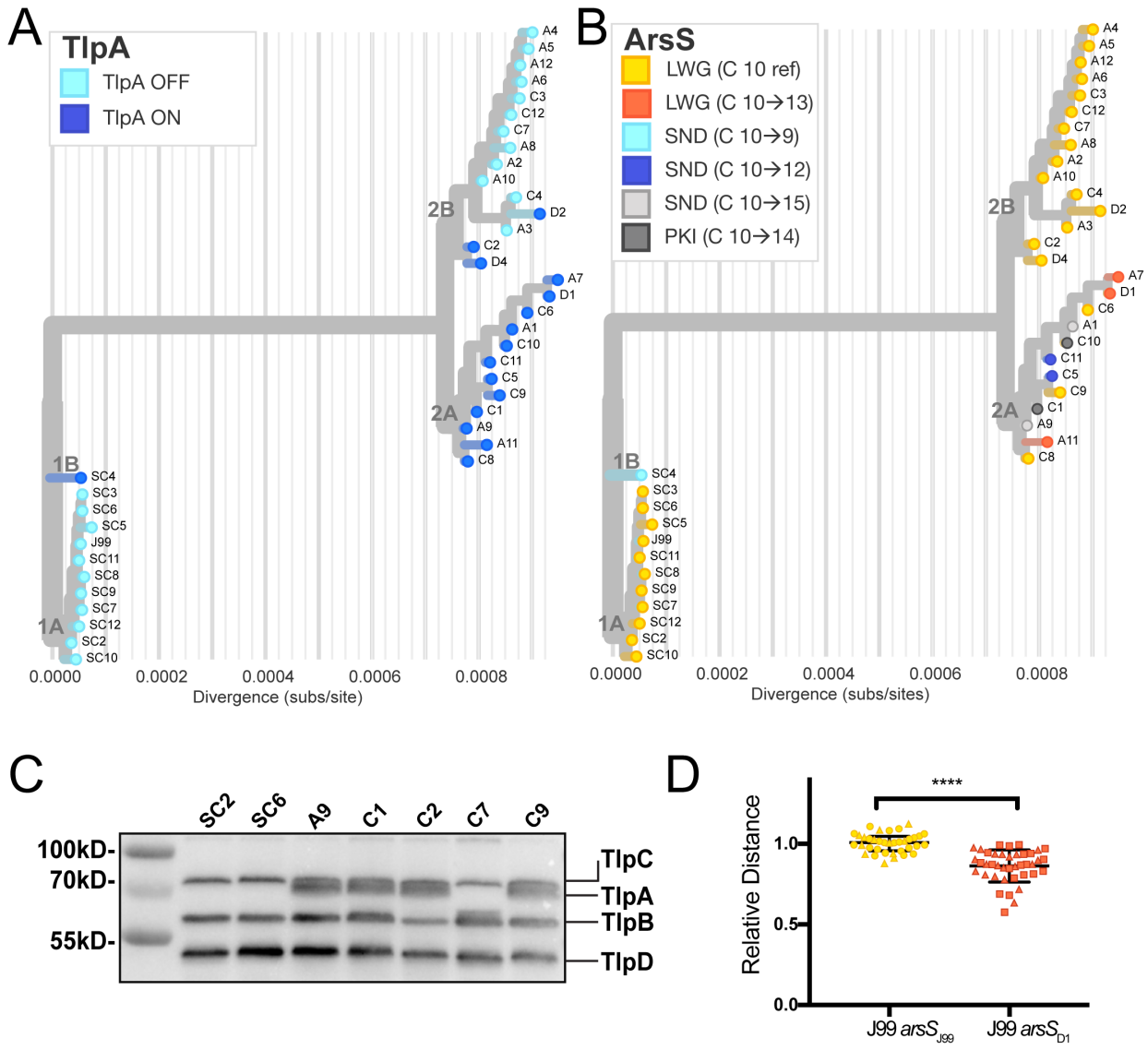


Figure 3.7. Indels generate in and out-of-frame variants of chemoreceptor TlpA and variant isoforms of histidine sensor kinase ArsS.

(A) The maximum-likelihood tree (Figure 2.4A) overlaid with allele variants of TlpA. Out of frame TlpA variants are colored in light blue and in frame variants are colored in dark blue according to the key. **(B)** Maximum likelihood tree overlaid with allelic variants

of ArsS. (**C**) Western blot analysis of Tlp chemoreceptors in isolates SC2, SC6, A9, C1, C2, C7, and C9. The position of each chemoreceptor is indicated on the right, and the molecular mass in kilodaltons (kD) are on the left. Predicted size of TlpD is 48.4kD, TlpB is 62.8kD, TlpA is 74.5kD, and TlpC is 75.2kD. (**D**) Each point represents the distance traveled in motility agar over 5 days relative to J99. Pooled replicates from at least two independent experiments are shown with shape of points representing individual clones tested. Bars indicate the median. The X-axis indicates *arsS* variants with J99 WT variant (LWG, 10 cytosines) in yellow and J99 D1 variant (LWG, 13 cytosines) in red. P-values were calculated from pooled experimental replicates using a student's test (**** $p<0.0001$).

3.4 Discussion

Motility, cell shape, chemotaxis are well established virulence factors essential for *H. pylori* colonization of the gastric environment. Less is known about the role these systems play at later stages of infection. Whole genome sequencing of paired samples from the same individual have revealed an enrichment in mutations in annotated motility and chemotaxis genes [90]. Additional studies have shown that engineered isogenic mutants that have lost helicity, elicit less gastric pathology compared to helical isolates in prolonged infection in a mouse model; and specific chemotaxis mutants were not able to expand to colonize corpus glands as typically observed in the chronic stage of infection [21,144]. Together, genetic and phenotypic examinations of cell shape, motility, and chemotaxis in single infections indicate these systems likely continue to serve important function beyond initial colonization.

In this study we leveraged the J99 collection of isolates to further characterize the phenotypic variation in cell shape, motility, and chemotaxis over time in a single host

and determine genetic underpinnings of phenotypic variation within these populations. We had previously identified and defined three cell shape phenotypes within this population—rod, short pitch helical, and long pitch helical (2.3.5). Loss of function mutations in cell shape determining genes conferred rod shape, however it was unknown which mutations conferred alterations in helical pitch.

Here we showed that a valine to alanine mutation in domain IIA of MreB, identified by genome-wide association, is sufficient to alter the helical pitch of ancestral isolate J99. The J99 allele MreB 284V also conferred altered shape that resembles the long pitch phenotype when transferred to the strain LSH100 background. MreB is an essential gene in *H. pylori* and is thought to be involved in maintaining or patterning helical cell shape due to its preferential localization to the minor helical axis [123]. In *Bacillus subtilis* and *Escherichia coli,* MreB has been shown to polymerize to form long filaments *in vivo* [145]. Intra-filament interactions are mediated by surface amino acids in domain IIA. Studies in *Caulobacter crescentus* show that mutagenesis of this region results in a monomeric form of MreB *in vitro*; and *in vivo*, renders gently curved cells spherical [146]. Further structural analysis of the filaments formed by variant versions of MreB in *H. pylori* described in this study, might provide additional insight into the mechanism by which MreB can alter helical cell pitch in *H. pylori* and other curved and helical bacteria. The alanine residue at position 284 is conserved among other *H. pylori* strains, suggesting that in most environments shorter helical pitch might be more advantageous.

Exploration of motility within isolates in the collection revealed that recent isolates had diminished motility compared to ancestral isolates. Previous work has shown that

rod shaped isolates travel shorter distances in soft agar than wildtype helical strains [124]. In this study we saw that the MreB alanine variant had similarly diminished motility, akin to rod isolates. This suggests that *H. pylori* may have an optimal shape for swimming, but that competing selective pressures in natural infections may create shape heterogeneity. It has been hypothesized that cell shape transitions may also serve as a mechanism to escape host immune response during chronic infection, but more work needs to be done to understand the role of shape beyond initial colonization [137,147].

Observed variation in motility was not completely explained by cell shape differences. Since variation in flagellar number is a major contributor to motility, we measured median flagella number in the population. Unexpectedly, we found that there was variability between isolates, but this was uncorrelated with motility phenotypes. The finding that some isolates had no flagella when grown up in liquid media, but were motile in soft agar may indicate that flagellar expression in these isolates is sensitive to environmental growth conditions.

In addition to variability in flagella number, it was also observed that isolates had differences in chemotactic ring formation in soft agar. Other studies examining genetic differences within a single individual over time found a significant amount of variation in chemoreceptors [90]. In our dataset, chemotactic behavior or flagella number were not uniquely associated with any of the 2,252 polymorphic loci. However, nonsynonymous SNPs in genes related to flagellar biosynthesis such as *fliF*, *flhF, flgE* and chemotaxis proteins including *cheA*, *tlpA*, *tlpB*, and *tlpC* were observed, indicating phenotypes may be a result of multiple mutations. Frameshift mutations in *arsS* and *tlpA* were also

detected. ArsS frameshifts occur in a homopolymeric repeat tract and generate unique isoform lengths that vary between isolates and clonal populations. Our data suggest this has the potential alter to motility by an unknown mechanism. Additional strains with these select variants need to be engineered in combination in order to understand the functional significance, but this has the possibility of uncovering genetic mechanisms *H. pylori* employs to regulate motility and chemotaxis during chronic infection.

3.5 METHODS

3.5.1 *H. pylori strains used in this study*

For this study we leveraged the collection of 39 J99 strains from a single individual that have been previously sequenced and described in section 2.3.1. The whole genome sequences are publicly on NCBI SRA database BioProject ID (PRJNA622860) and a full list of the variant calls referenced in this study can be found at <*https://github.com/salama-lab/Hp-J99*>. LSH100 is a mouse adapted derivative of the clinical isolate G27 [148].

3.5.2 *Culture conditions for H. pylori*

All *H. pylori* isolates were grown on horse blood agar (HB agar), solid media, or shaking liquid cultures. HB agar plates contain 4% Columbia agar base (Oxoid, Hampshire, UK), 5% defibrinated horse blood (Hemostat Labs, Dixon, CA), 10 mg/ml vancomycin (Thermo Fisher Scientific, Waltham, MA), 2.5 U/ml polymyxin B (Sigma-Aldrich, St.Louis, MO), 8 mg/ml amphotericin B (Sigma-Aldrich), and 0.2% β-cylodextrin (Thermo Fisher). To induce filamentation, log-phase cultures grown in broth were

diluted to an OD_{600} of 0.2 (OD_{600}), resuspended in broth containing 2 mg/l aztreonam, and shaken for 3 hrs. All plates and flasks were grown at 37°C under micro-aerobic conditions in 10% CO_2, 10% O_2, 80% N_2, as previously described [111]. For resistance marker selection, HB agar plates were supplemented with 15 µg/ml chloramphenicol, or 30 mg/ml sucrose, as appropriate.

3.5.3 *H. pylori strain construction*

Six J99 mutants were constructed (J99 Δ*motB*, J99 Δ*csd4,* J99 *mreB*$_{J99}$, J99 *mreB*$_{284A}$, LSH100 *mreB*$_{J99}$, LSH100 *mreB*$_{D1}$) and are listed in Table 3.2. Isogenic knockout mutant, J99 Δ*motB and* J99 Δ*csd4*, were constructed using a vector-free allelic replacement strategy. Upstream and downstream genomic regions flanking the gene were amplified and ligated to a catsacB cassette, which confers both chloramphenicol resistant (*cat*) and sucrose sensitivity (*sacB*). Positive clones were selected with 15 µg/ml chloramphenicol, as previously described [87,88]. To create the MreB allelic swaps, we integrated variant alleles by ligating a chloramphenicol acetyl transferase resistance cassette (*cat*) to the *mreB* allelic variants amplified by PCR [52]. Mutants were selected for using the plate conditions and methods described above. All mutants were validated via diagnostic PCR and Sanger sequence. Primers used for generating *H. pylori* mutants are listed in Table 3.3.

Table 3.2.Strains used in this study

Strains	Description	Reference
LSH100	Mouse derivative of G27	Lowenthal et al. [139]
LSH100 mreB$_{J99}$	LSH100 with J99 allele (MreB 284V variant)	This work
LSH100 mreB$_{D1}$	LSH100 with D1 (MreB 284A variant)	This work

J99	Clinical isolate	Alm et al.[64]
J99 ⊿csd4	J99 with catsacB cassette at *csd4* native locus	This work
J99 ⊿motB	J99 with catsacB cassette at *motB* native locus	This work
J99 mreB$_{J99}$	J99 with *mreB* J99 (MreB 284V) variant	This work
J99 mreB$_{284A}$	J99 with *mreB* 284A variant from isolate D1	This work
J99⊿ arsS	J99 with *arsS* knocked out	This work
J99 arsS $_{D1}$	J99 with D1 variant of *arsS* (LWG 13)	This work
J99 arsS$_{J99}$	J99 with J99 variant of *arsS* (LWG 10)	This work

Table 3.3. Primers used in this study

Primer name	Sequence (5'-3')	Target/Use
mreB_F	ATGATTTTTAGCAAATTAATCGGTTTG	Amplify *mreB* gene, upstream of catR cassette
mreC_DN_R	ATTTCATTGAAAGGGGTTAAAAACGCGCT	Amplify downstream flanking region of *mreB*
mreB_R	CACCAATATCCACGATCAAACTCC	Amplify downstream flanking region of mreB
mreB_catR_UP_R	atccacttttcaatctatatcGCATTATTCACTAAAACCCACA	Amplify upstream flanking region of *mreB*, homology to catsacB cassette
mreB_catR_DN_F	cccagtttgtcgcactgataaATGCGTTTTTATTTTAAATTCCTTT	ampliy *mreB* downstream flanking region, homology to catsacB cassette
mreC_R	GGTCGCTGATATAAGAAGAGCTGCC	Primer used for sequencing *mreB* variant
TlpA_F	CTATAGCTCATGCCCAAACAATCTTGCCAT	*tlpA* sequencing primer
TlpA_R	TCCCCTAACAATTCCTTACCCTTAGCTTTATC	*tlpA* sequencing primer
arsS_F	TTATCGTTAGACACAGAAGCGTTT	*arsS* sequencing primer
arsS_R	AGATAAGGATAGCCGAGAAGTGTATAT	*arsS* sequencing primer

3.5.4 *Analysis of cell morphology*

Phase contrast microscopy and quantitative analysis using CellTool software package was performed as previously described [20]. Bacterial cell masks were generated using the thresholding function in ImageJ. Average side curvature, wavenumber, and

centerline axis length were derived from thresholded images of bacteria (>100 cells/strain) using the CellTool software package [149].

3.5.5 *Modeling of variant strains of MreB*

The J99 MreB amino acid sequence was aligned to *C. crescentus* MreB sequence in Jalview (<http://www.jalview.org/>). Using the alignment file, the J99 sequence was superimposed onto *C. crescentus* MreB crystal structure (PDB 4CZL) using Chimera (<https://www.cgl.ucsf.edu/chimera/download.html>) and figure was colored and formatted in Pymol v. 2.4.2 [150].

3.5.6 *Quantification of motility and chemotaxis with soft agar assays*

Soft agar plates (Brucella broth (Thermo Fisher Scientific, Waltham, MA), 0.35% agar (Thermo Fisher Scientific), 10% FBS (Gemini BioProducts)) were prepared two days in advance of inoculation as previously described [18]. Bacteria were grown in liquid culture to an OD_{600} of 0.5 – 0.7. Each strain was diluted back to OD_{600} of 0.4 before inoculation with sterilized straight pins and incubated for 5 days in 10% CO_2. Plates were imaged using BioRad Gel Documentation System using white light at consistent dimensions (13.4 cm x 10 cm) and exposed for 0.755 seconds. Halo diameters were measured from images using Fiji <https://imagej.net/Fiji>.

3.5.7 *Flagella enumeration from transmission electron micrographs*

Cells were collected by pelleting 1 OD_{600} in a microcentrifuge for 5 min at 2000 rpm and fixed in half-strength Karnovsky's fixative (2.5% glutaraldehyde, 2% paraformaldehyde in 0.1 M cacodylate buffer) to a concentration of 10 OD_{600}. Cells were then applied to glow-discharged copper grids, and negatively stained with 0.75% uranyl acetate.

Images were acquired with JEOL 1400 transmission electron microscope using a Gatan UltraScan 1000xp camera with 2K x 2K resolution. Flagella were enumerated from images with medians calculated from 80-100 cells.

3.5.8 *Immunoblotting H. pylori whole cell lysate for Tlp proteins*

H. pylori cell lysates were prepared by pelleting 1 OD_{600} of log phase (0.3–0.7 OD_{600}) in a microcentrifuge 2 min at max speed. Cells were resuspended in 2x protein sample buffer to a concentration of 10 OD_{600} per ml and boiled for 10 min. Protein lysate was separated on a 10% SDS-PAGE and transferred onto PVDF membranes using the BioRad Turbo-transfer system according to the manufacturer's instructions. Membranes were blocked for 1 h at RT with 5% non-fat milk in TBS-T (0.5 M Tris,1.5 M NaCl, pH 7.6, plus 0.05% Tween 20). After blocking, membranes were incubated with anti-GST-TlpA22, a pan-Tlp antibody (gift from Dr. Karen Ottemann, University of California Santa Cruz [151]), at a 1:1000 dilution in 5% non-fat milk at RT for 1 h. Four 10 min washes with TBS-T were followed by a 1 h incubation at RT with anti-rabbit horseradish peroxidase-conjugated α-immunoglobulin G (Santa Cruz Biotechnology) antibody at 1:20,000 dilution in TBS-T. After four more TBS-T washes, antibody detection was performed with ECL Plus (Pierce) detection kits, following the manufacturer's protocol and imaged directly on film or with the BioRad Gel Documentation System.

3.5.9 *SNP associations and statistics*

A Fisher's exact test was used to identify polymorphic sites significantly associated with the phenotypes described in the study and Manhattan plots showing the p-values at all polymorphic positions was generated in R v4.0.2 <https://www.r-project.org/>. P-values

were corrected for multiple testing using Benjamini and Hochberg false discovery rate methods [113]. Adjusted p-values < $1x10^{-7}$ were considered statistically significant. The complete list of genes associated with motility or chemotaxis were curated based off annotations in PATRIC <https://www.patricbrc.org> or the microbial genome database <http://mbgd.genome.ad.jp/> (Table 3.4). Statistical analyses were performed according to tests specified above and in each figure legend using Prism v7 software (GraphPad) or R v4.0.2. P-values less than or equal to 0.05 were considered statistically significant and are marked with asterisks (*, p<0.05, **, p<0.01; ***, p<0.001; ****, p<0.0001; n.s., not significant)

Chapter 4. *COLONIZATION OF THE METAPLASTIC STOMACH DEPENDS ON STRAIN SPECIFIC GENOTYPES*

4.1 PREFACE

This chapter includes experiments performed and contributed by Dr. Valerie O'Brien as a part of a manuscript in preparation on which we are co-authors. I performed sequence analysis and *in vitro* tissue culture experiments, cytotoxicity experiments, and RNA sequencing with the help of Roberto Ramírez and Jessica Cai. Infections with PMSS1 and SS1, quantification of gland occupation, and writing associated with these sections were done by Dr. Valerie O'Brien.

4.2 INTRODUCTION

H. pylori stomach infection is the leading risk factor for the development of gastric malignancies, with an estimated 75% of gastric cancer cases attributed to active

infection [152]. Although the mechanisms by which *H. pylori* induces gastric cancer is incompletely known, the tissue changes preceding cancer development have been well described [34]. All infections lead to chronic inflammation of the stomach lining (gastritis). A small fraction of these cases progress to atrophic gastritis (loss of acid secreting parietal cells), SPEM (spasmolytic polypeptide expressing metaplasia) and/or IM (intestinal metaplasia), which then can progress to dysplasia and cancer [153,154]. Eradication of *H. pylori* prior to the development of metaplasia reduces risk of gastric cancer, which supports a model in which *H. pylori* promotes early tissue alterations leading to carcinogenesis [155].

The dynamics between host-and-pathogen shifts during the course of infection as acid homeostasis, nutrient availability, and immune responses fluctuate [91]. Infections are typically established in the antrum where pH is closer to neutral. However, loss of acid secreting parietal cells due to atrophy of the corpus glands promotes expansion of *H. pylori's* niche from the antrum to the corpus in both human and animal models [16,21]. Within these different topographic regions, *H. pylori* can be found in the protective mucus layer above the epithelium and within glands [156–158]. At the epithelial interface, *H. pylori* secretes effectors directly into host cells as well as into the surrounding environment in order to extract nutrients for growth, resulting in a robust inflammatory response [47]. It is thought that by-products of inflammation generated from infection lead to accumulation of mutations in gastric epithelial cells, which are sufficient to activate oncogenic pathways in some of individuals [31].

In addition to driving inflammation association with mutation, we recently showed that *H. pylori* presence alters KRAS driven induction of SPEM, IM, and dysplasia

[159,160]. While these studies suggest that *H. pylori* impacts the trajectory of pre-malignant tissue changes, tissue changes in turn impose selective pressures on the infecting bacterial populations. Bacterial persistence in response to tissue remodeling in the gastric environment is facilitated by genomic diversification through mutation as well as inter and intra genomic recombination [86]. Comparisons of paired isolates collected longitudinally from the same individual document extensive genetic diversification reflecting adaptation to niches within a single host [61,90,94]. Examination of sequence diversity between individual isolates can identify features that promote adaptation to specific stomach environments, impact tissue interactions, and disease progression. However, few studies have characterized how *H. pylori* specifically interacts with, and may genetically adapt to, the metaplastic tissue environment.

In Chapter 2, we described the J99 collection of isolates, originating from a single individual at two time points during disease. During the six years in between sampling, this patient's disease changed from a duodenal ulcer and inflammation localized to the antrum to corpus inflammation and lower acid secretion resulting from atrophy of the corpus glands. The genetic alterations in these infecting populations reported in Chapter 2, therefore represent adaptation over a period where tissue changes consistent with metaplasia had occurred. Therefore, we can leverage these strains to access their ability to acutely colonize the metaplastic stomach environment in mice. In this study, we utilized a mouse model of KRAS driven metaplasia together with *in-vitro* co-cultures to explore bacterial genotypes that determine niche tropism and colonization in the metaplastic stomach.

4.3 RESULTS

We previously used *Mist1-Kras* mice to explore how *H. pylori* infection could impact metaplasia development in the stomach corpus [160]. In these mice, administration of tamoxifen (TMX) induces expression of an active *Kras* allele (G12D) in the chief cells, leading to rapid onset of spasmolytic polypeptide-expressing metaplasia (SPEM) in the corpus within four weeks, which progresses to intestinal metaplasia (IM) by 12 weeks [161]. In the present study we either induced active KRAS (KRAS+), or sham-induced mice (KRAS-), and then performed acute *H. pylori* infections to assess how *H. pylori* interacted with metaplastic vs. healthy glands (Figure 4.1). We first tested *H. pylori* strain PMSS1, which robustly colonizes the healthy mouse stomach [21,162]. Mice were infected for one week during SPEM (four weeks after KRAS induction), SPEM with intestinalizing characteristics (SPEM-IC, eight weeks after KRAS induction), or IM (12 weeks after KRAS induction), with sham-induced mice as a control at each time point (Figure 4.2A). PMSS1 robustly colonized the KRAS- mice at each time point, in keeping with its previously established ability to colonize wild-type mice. Interestingly, PMSS1 also robustly colonized KRAS+ mice at each time point, demonstrating that this *H. pylori* strain could also survive in the metaplastic stomach. Interestingly, at 12 weeks, some KRAS+ mice showed one-two log higher bacterial loads than sham induced mice. We hypothesized that this could result from *H. pylori* expansion from the antrum into corpus metaplastic glands.

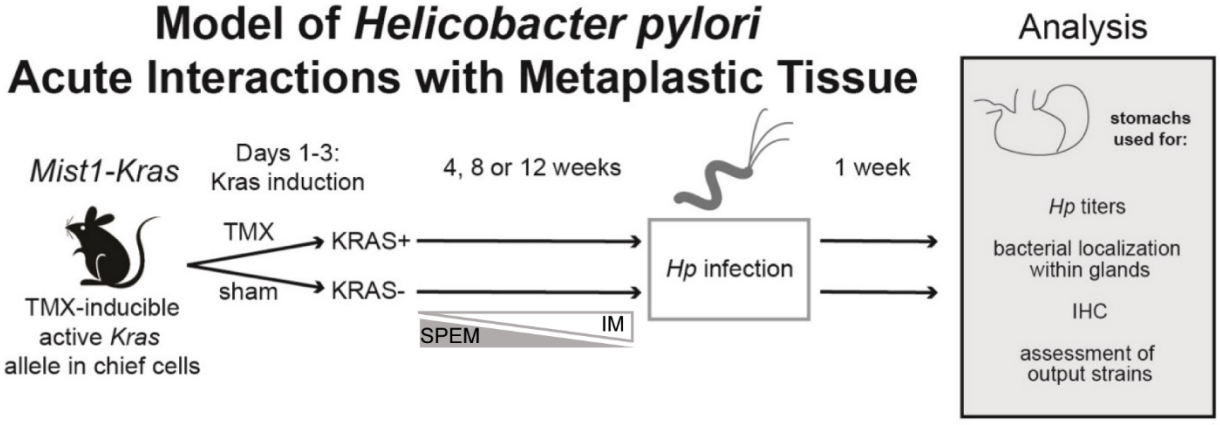


Figure 4.1 Mouse model to assess *H. pylori* interactions with healthy vs. metaplastic tissue.

Mist1-Kras mice were given tamoxifen (TMX) to induce active KRAS in the chief cells, which leads to metaplasia development, or were sham-induced (corn oil). After the onset of SPEM (4 wks), SPEM with intestinalizing characteristics (8 wks), or intestinal metaplasia (12 wks), mice were infected with *Helicobacter pylori* for one week to assess acute *H. pylori*-tissue interactions. IHC, immunohistochemistry.

In humans, *H. pylori* infection initially localizes to the antrum, which lacks acid-producing parietal cells, and then gradually can spread to the corpus (main body of the stomach) over a period of years to decades [32]. A similar phenomenon is observed in mice: PMSS1 localized to the antrum after one week of infection in wild-type C57BL/6 mice and then spread to the corpus within one month [21]. We used immunofluorescence microscopy to detect PMSS1 within thick sections of formalin-fixed, paraffin-embedded tissue (Figure 4.2B-E). We previously used immunofluorescence microscopy to detect and quantify *H. pylori* within gastric glands based on fluorescent voxels [21]. Here we used this method to compare antral and corpus gland occupation between KRAS- and KRAS+ mice infected with

PMSS1 (Figure 4.2E). We examined stomach sections from two to three mice per time point. At each time point, more *H. pylori* cells were present within the glands of KRAS+ mice than KRAS- mice, suggesting that the metaplastic changes to the gland architecture and microenvironment may favor *H. pylori* gland colonization. Consistent with our hypothesis, far more bacteria were detected in corpus glands than antral glands of KRAS+ mice, whereas KRAS- mice maintained the expected antral predominance. Thus, *H. pylori* strain PMSS1 can robustly colonize the corpus during a one week infection if metaplasia has developed.

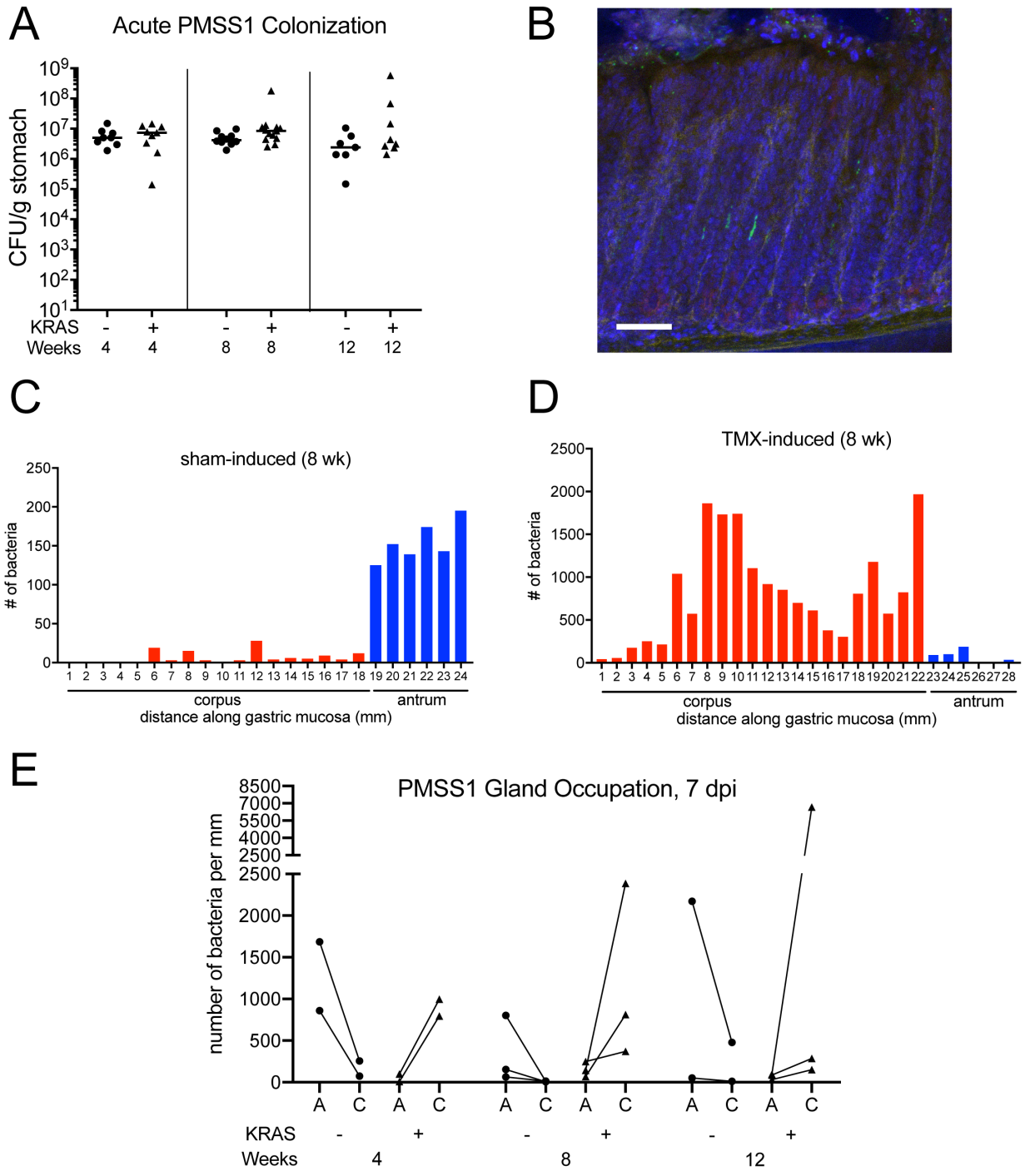


Figure 4.2. *H. pylori* expands colonization to corpus during various stages of gastric pre-neoplastic progression

Transgene expression in Mist1-Kras mice was induced with tamoxifen (TMX), or sham-induced (corn oil). Four, eight or 12 weeks later (corresponding to SPEM, SPEM

with intestinalizing characteristics [SPEM-IC], and IM, respectively), mice were infected with infected with *H. pylori* strain PMSS1 for one week and stomachs were assessed. (**A**) Bacterial titers are shown. CFU, colony-forming units. Data are from N=2 independent experiments with n=7-13 mice per group. (**B-D**) Thick stomach sections (200 μm) were stained for *H. pylori*. Using confocal microscopy, Z-stacks were collected along the entire length of the stomach. Volumetric analysis was used to enumerate bacteria based on fluorescent voxels. (**B**) Representative image of stomach section from the antrum of a sham-induced (oil) mouse infected with *H. pylori* strain PMSS1 for 1 week. Image is a maximum intensity projection of Z stack, with nuclei labeled with DAPI (blue), *H. pylori* (green), and neck cells labeled with GSII in (red). Scale bar, 50 μm. (**C-D**) Example of gland analysis for PMSS1 in sham induced (oil) (**C**) and TMX induced (KRAS+) (**D**) mice after 8 weeks of infection. Graph shows the estimated number of bacteria detected by immunofluorescence within glands along the length of the stomach in microns. Red bars indicate the corpus, and blue bars indicate the antrum. Panels C and D show representative data from one mouse for each condition. (**E**) The sum of bacteria within the antrum ('A') and corpus ('C') of each mouse is shown.

4.3.1 *Differential colonization of metaplastic stomach and glands by genetically diverse H. pylori strains*

Next we tested a panel of *H. pylori* isolates to look for strains that might differ in their ability to colonize the metaplastic vs. healthy stomach (Figure 4.3A). NSH57 is a mouse-adapted version of the well-studied clinical isolate G27 but is a relatively poor colonizer of mice [56,110]. This strain colonized 0/5 KRAS- mice and 1/9 KRAS+ mice in our study. Strain 7.13 is a gerbil-adapted carcinogenic *H. pylori* strain derived from the duodenal ulcer strain B128 [163]. Titers of 7.13 did not differ between KRAS- and KRAS+ mice. However, only 4/7 KRAS- mice and 5/8 KRAS+ mice were colonized, and median titers were three to four logs lower than titers of PMSS1, demonstrating that

7.13 is overall not a robust colonizer of mice, regardless of gastric metaplasia status. Thus induction of metaplasia does not enhance stomach colonization by these strains.

We next tested infection of representative isolates from the J99 strain collection of multiple *H. pylori* isolates from a single human host, collected at two time points spanning six years [62,64]. Notably, at the 1994 time point the patient had a duodenal ulcer, which progressed to atrophic gastritis (a precursor to metaplasia) by the time of the 2000 sampling [164]. Based on shared genetic variation, these isolates fall into 4 distinct subgroups represented by J99 (1A) and SC4 (1B) both isolated in 1994 and D1 (2A) and C2 (2B) isolated from 2000 [165].

The J99 strain was a relatively poor colonizer of mice, with titers recovered in only 4/10 KRAS- mice and 3/11 KRAS+ mice (Figure 4.3A). Strain SC4, which belongs to the same ancestral subgroup as strain J99 [165], was also a relatively poor colonizer of KRAS- mice (recovered from 2/6 KRAS- mice) and more variable in KRAS+ mice (recovered from 2/3 KRAS+ mice). Strain D1 was recovered from 8/11 KRAS- mice and 12/13 KRAS+ mice, demonstrating that it could robustly colonize mice. D1 titers were about one log higher in KRAS+ mice than in KRAS- mice ($P < 0.01$, Mann-Whitney U test), suggesting that this strain preferentially colonized the metaplastic stomach. The other strain tested from the later time point, C2, was recovered from only 5/10 KRAS- mice and 5/10 KRAS+ mice. We observed that the C2 strain was the most heterogeneous in overall bacterial burdens, with titers ranging from undetectable to as high as almost 10^7 CFU/g stomach. We thus hypothesized that C2 might be poised to adapt to the metaplastic stomach environment.

To test this hypothesis, KRAS+ mice were infected with the C2 “stock” strain (as in Figure 4.3A) or a C2 “adapted” strain saved from a KRAS+ mouse shown in red in Figure 4.3A. Strikingly, titers of the C2 adapted strain were far greater than the C2 stock strain in KRAS+ mice (Figure 4.3B). The C2 stock strain was only recovered from 1/4 mice, whereas the adapted strain was recovered from 6/6 mice. Thus, C2 appears poised to adapt *in-vivo* to efficiently colonize the metaplastic stomach.

Because PMSS1 exhibited robust colonization of corpus glands in KRAS+ mice (Figure 4.2E, Figure 4.3A), we tested whether other *H. pylori* isolates could also colonize KRAS+ corpus glands. Due to decreased bacterial loads detected across the other strains compared to PMSS1, the signal was too weak to detect and quantify using fluorescent voxels from thick sections as in Figure 4.2. However, *H. pylori* could be visualized within the corpus glands in KRAS+ mice across all strains. Thin stomach sections from mice infected with different *H. pylori* strains were stained with an anti-*H. pylori* antibody and corpus tissue was semi-quantitatively scored in a blinded fashion, where 0 indicates no bacteria detected in the corpus, 1 indicates few bacteria detected in the corpus, and 2 indicates moderate to abundant bacteria detected in the corpus (Figure 4.4A). Only mice with detectable CFU were included in these experiments. PMSS1 robustly colonized KRAS+ corpus glands and bacteria were abundant (Figure 4.2A). Few bacteria of any strain tested were seen in the antral glands (not shown).

The J99 and SC4 strains were poor colonizers of KRAS+ corpus glands (Figure 4.4). Interestingly, the D1 strain fared much better, with bacteria seen in corpus glands of about 80% of mice tested, though gland colonization was not as robust as with strain

PMSS1. Again, the C2 strain was the most heterogeneous, with 3/6 mice exhibiting no bacteria in corpus glands, 1/6 exhibiting a few bacteria throughout the corpus and 2/6 exhibiting moderate or abundant bacteria throughout the corpus. The C2 adapted isolate established colonization in a greater proportion of mice. But in mice that were colonized, both stock and adapted C2 strains were detected in the corpus glands of mice with titers above the limit of detection (Figure 4.4), suggesting that the higher titers in the adapted C2 strain could reflect enhanced ability to colonize the corpus glands within the metaplastic stomach.

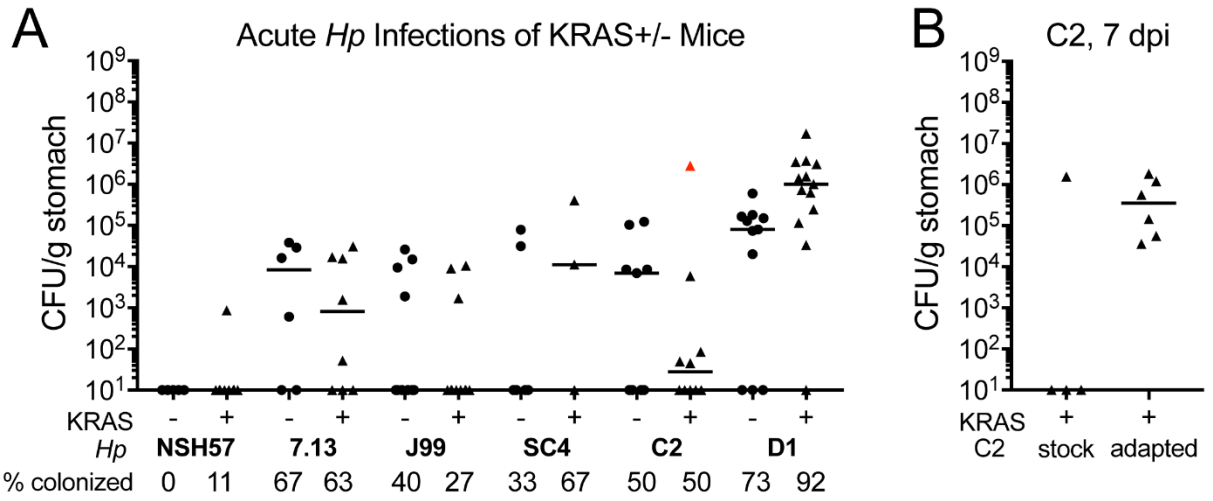


Figure 4.3. Diverse *H. pylori* isolates exhibit different propensities for colonization of the metaplastic vs. healthy stomach.

Mist1-Kras mice were sham-induced or induced with TMX. After eight weeks, mice were infected with *H. pylori* for one week. (**A**) Shown are stomach titers after one week of infection. Data points represent actual values from each individual mouse, bars represent the median, and mice with no detectable CFU are plotted at the limit of detection. Percentage of total mice infected is listed below isolate name. Each *H. pylori* strain was tested in N=1-3 independent experiments with at least 3 mice per group. The mouse from which the C2 "adapted" strain was obtained is indicated in red. (**B**) Recent isolate C2 can be adapted to colonize metaplastic stomach. *Mist1-Kras* mice

were induced with TMX. After eight weeks, mice were infected with either the C2 stock or adapted strain for one week. Data points represent actual values from each individual mouse, bars represent the median, and mice with no detectable CFU are plotted at the limit of detection.

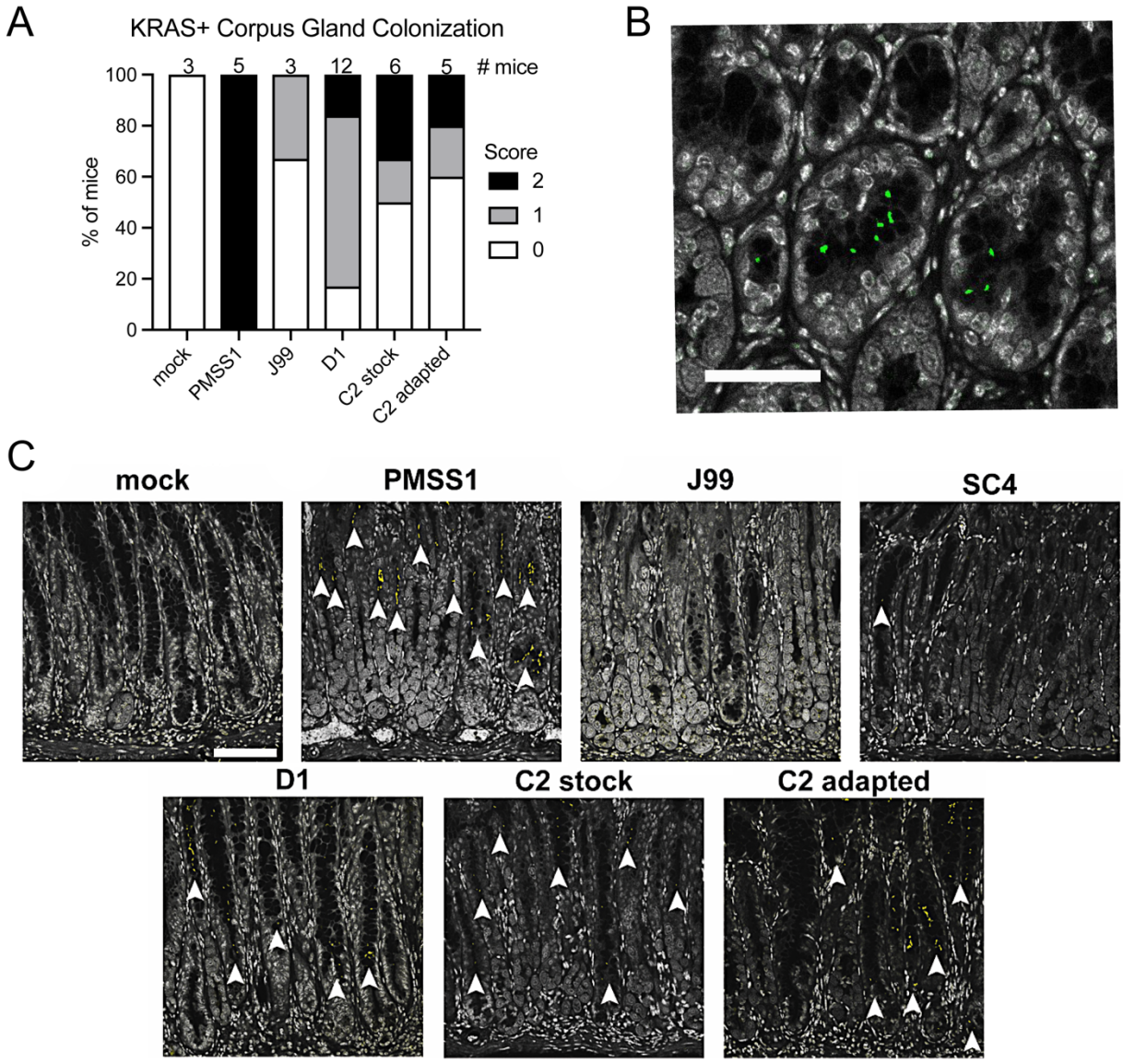


Figure 4.4. Strains that established colonization in metaplastic stomach after 1 week can be detected in corpus glands at different densities.

Thin stomach sections (4 μm) were stained for *H. pylori* and the corpus was assessed for bacterial colonization. (**A**) Shown is a corpus gland colonization score for each *H. pylori* strain tested. The number of mice examined is indicated at the top of the graph. (**B**) Representative image of a colonized gland is shown. Grey indicates nuclei

(stained with DAPI) and green indicates *H. pylori* cells (stained with anti-*H. pylori* antibody). Scale bar, 50μm. (**C**) Representative images of corpus tissue are shown. Grey indicates nuclei and yellow indicates *H. pylori* cells. Arrowheads show examples of colonized glands. Scale bar, 100 μm.

4.3.2 *Sequence comparisons of isolates D1, C2 stock, and C2 adapted reveal mutations in outer membrane adhesins*

Isolates from the J99 collection exhibit differential colonization patterns in KRAS induced and uninduced mice, despite having relatively few differentiating polymorphisms compared to isolates from different individuals. Both D1 and C2 have been sequenced in Chapter 2 and short read sequence data is publicly available [165]. Therefore in order to investigate potential genetic determinants of colonization in the KRAS- and KRAS+ mice, we examined all 349 polymorphic loci between isolates D1 and C2. A total of 63 of these variant positions were detected within intergenic regions. For this study, we focused on the 175 polymorphisms that resulted in alterations to amino acid sequence, which includes 123 nonsynonymous polymorphisms and 52 indels detected across 68 genes. Because *H. pylori* is highly recombinogenic and encodes several paralogous proteins, we predicted that many of these alterations reflect inter or intra genomic recombination events or mis-mapping of short reads to paralogs with high sequence identity. Therefore we leveraged ClonalFrameML to bioinformatically predict genes with mutations predicted to be introduced through importation of divergent DNA [98]. As expected, more than half of these alterations (65%) were predicted have been introduced via recombination events and were detected across only 14 genes. Polymorphisms in these genes were validated separately and excluded from the initial variant list (Table 4.3). The remaining 46 nSNPs

and 16 indels differentiating D1 and C2 are listed in Table 4.1 and Table 4.2. Genes with polymorphic loci have a range of functions. There are mutations in two fucosyltransferases--*fucA* and *fucU,* which are involved in modifying bacterial outer membrane proteins to mimic surface sugars on the host epithelium in order to evade immune detection [166]; and several mutations in genes involved in DNA uptake, metabolism, and repair including *comB8*, *comM*, *hsdR*, *hsdM*, and *hsdS*. It remains unclear which mutations or combination of mutations contribute to differences in colonization loads and penetration into the glands observed between strains D1 and C2.

Due to limitations mapping short read data to repetitive regions in the reference, we generated long-read assemblies of isolate D1 using Pacific Biosciences (PacBio) sequencing technology [167]. Long read assemblies revealed some of these polymorphisms predicted to be introduced through recombination were a result of short reads misaligning to the reference strain in repetitive regions. However, 10 of the 14 genes did have mutations or significant structural differences that were detected using both short and long read data (Table 4.3). Six out of ten encode outer membrane proteins, including *sabA, sabB, homA, hopJ, hopK, and hopQ*. Genes jhp0440, jhp1031, and *hopQ* had significantly altered protein lengths. Isolate D1 had an alternate start site rendering the protein 17 amino acids shorter than the C2 *hopQ* protein.

The adapted C2 isolate, which colonized the metaplastic stomach more robustly than stock C2, was also sequenced using the PacBio long read platform to determine adaptations that may have occurred during passage in the KRAS+ mouse stomach. The only polymorphic loci between strain C2 stock and C2 adapted is in outer membrane protein SabB. SabB has paralogs SabA and HopQ, which facilitate binding to sialic acid

and CEACAM, respectively [87,168]. Gene conversion events and phase variation introducing ON and OFF variants is thought to modulate adherence to inflamed tissue during the course of chronic infection [169,170]. The C2 adapted variant of *sabB* has 6 mutations in the C-terminal region that are shared with the D1 variant of *sabB*, but not C2 stock. Because they are clustered in one genomic region, they likely arose from a single recombination event rather than de-novo mutation. The altered bases within *sabB* in C2 adapted align to its paralog *sabA* in the C2 stock, indicating that these changes were most likely introduced through an intra-genomic recombination event between *sabA* and *sabB*. Only one of these polymorphisms results in an amino acid change—a mutation from threonine, an amino acid with a polar side chain, to alanine an amino acid with a hydrophobic side chain at position 553.

Table 4.1 Nonsynonymous SNPs between isolates D1 and C2 (n=46)

Gene ID	Annotation	Description	nSNPs
jhp0034	*comB8*	DNA transformation competency	1
jhp0066	*ureI*	urease accessory protein (urea transporter)	1
jhp0104	*fucA*	L-fuculose-1-phosphate aldolase	1
jhp0116	*rpl20*	50S ribosomal protein L20	1
jhp0139		*unknown*	1
jhp0151	*arsS*	signal-transducing protein, histidine kinase	1
jhp0267		*unknown*	1
jhp0296		*unknown*	1
jhp0314	*minD*	cell division inhibitor	1
jhp0325	*fliF*	flagellar basal-body M-ring protein	1
jhp0334	*kgtP*	alpha-ketoglutarate permease	1
jhp0402		*predicted 5'-3' exonuclease*	1
jhp0429	*hopJ*	putative outer membrane protein	1
jhp0468		*unknown*	1
jhp0495	*cagA*	cag pathogenicity island protein, cytotoxin immunodominant antigen	2
jhp0554	*acrB*	acriflavine resistance protein, putative efflux transporter	1
jhp0575	*hydB*	quinone-reactive Ni/Fe hydrogenase, large subunit hydrogenase	1

jhp0620	*wbpB*	putative lipopolysaccharide biosynthesis protein	1
jhp0654		predicted tRNA threonylcarbamoyladenosine biosynthesis protein	1
jhp0728	*comM*	predicted DNA transformation competence	1
jhp0785	*hsdS*	type I restriction enzyme (specificity subunit)	1
jhp0809	*katA*	catalase	1
jhp0899	*lgt*	prolipoprotein diacylglyceryl transferase	1
jhp0929		plasticity region	1
jhp0934		plasticity region	2
jhp0935		plasticity region	1
jhp0936		plasticity region	1
jhp0938		plasticity region	1
jhp0956		plasticity region	1
jhp0967	*metS*	methionyl-tRNA synthetase	1
jhp0982	*rps1*	30S ribosomal protein S1	1
jhp1002	*fucU*	alpha-(1,3)-fucosyltransferase	3
jhp1065	*atpF'*	ATP synthase F0, subunit b' ATP synthase B'	1
jhp1083	*hopI*	putative outer membrane protein	1
jhp1121	*rpoB*	DNA-directed RNA polymerase, beta subunit	1
jhp1164	*babB*	outer membrane protein, adhesin	1
jhp1178	*pyrE*	orotate phosphoribosyltransferase	1
jhp1359		predicted ABC transport system permease	1
jhp1371	*uvrD*	putative DNA helicase II	1
jhp1392		unknown	1
jhp1423	*hsdM*	type I restriction enzyme (modification subunit)	1
jhp1424	hsdR	type I restriction enzyme (restriction subunit)	1

Table 4.2 Insertion or Deletion Events between isolates C2 and D1 (n=16).

Gene ID	Annotation	Description	Indels
jhp0151	*arsS*	signal-transducing protein, histidine kinase	1
jhp0241		conserved hypothetical secreted protein	1
jhp0451	*pldA*	putative phospholipase A1 precursor (DR-phospholipase A)	1
jhp0709	*amiA*	putative N-acetylmuramoyl-L-alanine amidase	1
jhp0741	*lex2B*	Lipoligosaccharide 5G8 epitope biosynthesis associated protein	1
jhp0888		oxygen-insensitive NAD(P)H nitroreductase	1
jhp0892		unknown	1
jhp0928		plasticity region	1
jhp0934		plasticity region	1
jhp0937		plasticity region	1

jhp0938		plasticity region	1
jhp1002	*fucU*	alpha-(1,3)-fucosyltransferase	1
jhp1171	*csd5*	Cell-shape determining protein	1
jhp1305		unknown	1
jhp1312		unknown	1
jhp1410		type III restriction enzyme (restriction enzyme)	1

Table 4.3. Genes with high sequence divergence between isolates D1 and C2 due to recombination

Gene ID	**Annotation**	**SNPs**	**nSNPs**	**Indels**
jhp1300		2	2	3
jhp0659	*sabB*	13	6	0
jhp0662	*sabA*	2	1	1
jhp0857	*hopK*	1	0	0
jhp0870		55	22	3
jhp0649	*homA*	55	18	3
jhp0429	*hopJ*	2	0	0

Genes with alternate stop and start sites

Gene ID	**Annotation**	**Description**
jhp1031	jhp1031	alternate stop, truncated protein in isolate D1
jhp0440	jhp0440	alternate stop, truncated protein in isolate D1
jhp1103	*hopQ*	alternate start site, truncated protein in isolate D1

4.3.3 *Select isolates from J99 collection sub-group 2A can induce novel rounding morphology in host cells dependent on a secreted factor*

Isolate D1 colonizes the healthy and metaplastic stomach at higher titers than the C2 stock and J99 isolate. The adapted strain C2, which colonizes KRAS+ mice at similar titers to D1, also appears have altered gland colonization compared to D1. This suggests that these isolates have fundamental differences in the ways in which they interact with host tissues. In order to parse out these differences, we examined host-pathogen interactions in an *in vitro* co-culture model of infections with the AGS gastric cancer cell line. Gastric cell lines cultured with *H. pylori* strains that harbor a CagPAI

typically exhibit altered morphology characterized by cell elongation and scattering, reflecting perturbations to host-signaling by bacterial cytotoxin, CagA [137]. All the isolates in the J99 collection, including both C2 and D1, have a functioning type IV secretion system and were able to induce cell elongation. However, at early time points, a few isolates within subgroup 2A, induced another novel cell rounding morphology in AGS cells, characterized by scattering without elongation that was not observed in AGS cells co-cultured with isolate C2 or J99 (Figure 4.5A). The phenotype was not completely penetrant with only a fraction of AGS cells exhibiting the rounding phenotype at any one time. Additional isolates originating from the same subgroup as D1 were also subject to co-culture and several isolates within subgroup 2A in addition to D1 were able to induce this phenotype including C11, C5, C9, A9, and C1. Rounded cells have overall decreased cell length compared to cells exhibiting the hummingbird phenotype and uninfected cells, therefore the phenotype was measured by taking length measurements of a population of AGS cells from images after 6hrs of incubation with *H. pylori*. We found that strain C11, a strain related to D1 and residing in the same subgroup (2A), strongly induces the rounding phenotype and we showed average cell length is decreased in cells treated with strain C11. Incubation of AGS cells with cell free bacterial supernatants from isolate C11 also show rounding, suggesting rounding is dependent on a secreted bacterial factor (Figure 4.5B).

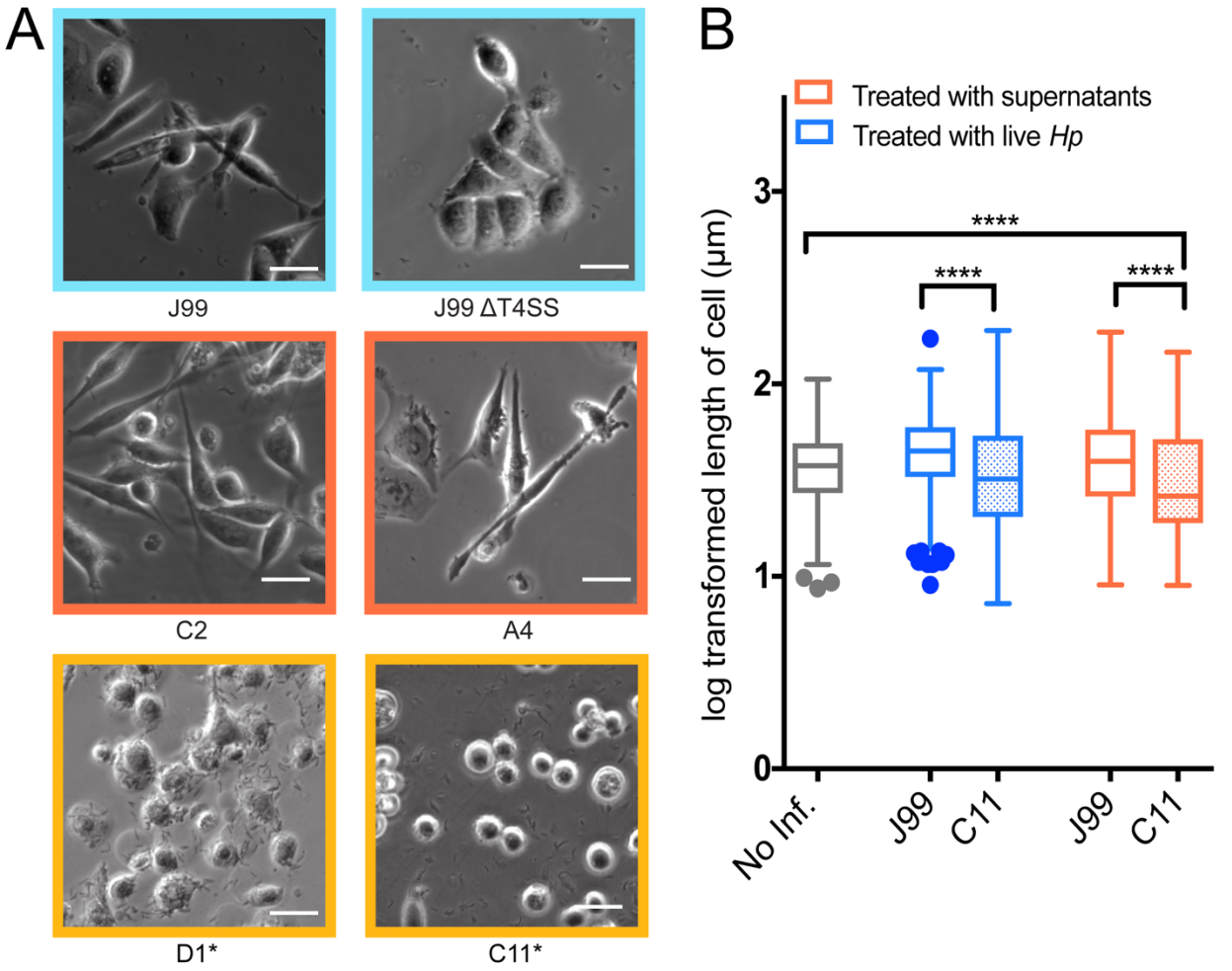


Figure 4.5. Novel host rounding phenotype induced by certain isolates is dependent on a secreted bacterial factor

(**A**) Each condition represents infected AGS cells (MOI =10) after 6hrs. Top row is J99 and J99 with inactive type IV secretion system (J99 Δ*cagE*). Middle and bottom rows are representative isolates from genetically distinct subgroups (2A= yellow, 2B= red). Cells displaying rounded phenotype are marked with an asterisk. A=antral isolate, C= corpus, and D=duodenal isolate. Scale bars= 10 μm. (**B**) Cell rounding quantified by median cell length of AGS cells with co-culture media only (no inf.), incubated with live *H. pylori* (blue) and *H. pylori* filtered supernatants (red) from a rounding (C11) and non-rounding isolate (J99). Cell length of control uninfected AGS cells treated with co-culture media alone are in gray. All values were log transformed to achieve a normal distribution. P-values were calculated using a student's t-test (****, p-value <0.0001). Results shown as a box plot encompassing 25th-75th percentile values and a line

indicating median log transformed length values from >350 cells collected from a single experiment using J99 and C11 as representative isolates for each phenotype.

We then tested to see if rounded cells were indicative of apoptosis or general cell death measured with Annexin/PI staining. However, we did not see increases in annexin positive cells at 6 or 24hrs post infection. We also did not observe any increase in lactate dehydrogenase (LDH) in co-culture supernatants indicative of other forms of cell death over a 48hr period of infection (Figure 4.6A). To identify transcriptional changes that accompany rounding, we performed an RNA-seq experiment comparing transcriptional responses of AGS cells treated with strain J99, a non-rounding isolate, and C11, a rounding isolate for 6 hours. The results identified unique gene expression programs in AGS cells exposed to a cell rounding isolate, suggesting distinct isolates within the collection induce disparate host transcriptional responses (Figure 4.6B-C). Some of the responses such as the increase in expression of inflammatory cytokines--CXCL8, CXCL6, CXCL1 in cells treated with C11 are likely due to increased function of the Cag T4SS, which we have previously shown [165]. However, many additional genes are differentially expressed, including ANKDR1 and CTGF which are both downstream targets of the host transcriptional regulator, Yap1 [172].

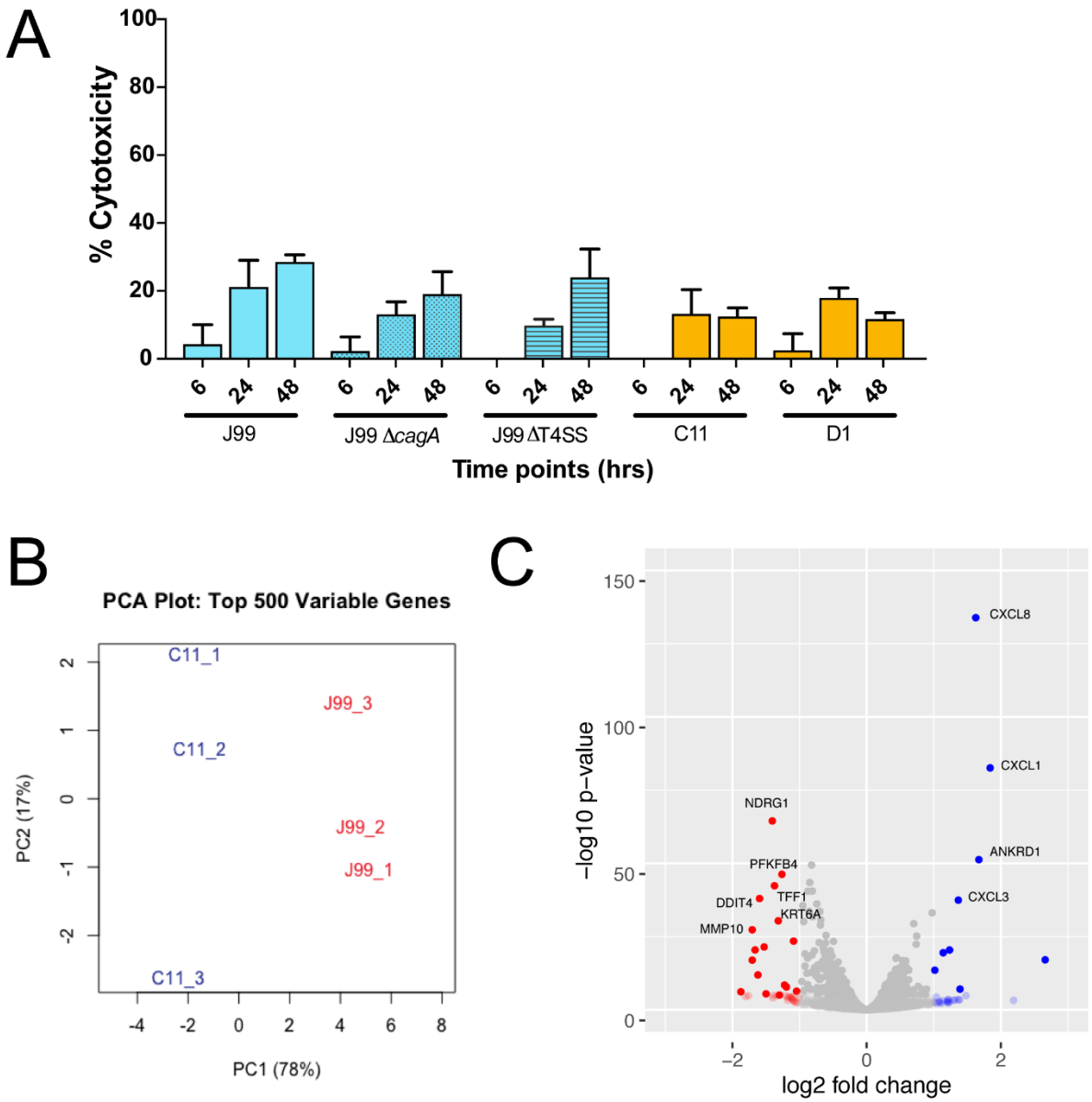


Figure 4.6. Rounding isolate induces unique transcriptional response in AGS cells that is not reflective of increased cell death

(**A**) Percent cytotoxicity measured by lactate dehydrogenase (LDH) release from AGS cells infected with J99, J99 Δ*cagA*, J99 ΔT4SS (blue), and recent isolates C11 and D1 (yellow) after 6, 24, and 48 hrs of infection. Percent cytotoxicity is calculated relative to a complete cell lysis control so that each value shows percentage cell death in the plate well. Error bars represent standard deviation from mean. Mean values were calculated from triplicate wells. (**B**) Principle component analysis on the top 500 differentially

expressed genes after 6 hrs treatment with rounding isolate (C11) and non-rounding isolate (J99). Each point represents an experimental replicate from bulk RNA-seq on AGS cells infected with C11 (blue) and J99 (red). (**C**) Volcano plot showing differential expression of C11 infected cells compared to J99 infected cells. Points highlighted in blue are significantly upregulated in C11 compared to J99 and points highlighted in red are significantly downregulated in C11 compared to J99. The X-axis represents the $\log_2$ fold change and the Y-axis denotes the negative $\log_{10}$ p-value. Labeled genes have FDR corrected p-value < 1E-25 and $\log_2$ fold change greater than 1 or less than -1.

4.4 DISCUSSION

It has been well established that prolonged *H. pylori* infection can induce premalignant tissue alterations, but little is known about the dynamics between host and pathogen at later stages of disease development. In this study we examined *H. pylori* strain specific determinants of stomach and gland colonization during metaplastic development utilizing the Mist1-Kras mouse model.

We found that *H. pylori* strains have different propensities for infecting +/- KRAS mice. Strains that can robustly colonize KRAS+ mice have an expanded niche in the corpus glands. This expansion of *H. pylori* into corpus during acute infections in the KRAS+ mice is similar to what is observed with long term (>4wks) chronic infections in mice [21,157]. The ability to infect KRAS+ mice can be genetically acquired by *H. pylori* during acute infection of metaplastic tissue.

J99 strains, originating from the same patient at two different stages of disease [65,165], exhibit differential colonization in KRAS+ mice. J99 is a relatively poor colonizer of both induced and uninduced KRAS mice. Strain SC4, isolated at the early timepoint but belonging to a different genetic subgroup from J99, may colonize KRAS+

more efficiently than J99, but additional experimental replicates will need to be performed. Isolate D1, collected from the patient during a period of gastric atrophy, robustly colonized KRAS+ mice. Surprisingly, isolate D1 could colonize the KRAS+ stomach at higher loads than healthy, uninduced mice. In contrast, isolate C2, originating from the same time point as D1, had relatively poor colonization in some mice but very robust colonization in others. We showed C2 could be adapted to colonize KRAS+ mice via a single one week passage. Additional follow up experiments will determine if C2 adapted strain also preferentially colonizes KRAS induced mice.

In human infections, *H. pylori* triggers metaplasia, perhaps as an unintended consequence of prolonged bacterial growth and inflammation. Once this gastric atrophy and metaplasia is initiated, it is thought that *H. pylori* does not further contribute to the development of cancer and that these tissue alterations are not favorable to *H. pylori* proliferation [155,173]. This is referred to as the 'hit and run' theory of gastric cancer development [103]. This implies eradication of *H. pylori* infection is unnecessary once metaplastic lesions have developed. However, a long-term study with a Colombian cohort shows that eradication of *H. pylori* after discovery of metaplastic legions reduces risk of gastric cancer, suggesting a new role for *H. pylori* in the development or acceleration of gastric cancer [174]. Host-pathogen interactions during pre-malignant tissue alterations are understudied, but in a recent study we showed that *H. pylori* interacts with tissue in ways that impact disease progression and skews immune response [160]. Other animal models which induce gastric atrophy and SPEM in the absence of inflammation via chemical ablation of parietal cells, show that *H. pylori* may preferentially interact with SPEM cells within the corpus glands [175]. Here we show

certain strains of *H. pylori* readily colonize the metaplastic stomach despite a substantial immune response triggered by genetic activation of KRAS and exhibit an expanded niche in the corpus glands. Together these data suggest that metaplastic environments may be suitable or even preferable for *H. pylori* growth and challenge the hit-and-run theory of carcinogenesis. This new model provides avenues to further explore *H. pylori* interactions with metaplastic tissue and cell types specific to metaplastic glands.

Not all of the clinical isolates we tested colonized the metaplastic stomach of KRAS+ mice, indicating there are bacterial determinants of metaplastic colonization. The host environment where clinical strains were isolated from may make certain isolates better adapted for colonization. Whole genome sequencing suggests that changes in the outer membrane protein SabB facilitated quick adaptation to the metaplastic stomach environment in isolate C2. SabB is a protein that is part of a paralogous family including HopQ, which binds to CEACAM, and SabA which bind to sialic acid. These proteins are some of the most highly variable in the *H. pylori* genome and can recombine to quickly adapt to changes to the gastric environment [88,170]. Inflamed tissue is more likely to express sialylated glycoconjugates than healthy tissue [168], which may suggest that *H. pylori* genetically alters *sabA* and *sabB* in order to modulate binding to inflamed tissue. Our data suggest, that *sabB* may play an unknown role in achieving colonization of tissue that is inflamed or exhibiting SPEM, however it is likely not the only player.

Differential colonization may reflect important differences in the ways that isolates interact with host tissues during different stages of disease, which could have implications for disease progression, intervention, and treatment strategies. *In vitro* co-

culture experiments with isolates from the J99 collection showed J99, C2 and D1, induced disparate host morphological responses. Isolate D1 and other related isolates from the same subgroup, appear to induce a rounded morphology in host cells dependent on a secreted bacterial factor. Isolate C2 did not exhibit this phenotype. Isolates D1 and C2 were previously found to have different phenotypic and genotypic properties [165], so the additional phenotypes identified strengthen the hypothesis that these isolates they likely occupy two different niches within the stomach.

It is remains unknown if *H. pylori* has evolved to exploit pre-neoplastic tissue changes such as SPEM, IM, and dysplasia in order to expand its niche and increase transmission, or if it is an unintended consequence of bacterial growth that is tolerated but not advantageous. Our findings suggest certain strain factors may promote colonization of the metaplastic stomach and these factors can vary between isolates in a single individual during the course of infection. Additional studies of bacterial genetic factors facilitating colonization during SPEM may be able to identify important predictors for risk of transmission or disease progression.

4.5 METHODS

Ethics Statement

All mouse experiments were performed in accordance with the recommendations in the National Institutes of Health Guide for the Care and Use of Laboratory Animals. The Fred Hutchinson Cancer Research Center is fully accredited by the Association for Assessment and Accreditation of Laboratory Animal Care and complies with the United States Department of Agriculture, Public Health Service, Washington State, and local

area animal welfare regulations. Experiments were approved by the Fred Hutch Institutional Animal Care and Use Committee, protocol number 1531.

4.5.1 *Culture conditions for H. pylori*

All *H. pylori* isolates were grown on solid media, horse blood agar (HB agar). HB agar plates contain 4% Columbia agar base (Oxoid, Hampshire, UK), 5% defibrinated horse blood (Hemostat Labs, Dixon, CA), 10 mg/ml vancomycin (Thermo Fisher Scientific, Waltham, MA), 2.5 U/ml polymyxin B (Sigma-Aldrich, St.Louis, MO), 8 mg/ml amphotericin B (Sigma-Aldrich), and 0.2% β-cylodextrin (Thermo Fisher). For HB agar plates used to grow *H. pylori* from homogenized mouse stomach, 5 mg/l cefsulodin (Thermo Fisher), 5 mg/l trimethoprim (Sigma) and 0.2 mg/ml of Bacitracin (Acros Organics, Fisher) are added to prevent outgrowth of mouse microflora. Shaking liquid cultures were grown in Brucella broth (Thermo Fisher Scientific, Waltham, MA) supplemented with 10% heat inactivated FBS (Gemini BioProducts, West Sacramento, CA).

4.5.2 *Mist1 Kras breeding, induction of transgene, and infection*

Mist1-CRERT Mist1-CreERT2 Tg/+; LSL-K-RAS G12D Tg/+ “Mist1-Kras” mice described previously [121] were obtained from University of Vanderbilt to establish a breeding colony at Fred Hutchinson Cancer Research Center [160]. Pairs of Mist1-CreERT2 Tg/+, KRAS G12D Tg/+ mice were bred in house to achieve cohorts with the Mist1-Kras genotype (Mist1CreERT2 tg/+, Kras G12D tg/+). Mice were housed in sterilized microisolator cages with irradiated rodent chow, autoclaved corn cob bedding,

and acidified, reverse-osmosis purified water. Mice were genotyped from ear punches as previously described [160]. Expression of the KRAS transgene was induced via subcutaneous administration 5 mg of tamoxifen (Sigma Aldrich) in corn oil (Sigma Aldrich) for three consecutive days. Sham induced mice were administered corn oil subcutaneously on the same schedule.

All mouse experiments were performed as previously described [160]. The inoculum for each infection was 5 x 10^7 cells. After stomach excision, the forestomach was removed and opened along the lesser curvature. Stomachs were divided in equal thirds or halves containing both antral and corpus regions. For culture stomachs sections were place in 0.5 mL of sterile BB10 media, weighed, and homogenized. Serial homogenate dilutions were plated on nonselective HB plates. After 5-9 days in tri-gas incubator, colony forming units (CFU) were enumerated and reported as CFU per gram of stomach tissue. P-values were calculated from pooled experimental replicates using a Mann-Whitney non-parametric test. For histologic analysis, stomach sections were fixed in 10% neutral-buffered formalin.

4.5.3 *Gland colonization analysis in thick stomach sections*

PMSS1 gland colonization was assessed as previously described [21] with the following modifications. One third of the stomach was embedded in 4% agarose in 1x phosphate-buffered saline (PBS, Gibco) and cut into 200 µm thick longitudinal sections using a Leica VT1200S Vibratome. Tissue sections were then permeabilized overnight at 4°C by gently rocking in blocking buffer comprising 3% bovine serum albumin (Sigma Aldrich), 1% saponin (Sigma Aldrich) and 1% Triton X-100 (Sigma Aldrich) in phosphate

buffered saline (PBS). Stomachs were incubated with 1:1,000 rabbit polyclonal anti-Hp PMSS1 antibody (gift of Manuel Amieva, Stanford University) and 1:2,000 GS-II 488 (conjugated lectin from *Griffonia simplicifolia*, Fisher) for two hours at 4^{o}C with gentle rocking. After three ten-minute washes in PBS, samples were incubated with 1:2,000 Alexa Fluor 647 donkey anti-rabbit IgG (Invitrogen) and 1:2,000 DAPI for two hours at room temperature with gentle rocking. After three ten-minute washes in PBS, sections were mounted onto glass slides with imaging spacers cut from Parafilm (Bemis) in ProLong Diamond Anti-Fade Reagent (Molecular Probes) and coverslips were sealed with VaLP (1:1:1 Vaseline:Lanolin:Paraffin). Stomach sections were imaged with an UltraView spinning disk microscope (PerkinElmer). Z-stacks were collected to visualize *H. pylori* within glands, and maximum intensity projections were generated. Z-stacks were assessed in Volocity (Quorum Technologies) and *H. pylori* within glands were enumerated based on fluorescent voxels as previously described [21].

4.5.4 *Immunofluorescence microscopy of thin stomach sections*

Tissue was deparaffinized with Histo-Clear solution (National Diagnostics) and rehydrated in decreasing concentrations of ethanol. Slides were boiled in a pressure cooker for 15 min in 10 mM sodium citrate (Thermo Fisher Scientific) for antigen retrieval. Slides were incubated with Protein Block, Serum Free (Agilent Dako) for 90 min at room temperature. The primary antibody (1:1,000 rabbit polyclonal anti-*H. pylori* PMSS1 antibody, a gift of Manuel Amieva, Stanford University [59]) was diluted in Protein Block, Serum Free and applied to the slides overnight at 4°C. After washing, the secondary antibody (1:500 Alexa Fluor 647 donkey anti-rabbit IgG, Invitrogen) and GSII

488 (1:2,000 conjugated lectin from *Griffonia simplicifolia*, Fisher) were diluted in Protein Block, Serum Free and slides were incubated for one hour at room temperature protected from light. Slides were mounted in ProLong Diamond antifade reagent with DAPI (Invitrogen) and allowed to cure for 24 hours at room temperature before imaging. Slides were imaged on a Zeiss LSM 780 laser-scanning confocal microscope using Zen software (Zeiss). The entire length of each section was inspected for *H. pylori* cells and three to five representative images of the corpus were taken.

4.5.5 *PacBio long read sequencing*

Genomic DNA from to be sequenced was purified using the Wizard Genomic DNA Purification Kit (Promega, Fitchburg, WI), concentration was determined using QuBit dSDNA HS (High Sensitivity) assay kit (Thermo Fisher), and purity was calculated using Nanodrop Spectrophotometer 1000. Libraries were generated with SMRTbell Express Template Prep Kit 2.0 and sequenced on a PacBio Sequel instrument.

4.5.6 *Bioinformatic analysis*

Short reads for isolates C2 stock and D1 were downloaded from the NCBI SRA database (BioProject accession: PRJNA622860). Single nucleotide variants differentiating D1 and C2 stock were determined by aligning short reads to J99 reference (AE001439) with BreSeq v0.35.0 software using default parameters [112]. ClonalFrameML was used to predict putative sites of recombination. Select isolates, D1 and C2 adapted, were sequenced on a PacBio Sequel instrument to generate long sequence reads. Closed reference sequences were generated using either SMRT Link web-based analysis suite <https://www.pacb.com/support/software-downloads/> or Flye

<https://github.com/fenderglass/Flye> assembly pipelines and genomes were annotated using Pathosystems Resource Integration Center (PATRIC, <https://www.patricbrc.org>) annotation tool with *Helicobacter pylori* (species ID: 210) as the reference. Recombination events predicted by short read assemblies were validated by cross comparing with sequences generated from long read assemblies. Genomic comparisons of C2 stock isolate and C2 adapted isolates was performed by aligning C2 stock short reads with Breseq v0.35.0. using C2 adapted PacBio assembly as the reference.

4.5.7 *Co-culture with AGS cells and cell rounding quantification*

AGS cells, from a human gastric adenocarcinoma cell line (ATCC CRL-1739), were grown in Dulbecco's modified Eagle's medium (DMEM) (Thermo-Fisher) supplemented with 10% heat-inactivated FBS (Gemini-Benchmark). For co-culture with *H. pylori*, AGS cells were seeded at 1 × 105 cells/well in 24-well plates 16 hrs prior to infection. The day of infection, medium was removed from AGS cells and mid-log-phase (optical density at 600 nm (OD) 0.3-0.6) *H. pylori* resuspended in DMEM–10% FBS–20% Brucella broth was added at multiplicity of infection of 10:1. To collect supernatants only, *H. pylori* was grown to mid-log-phase, cells were pelleted by centrifugation at 3000 rpm in a Hettich Rotana 460R swinging bucket centrifuge, and supernatants were collected. Collected supernatants were then filtered through a 0.22 micron filter (VWR, Radnor, PA) to remove any additional live bacteria and applied to cells at a concentration 4x that of inoculation containing live bacteria to mimic predicted doublings that would occur after 6 hrs of infection. At least 5-7 Images were acquired with phase

contrast microscope at 100x magnification (10x objective, 10x ocular). Cell length measurements were collected from >100 cells and measured using Fiji image analysis software. Absolute length measurements were log transformed to achieve a normal distribution. LDH release assays were performed using the co-culture conditions described above and measured at 6, 24, and 48 hrs post infection using the Promega CytoTox 96 Non-Radioactive Cytotoxicity Assay (Madison, WI) according to manufacturer's instructions. Percent cytotoxicity was calculated relative to a complete lysis control.

4.5.8 *RNA sequencing*

Co-culture with AGS cells was set up as described above with three replicate wells per isolate. After 6 hrs of infection, adherent cells were gently removed from well using a cell scraper (Fisher) and collected via centrifugation at 3000 rpm in a Hettich Rotana 460R swinging bucket centrifuge. RNA was extracted from each of the pellets using the Qiashredder and RNeasy extraction kits (Qiagen, Hilden, Germany). Samples of at least 1000 ng of RNA were submitted to the Fred Hutch Genomics core for RNASeq library preparation and sequencing on an Illumina HiSeq. Principle component analysis was performed in R v4.0.2. using the top 500 differentially regulated genes. Volcano plots were constructed by taking the $\log_2$ of the fold change and the negative $\log_{10}$ of the unadjusted p-value for each gene.

4.5.9 *Statistical analysis*

Statistical analyses were performed according to test specified above and in each figure legend using Prism v7 software (GraphPad) or R v4.0.2. P-values less than or equal to

0.05 were considered statistically significant and are marked with asterisks (*, p<0.05, **, p<0.01; ***, p<0.001; ****, p<0.0001; n.s., not significant).

Chapter 5. *CONCLUSION AND FUTURE DIRECTIONS*

5.1 STATE OF THE FIELD

Before starting this thesis work, prior research had identified strain differences that determine disease risk. Genetic diversity of *H. pylori* isolates within a single host had been identified, but not extensively characterized. The work presented in this thesis began with the hypothesis that genetic alterations among *H. pylori* populations within host impact phenotypes related to pathogenesis and our belief that clinical outcomes from infection are due not only to factors of the host and the genetics of the infecting strain, but also to the trajectory of continued evolution within that single host.

In 2000, Israel et al. analyzed microarray data from 42 *H. pylori* isolates from an individual at two time points and detected significant genetic diversity in this patient [65]. In the 20 years since this original study was published, others had utilized next generation sequencing to study isolates collected from single hosts [61,70,94]. However, historically there have been several challenges to these studies. First, many isolates originate from individuals who have been treated with antibiotics, resulting in population bottlenecks. The study of *H. pylori* adaptation to antibiotics is an interesting and important question that can inform clinical treatment strategies, but confounds examination of selective pressures imparted by the host. Several other studies used paired isolates to calculate the rates of evolution, but do not tell us about population

structure at each time point [61,67]. *H. pylori* is also highly recombinogenic and there are challenges in identifying genetic material that has been horizontally acquired. Computational tools such as ClonalFrame have been developed to identify regions of the genome that exhibit high diversity due to introduction of divergent DNA from a recombination event [98], however recombination events occurring between highly related regions are difficult to identify and can confound evolutionary analysis [176].

The J99 collection presents a unique opportunity to study isolates from two different time points in the absence of antibiotic eradication. When I began this project in 2016, the lab had re-sequenced these isolates using new sequencing technologies (Illumina MiSeq, short read sequencing), which was able to detect additional diversity across the entire genome. I wanted capture and characterize a more complete picture of within host genetic diversity, but also pair this analysis with examination of phenotypes associated with pathogenesis. We did not know if genetic diversity would confer quantifiable phenotypic differences, however we were able to show here that these isolates differ in many phenotypic characteristics including cell morphology, motility, induction of inflammatory cytokines, and colonization in two different mouse models. Sequencing of multiple isolates from a single individual allows quantification of recombination and identification of population structure, which underlies adaptation to niches.

5.2 ISOLATES FORM DISTINCT SUB-POPULATIONS WITHIN HOST WITH DISTINCT GENETIC AND PHENOTYPIC CHARACTERISTICS

As outlined in Chapter 2, we found that isolates within a single individual formed sub-populations reflecting adaptation to distinct niches independent from gastric region of origin. Consistent with our findings, another study by Allioud *et al*. that sequenced *H. pylori* isolates collected from single individuals, also detected genetic variation, but did not find any genetic signature of stomach region specific adaptation [90]. While Allioud *et al*. propose frequent mixing and migration as a potential explanation for this, we find that isolates in our collection do form a genetically distinct sub-populations consistent with adaptation to niches present in multiple gastric regions. This prompts the question- what niches within the host are these populations adapting to? One possible hypothesis is that isolates are adapting to different compartments within the glands. As outlined in Figure 1.1 the glands are comprised of surface mucous cells, precursor, and stem cells which are located in the middle and base providing distinct microenvironments for bacterial growth. In support of this, Fung et al. has shown that *H. pylori* forms adherent microcolonies that can both occupy the surface epithelial cells, but also penetrate deep within the glands [156]. Another hypothesis is that these isolates adapt both mucus and epithelium associated populations, as *H. pylori* has been found in the mucus and also adhered to the epithelium. Mouse models have been used to determine total stomach colonization from homogenized sections, however these methods cannot distinguish colonization in the glands versus mucus. Immunofluorescence of stomach sections as performed and described in Figure 4.4 can detect gland associated populations, however sectioning methods do not preserve the mucus above the epithelium. Recent methods of culturing organoids showed that an air-liquid interface can promote mucus secretion, potentially providing the opportunity to study mucus and cell associated

populations in an *in vitro* model [177]. Utilization of these models together with animal infections, may help elucidate strain specific factors and phenotypes that determine niche tropism or adaptation.

Considering the relatively high number of mutations detected within the collection, we did not know if we would be able to link genetic determinants with the observed phenotypes. However, in the preceding chapters, I have shown that we can use datasets like these as a tool to identify phenotypes and link these phenotypes with genetic alterations that occurred over the six year period of adaptation. These findings can serve as the foundation for mechanistic studies of protein function. Others in the lab have created a mechanistic model of cell shape that includes cell shape proteins Csd4, Csd5, and Csd6 [20,136], which were found to have loss of function mutations in the J99 collection. However, we also found MreB 284V mutation which alters the parameters of helical cell shape. Very little is known about the role of MreB in generating shape for helical bacteria, however we recently showed this protein is essential in *H. pylori* and localizes to the minor helical axis on the bacterial cell body [123]. We also identified mutations in CagPAI protein CagY that impacts IL-8 release from AGS cells. There have been some preliminary studies suggesting this repeat region may be involved in binding to host integrins [83], however the mechanisms of this are still poorly understood.

While some studies have shown that isolates from the same individual can be genetically differentiated, this body of work argues that isolates can also be phenotypically differentiated and the phenotypic impacts of isolate genetic diversity within host should be characterized. From the *in vitro* phenotyping, we see that recent

isolates elicit greater IL-8 secretion than the ancestral isolates on average. Considering isolates with greater ability to initiate pro-inflammatory cytokine secretion from epithelial cells are associated with worse patient outcomes, it is an important finding that isolates can become more pro-inflammatory during the course of infection. This could have consequences for disease progression and outcome. Other phenotypes important for colonization--helical cell shape and motility, are lost or decreased in some of the recent isolates. This could suggest that these characteristics are dispensable or detrimental in specific host microenvironments as infection progresses. Representative isolates from certain sub-populations appear to be superior colonizers of mice. It is possible that this population has increased potential to colonize a new host and may be involved in transmission. However, it is also possible these differences are specific to the mouse model.

Considering our study is one of the few that has looked at large populations of isolates within a single treatment naïve host, we need more datasets that look at isolates from a single individual in order to identify population wide trends. It is very likely that within host bacterial adaptation varies between hosts.

5.3 CONSEQUENCES OF WITHIN HOST DIVERSITY FOR DISEASE PROGRESSION AND TREATMENT

Getting back to the larger picture, what impact does this trajectory of within host evolution have on disease progression or maintenance of homeostasis? There has been a large debate about the best treatment practices for individuals carrying *H. pylori*. Universal eradication of infection is imprudent given the consequences of antibiotic overuse and the fact that the majority of individuals never develop disease associated

with infection [76]. Some studies have shown *H. pylori* may provide a benefit to human hosts in that it can be protective against reflux, allergies, and esophageal cancer [178]. Therefore, one way to limit the burden of *H. pylori* associated disease is to develop better models to stratify disease risk.

Certain strain specific genetic markers, such as the *cag* pathogenicity island, that carry increased disease risk have been identified. However, the work presented here shows that within host, isolates can have diverse genetic and phenotypic characteristics and these can be modified over time. While the exact implications of this on disease outcome and risk remains largely unknown, recording changes in certain bacterial genotypes through infection may be useful in predicting disease risk and timing of intervention. Using sequencing to detect bacterial genotypes at various time points during infection has been proposed as a method to monitor factors like antibiotic resistance in order to determine best intervention strategies [66,77]. However, due to the invasive nature of methods used to extract bacteria and monitor resistance (endoscopy), these strategies have not been brought to fruition. Our lab has shown more sensitive detection methods like droplet digital PCR can be used to detect bacterial loads and certain genotypes from stool, making this a more feasible option [77]. This method can also be used to collect additional epidemiological information to monitor bacterial genotypes in single individual over time.

5.4 Future directions and concluding thoughts

While it is exciting to think about the clinical implications of the findings presented, this dissertation just represents a single case study. There is much left to explore to create a clear concept of the ways in which genetic changes confer altered

phenotype, and potentially shape disease outcome and progression. Even in this dataset alone, we still do not understand genetic determinants of AGS cell rounding or colonization in the WT and metaplastic mouse models. Undoubtedly, additional work on this dataset and others similar to it will create a better understanding of the different strategies *H. pylori* has adapted to modulate interaction with host tissues. Sequencing and bioinformatic tools continue to improve, providing increasingly detailed views of *H. pylori* diversity in single individuals. I expect that just as this dissertation improved upon tools in the original analysis by Israel *et. al*, new technology will elucidate difficult to map regions of the genome.

At the heart of this project, was a quest to understand the longstanding relationship between the gastric bacterium *H. pylori* and its human host. Specifically, I wanted to shed light on the trajectory of evolution that can tip the scale from symbiont to pathogen and from health to disease. Historically, bacteria have been viewed as an agent of disease, an adversary to human health. However, we are coming to understand the essential role bacteria play in our lives and the delicate balance that must be maintained for our health. *H. pylori* infection in the human stomach is an excellent model to study properties of both the host and bacterium that shape this relationship. In this study we have shown that *H. pylori* properties change throughout infection, potentially in ways that are pathoadaptive. But, this is just a single snapshot of a dynamic relationship in a single individual. By ignoring the dynamics of this relationship, we are limiting ourselves to a narrow scientific view of the biology of infection. I hope many others will share my curiosity of this intimate association so we

can continue to understand all aspects of the complex and ongoing relationship between *H. pylori* and its human host.

BIBLIOGRAPHY

1. Warren JR, Marshall B. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. Vol. 1, Lancet (London, England). England; 1983. p. 1273–5.

2. Hooi JKY, Lai WY, Ng WK, Suen MMY, Underwood FE, Tanyingoh D, et al. Global Prevalence of *Helicobacter pylori* Infection: Systematic Review and Meta-Analysis. Gastroenterology. 2017 Aug;153(2):420–9.

3. Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P, et al. An African origin for the intimate association between humans and *Helicobacter pylori*. Nature. 2007;445(7130):915–8.

4. Kusters JG, Van Vliet AHM, Kuipers EJ. Pathogenesis of *Helicobacter pylori* infection. Clin Microbiol Rev. 2006;19(3):449–90.

5. Blaser MJ, Atherton JC. *Helicobacter pylori* persistence : biology and disease. Sci Med. 2004;113(3):321–33.

6. Rawla P, Barsouk A. Epidemiology of gastric cancer: Global trends, risk factors and prevention. Prz Gastroenterol. 2019;14(1):26–38.

7. Wroblewski LE, Peek RM, Wilson KT. *Helicobacter pylori* and Gastric Cancer : Factors That Modulate Disease Risk. Clin Microbiol Rev. 2010;23(4):713–39.

8. Kodaman N, Pazos A, Schneider BG, Blanca Piazuelo M, Mera R, Sobota RS, et al. Human and *Helicobacter pylori* coevolution shapes the risk of gastric disease. Proc Natl Acad Sci U S A. 2014;111(4):1455–60.

9. Ghoshal UC, Chaturvedi R, Correa P. The enigma of *Helicobacter pylori* infection and gastric cancer. Indian J Gastroenterol Off J Indian Soc Gastroenterol. 2010 Jun;29(3):95–100.

10. Blaser MJ, Perez-perez GI, Kleanthous H, Cover TL, Peek RM, Chyou PH, et al. Infection with *Helicobacter pylori* Strains Possessing cagA Is Associated with an Increased Risk of Developing Adenocarcinoma of the Stomach. Cancer Res. 1995;55:2111–6.

11. Ap UAS, El-omar EM, Carrington M, Chow W, Lissowska J, Yuan C, et al. Interleukin-1 polymorphisms associated with increased risk of gastric cancer. Nature. 2000;404(March):398–403.

12. Amieva M, Alto P, Alto P. Pathobiology of *Helicobacter pylori*-induced Gastric Cancer Manuel. Gastroenterology. 2017;150(1):64–78.

13. Mégraud F. Transmission of *Helicobacter pylori*: faecal-oral versus oral-oral route. Aliment Pharmacol Ther. 1995;9 Suppl 2:85–91.

14. Krakowka S, Eaton KA, Rings DM, Morgan DR. Gastritis induced by *Helicobacter pylori* in gnotobiotic piglets. Rev Infect Dis. 1991;13(7):S681–5.

15. Celli JP, Turner BS, Afdhal NH, Ewoldt RH, McKinley GH, Bansil R, et al. Rheology of gastric mucin exhibits a pH-dependent sol-gel transition. Biomacromolecules. 2007 May;8(5):1580–6.

16. Keilberg D, Ottemann KM. How *Helicobacter pylori* senses, targets and interacts with the gastric epithelium. Environ Microbiol. 2016;18(3):791–806.

17. Gu H. Role of Flagella in the Pathogenesis of *Helicobacter pylori*. Curr Microbiol. 2017;74(7):863–9.

18. Ottemann KM, Lowenthal AC. *Helicobacter pylori* uses motility for initial colonization and to attain robust infection. Infect Immun. 2002;70(4):1984–90.

19. Eaton KA, Suerbaum S, Josenhans C, Krakowka S. Colonization of gnotobiotic piglets by *Helicobacter pylori* deficient in two flagellin genes. Infect Immun. 1996;64(7):2445–8.

20. Sycuro LK, Pincus Z, Gutierrez KD, Biboy J, Stern CA, Vollmer W, et al. Peptidoglycan crosslinking relaxation promotes *Helicobacter pylori*'s helical shape and stomach colonization. Cell. 2010;141(5):822–33.

21. Martínez LE, O'Brien VP, Leverich CK, Knoblaugh SE, Salama NR. Nonhelical *Helicobacter pylori* mutants show altered gland colonization and elicit less gastric pathology than helical bacteria during chronic infection. Infect Immun. 2019;87(7):1–15.

22. Saadat I, Higashi H, Obuse C, Umeda M, Murata-kamiya N, Saito Y, et al. *Helicobacter pylori* CagA targets PAR1 / MARK kinase to disrupt epithelial cell polarity. Nature. 2007;447(May):3–7.

23. Gall A. TIFA Signaling in Gastric Epithelial Cells Response to *Helicobacter pylori* Infection. 2017;1–16.

24. Viala J, Chaput C, Boneca IG, Cardona A, Girardin SE, Moran AP, et al. Nod1 responds to peptidoglycan delivered by the *Helicobacter pylori* cag pathogenicity island. Nat Immunol. 2004;5(11):1166–74.

25. Keates S, Hitti Y, Upton M, Kelly C. *Helicobacter pylori* infection activates NF-kappa B in gastric epithelial cells. Gastroenterology. 1997;113(4):1099–109.

26. Bagnoli F, Buti L, Tompkins L, Covacci A, Amieva MR. *Helicobacter pylori* CagA induces a transition from polarized to invasive phenotypes in MDCK cells. Proc Natl Acad Sci. 2005;102(45).

27. Palframan SL, Kwok T, Gabriel K. Vacuolating cytotoxin A (VacA), a key toxin for *Helicobacter pylori* pathogenesis. Front Cell Infect Microbiol. 2012;2(July):92.

28. Hoy B, Löwer M, Weydig C, Carra G, Tegtmeyer N, Geppert T, et al. *Helicobacter pylori* HtrA is a new secreted virulence factor that cleaves E-cadherin to disrupt intercellular adhesion. EMBO Rep. 2010;11(10):798–804.

29. Tegtmeyer N, Wessler S, Necchi V, Rohde M, Harrer A, Rau TT, et al. *Helicobacter pylori* Employs a Unique Basolateral Type IV Secretion Mechanism for CagA Delivery. Cell Host Microbe. 2017;22(4):552-560.e5.

30. Wüstner S, Anderl F, Wanisch A, Sachs C, Steiger K, Nerlich A, et al. *Helicobacter pylori* γ-glutamyl transferase contributes to colonization and differential recruitment of T cells during persistence. Sci Rep. 2017;7(1):1–12.

31. Hanada K, Uchida T, Tsukamoto Y, Watada M, Yamaguchi N, Yamamoto K, et al. *Helicobacter pylori* Infection Introduces DNA Double-Strand Breaks in Host Cells. 2014;82(10):4182–9.

32. Sipponen P, Maaroos HI. Chronic gastritis. Scand J Gastroenterol. 2015;50(6):657–67.

33. McColl KEL, El-Omar E, Gillen D. Interactions between *H. pylori* infection, gastric acid secretion and anti-secretory therapy. Br Med Bull. 1998;54(1):121–38.

34. Blanca Piazuelo PC. The Gastric Precancerous Cascade. J Clin Exp Pathol. 2013;03(03):2–9.

35. Baltrus DA, Blaser MJ, Guillemin K. *Helicobacter pylori* Genome Plasticity. Genome Dyn. 2009;6:75–90.

36. Go MF, Kapur V, Graham DY, Musser JM. Population genetic analysis of *Helicobacter pylori* by multilocus enzyme electrophoresis: Extensive allelic diversity and recombinational population structure. J Bacteriol. 1996;178(13):3934–8.

37. Alm RA, Trust TJ. Analysis of the genetic diversity of *Helicobacter pylori:* The tale of two genomes. J Mol Med. 1999;77(12):834–46.

38. Oleastro M, Ménard A. The Role of *Helicobacter pylori* Outer Membrane Proteins in Adherence and Pathogenesis. Biology (Basel). 2013;2(3):1110–34.

39. Kersulyte D, Lee WK, Subramaniam D, Anant S, Herrera P, Cabrera L, et al. *Helicobacter pylori*'s plasticity zones are novel transposable elements. PLoS One. 2009;4(9).

40. Montano V, Didelot X, Foll M, Linz B, Reinhardt R, Suerbaum S, et al. Worldwide population structure, long-term demography, and local adaptation of *Helicobacter pylori*. Genetics. 2015;200(3):947–63.

41. Jones KR, Whitmire JM, Merrell DS. A tale of two toxins: *Helicobacter pylori* CagA and VacA modulate host pathways that impact disease. Front Microbiol. 2010;1(NOV):1–17.

42. Dossumbekova A, Prinz C, Mages J, Lang R, Kusters JG, Van Vliet AHM, et al. *Helicobacter pylori* HopH (OipA) and Bacterial Pathogenicity: Genetic and Functional Genomic Analysis of hopH Gene Polymorphisms. J Infect Dis. 2006;194(10):1346–55.

43. Bridge DR, Merrell DS. Polymorphism in the *Helicobacter pylori* CagA and VacA toxins and disease © 2013 Landes Bioscience.2013;(April):101–17.

44. Ferreira RM, Machado JC, Figueiredo C. Clinical relevance of *Helicobacter pylori* vacA and cagA genotypes in gastric carcinoma. Best Pract Res Clin Gastroenterol. 2014;28(6):1003–15.

45. Hatakeyama M. Oncogenic mechanisms of the *Helicobacter pylori* CagA protein. Nat Rev Cancer. 2004;4(9):688–94.

46. Azuma T, Yamakawa A, Yamazaki S, Ohtani M, Ito Y, Muramatsu A, et al. Distinct diversity of the cag pathogenicity island among *Helicobacter pylori* strains in Japan. J Clin Microbiol. 2004;42(6):2508–17.

47. Peek RM, Crabtree JE. *Helicobacter* infection and gastric neoplasia. J Pathol. 2006;208(2):233–48.

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

49. Wang G, Ge Z, Rasko DA, Taylor DE. Lewis antigens in *Helicobacter pylori:* Biosynthesis and phase variation. Mol Microbiol. 2000;36(6):1187–96.

50. Bergman M, Prete G Del, Kooyk Y Van, Appelmelk B. *Helicobacter pylori* phase variation , immune modulation and gastric autoimmunity. Nat Rev Microbiol. 2006;4(February):1194–7.

51. Björkholm B, Sjölund M, Falk PG, Berg OG, Engstrand L, Andersson DI. Mutation

frequency and biological cost of antibiotic resistance in *Helicobacter pylor*i. Proc Natl Acad Sci U S A. 2001;98(25):14607–12.

52. Wang GE, Wilson TJM, Jiang QIN, Taylor DE, Al WET. Spontaneous Mutations That Confer Antibiotic Resistance in *Helicobacter pylori*. Antimicrob Agents Chemother. 2001;45(3):727–33.

53. Lee H, Popodi E, Tang H, Foster PL. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. Proc Natl Acad Sci U S A. 2012;109(41).

54. García-Ortíz MV, Marsin S, Arana ME, Gasparutto D, Guérois R, Kunkel TA, et al. Unexpected role for *Helicobacter pylori* DNA Polymerase I as a source of Genetic Cariability. PLoS Genetics. 2011;7(6):1–8.

55. Eutsey R, Wang G, Maier RJ. Role of a MutY DNA glycosylase in combating oxidative DNA damage in *Helicobacter pylori*. DNA Repair (Amst). 2007;6(1):19–26.

56. Dorer MS, Cohen IE, Sessler TH, Fero J, Salama NR. Natural competence promotes H*elicobacter pylori* chronic infection. Infect Immun. 2013;81(1):209–15.

57. Baltrus DA, Guillemin K, Phillips PC. Natural transformation increases the rate of adaptation in the human pathogen *Helicobacter pylori*. Evolution (N Y). 2008;62(1):39–49.

58. Dunne C, Dolan B, Clyne M. Factors that mediate colonization of the human stomach by *Helicobacter pylori*. World J Gastroenterol. 2014;20(19):5610–24.

59. Howitt MR, Lee JY, Lertsethtakarn P, Vogelmann R, Joubert L-M, Ottemann KM, et al. ChePep Controls *Helicobacter pylori* Infection of the Gastric Glands. MBio. 2011;2(4):1–10.

60. Linz B, Windsor HM, McGraw JJ, Hansen LM, Gajewski JP, Tomsho LP, et al. A mutation burst during the acute phase of *Helicobacter pylori* infection in humans and rhesus macaques. Nat Commun. 2014;5(May):1–8.

61. Kennemann L, Didelot X, Aebischer T, Kuhn S, Drescher B, Droege M, et al. *Helicobacter pylori* genome evolution during human infection. Proc Natl Acad Sci. 2011;108(12):5033–8.

62. Israel DA, Salama N, Krishna U, Rieger UM, Atherton JC, Falkow S, et al. *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. Proc Natl Acad Sci U S A. 2001;98(25):14625–30.

63. Krishna U, Romero-Gallo J, Suarez G, Azah A, Krezel AM, Varga MG, et al.

Genetic evolution of a *Helicobacter pylori* acid-sensing histidine kinase and gastric disease. J Infect Dis. 2016;214(4):644–8.

64. Alm RA, Ling LL, Moir DT, King BL, Brown ED, Doig PC, et al. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen. Nature. 1999;397(February):176–80.

65. Israel DA, Salama N, Krishna U, Rieger UM, Atherton JC, Falkow S, et al. *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. Proc Natl Acad Sci U S A. 2001;98(25):14625–30.

66. Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. Within-host evolution of bacterial pathogens. Nat Rev Microbiol. 2016;14(3):150–62.

67. Prouzet-Mauléon V, Hussain MA, Lamouliatte H, Kauser F, Mégraud F, Ahmed N. Pathogen evolution in vivo: Genome dynamics of two isolates obtained 9 years apart from a duodenal ulcer patient infected with a single *Helicobacter pylori* strain. J Clin Microbiol. 2005;43(8):4237–41.

68. Linz B, Windsor HM, Gajewski JP, Hake CM, Drautz DI, Schuster SC, et al. *Helicobacter pylori* genomic microevolution during naturally occurring transmission between adults. PLoS One. 2013;8(12):1–9.

69. Furuta Y, Konno M, Osaki T, Yonezawa H, Ishige T, Imai M, et al. Microevolution of virulence-related genes in *Helicobacter pylori* familial infection. PLoS One. 2015;10(5):1–17.

70. Falush D, Kraft C, Taylor NS, Correa P, Fox JG, Achtman M, et al. Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: Estimates of clock rates, recombination size, and minimal age. Proc Natl Acad Sci. 2001;98(26):15056–61.

71. Parsonnet J, Friedman G, Vandersteen D. *Helicobacter pylori* infection and the risk of gastric carcinoma. N Engl J Med. 1991;325(16):1127–31.

72. Ohata H, Kitauchi S, Yoshimura N, Mugitani K. Progression of chronic atrophic gastritis associated with *Helicobacter pylori* infection increases risk of gastric cancer. Int J Cancer. 2004;109:138–43.

73. Blaser MJ, Berg DE. *Helicobacter pylori* genetic diversity and risk of human disease. J Clin Invest. 2001;107(7):767–73.

74. Salama NR, Gonzalez-Valencia G, Deatherage B, Aviles-Jimenez F, Atherton JC, Graham DY, et al. Genetic Analysis of *Helicobacter pylori* Strain Populations Colonizing the Stomach at Different Times Postinfection †. J Bacteriol. 2007;189(10):3834–45.

75. Suerbaum S, Josenhans C. *Helicobacter pylori* evolution and phenotypic diversification in a changing host. Nat Rev Microbiol. 2007;5(6):441–52.

76. Dang BN, Graham DY. *Helicobacter pylori* infection and antibiotic resistance: a WHO high priority? Nat Publ Gr. 2017;14(7):383–4.

77. Sun L, Talarico S, Yao L, He L, Self S, You Y, et al. Droplet digital PCR-based detection of clarithromycin resistance in *Helicobacter pylori* isolates reveals frequent heteroresistance. J Clin Microbiol. 2018;56(9):1–11.

78. Cover TL. *Helicobacter pylori* diversity and gastric cancer risk. MBio. 2016;7(1):1–9.

79. Olbermann P, Josenhans C, Moodley Y, Uhr M, Stamer C, Vauterin M, et al. A global overview of the genetic and functional diversity in the *Helicobacter pylori* cag pathogenicity island. PLoS Genet. 2010;6(8).

80. Backert S, Brandt S, Kwok T, Hartig R, Ko W. NF- kB activation and potentiation of proinflammatory responses by the *Helicobacter pylori* CagA protein. Proc Natl Acad Sci. 2005;102(26):9300–5.

81. Barrozo RM, Cooke CL, Hansen LM, Lam AM, Gaddy JA, Johnson EM, et al. Functional Plasticity in the Type IV Secretion System of *Helicobacter pylori.* PLoS Pathog. 2013;9(2).

82. Barrozo RM, Hansen LM, Lam AM, Skoog EC, Martin ME, Cai LP, et al. CagY is an Immune-Sensitive Regulator of the *Helicobacter pylori* Type IV Secretion System. Gastroenterology. 2016;(October):1–12.

83. Skoog EC, Morikis VA, Martin ME, Foster GA, Cai LP, Hansen LM, et al. CagY-dependent regulation of type IV secretion in *Helicobacter pylori* is associated with alterations in integrin binding. MBio. 2018;9(3):1–16.

84. Brandt S, Kwok T, Hartig R, König W, Backert S. NF-kappaB activation and potentiation of proinflammatory responses by the *Helicobacter pylori* CagA protein. Proc Natl Acad Sci U S A. 2005;102(26):9300–5.

85. Lamb A, Yang XD, Tsang YHN, Li JD, Higashi H, Hatakeyama M, et al. *Helicobacter pylori* CagA activates NF-κB by targeting TAK1 for TRAF6-mediated Lys 63 ubiquitination. EMBO Rep. 2009;10(11):1242–9.

86. Salama NR, Hartung ML, Müller A. Life in the human stomach: Persistence strategies of the bacterial pathogen *Helicobacter pylori*. Nat Rev Microbiol. 2013;11(6):385–99.

87. Javaheri A, Kruse T, Moonens K, Mejías-luque R, Debraekeleer A, Asche CI, et al. *Helicobacter pylori* adhesin HopQ engages in a virulence-enhancing interaction with human CEACAMs. Nat Microbiol. 2016;2(1):1–12.

88. Solnick J V., Hansen LM, Salama NR, Boonjakuakul JK, Syvanen M. Modification of *Helicobacter pylori* outer membrane protein expression during experimental infection of rhesus macaques. Proc Natl Acad Sci U S A. 2004;101(7):2106–11.

89. Morelli G, Didelot X, Kusecek B, Schwarz S, Bahlawane C, Falush D, et al. Microevolution of *Helicobacter pylori* during prolonged infection of single hosts and within families. PLoS Genet. 2010;6(7):1–12.

90. Ailloud F, Didelot X, Woltemate S, Pfaffinger G, Overmann J, Bader RC, et al. Within-host evolution of *Helicobacter pylori* shaped by niche-specific adaptation, intragastric migrations and selective sweeps. Nat Commun. 2019;10(1).

91. Amieva M, Jr RMP. Pathobiology of *Helicobacter pylori* – Induced Gastric Cancer. Gastroenterology. 2016;150(1):64–78.

92. Nilsson C, Bjo B, Skoglund A, Ba HK, Engstrand L. A Changing Gastric Environment Leads to Adaptation of Lipopolysaccharide Variants in *Helicobacter pylori* Populations during Colonization. PLoS One. 2009;4(6):1–9.

93. Cao Q, Didelot X, Wu Z, Li Z, He L, Li Y, et al. Progressive genomic convergence of two *Helicobacter pylori* strains during mixed infection of a patient with chronic gastritis. Gut. 2015;64(4):554–61.

94. Didelot X, Nell S, Yang I, Woltemate S, Van Der Merwe S, Suerbaum S. Genomic evolution and transmission of *Helicobacter pylori* in two South African families. Proc Natl Acad Sci U S A. 2013;110(34):13880–5.

95. Uchiyama I, Mihara M, Nishide H, Chiba H, Kato M. MBGD update 2018: Microbial genome database based on hierarchical orthology relations covering closely related and distantly related comparisons. Nucleic Acids Res. 2019;47(D1):D382–9.

96. Solnick J V. Dynamic Expression of the BabA Adhesin and its BabB paralog during *Helicobacter pylori* infection in Rhesus Macaques. Infect Immun. 2017;(April).

97. Åberg A, Gideonsson P, Vallström A, Olofsson A, Öhman C, Rakhimova L, et al. A Repetitive DNA Element Regulates Expression of the *Helicobacter pylori* Sialic Acid Binding Adhesin by a Rheostat-like Mechanism. PLoS Pathog. 2014;10(7).

98. Didelot X, Wilson DJ. ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. PLoS Comput Biol. 2015;11(2):1–18.

99. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics. 2018;2–6.

100. Alves JM, Prieto T, Posada D. Multiregional Tumor Trees Are Not Phylogenies. Trends in Cancer. 2017;3(8):546–50.

101. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. Virus Evol. 2018;4(1):1–9.

102. Kaplan-türköz B, Jiménez-soto LF, Dian C, Ertl C, Remaut H, Louche A. Structural insights into *Helicobacter pylori* oncoprotein CagA interaction with β 1 integrin. 2012;

103. Hatakeyama M. Structure and function of *Helicobacter pylori* CagA, the first-identified bacterial protein involved in human cancer. Proc Japan Acad Ser B. 2017;93(4):196–219.

104. Dorer MS, Sessler TH, Salama NR. Recombination and DNA Repair in *Helicobacter pylori*. Annu Rev Microbiol. 2008;23(1):1–7.

105. Sakoparnig T, Field C, van Nimwegen E. Whole genome phylogenies reflect long-tailed distributions of recombination rates in many bacterial species. bioRxiv. 2019;601914.

106. Jorth P, Staudinger BJ, Wu X, Bruce JE, Timothy L, Singh PK, et al. within Cystic Fibrosis Lungs Regional Isolation Drives Bacterial Diversification within Cystic Fibrosis Lungs. Cell Host Microbe. 2015;18(3):307–19.

107. Moyat M, Velin D. Immune responses to *Helicobacter pylori* infection. World J Gastroenterol. 2014;20(19):5583–93.

108. Martinez LE, Hardcastle JM, Wang J, Pincus Z, Hoover TR, Bansil R, et al. To Maintain Robust Motility in Viscous Environments. Mol Microbiol. 2017;99(1):88–110.

109. Lee A, O'Rourke J, De Ungria MC, Robertson B, Daskalopoulos G, Dixon MF. A standardized mouse model of *Helicobacter pylori* infection: Introducing the Sydney strain. Gastroenterology. 1997;112(4):1386–97.

110. Baldwin DN, Shepherd B, Kraemer P, Hall MK, Sycuro LK, Pinto-Santini DM, et al. Identification of *Helicobacter pylori* genes that contribute to stomach colonization. Infect Immun. 2007;75(2):1005–16.

111. Humbert O, Salama NR. The *Helicobacter pylori* HpyAXII restriction-modification system limits exogenous DNA uptake by targeting GTAC sites but shows

asymmetric conservation of the DNA methyltransferase and restriction endonuclease components. Nucleic Acids Res. 2008;36(21):6893–906.

112. Barrick JE, Colburn G, Deatherage DE, Traverse CC, Strand MD, Borges JJ, et al. Identifying structural variation in haploid microbial genomes from short-read resequencing data using breseq. 2014;1–17.

113. Hochberg Y. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing Author (s): Yoav Benjamini and Yosef Hochberg Source : Journal of the Royal Statistical Society . Series B (Methodological), Vol . 57 , No . 1 (1995), Publi. 1995;57(1):289–300.

114. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A. 1979;76(10):5269–73.

115. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. PopGenome: An efficient swiss army knife for population genomic analyses in R. Mol Biol Evol. 2014;31(7):1929–36.

116. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2.

117. Dixit PD, Pang TY, Studier FW, Maslov S. Recombinant transfer in the basic genome of *Escherichia coli*. Proc Natl Acad Sci U S A. 2015;112(29):9070–5.

118. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: Multiple Alignment of Conserved Genomic Sequence with Rearrangements. Genome Res. 2004;14:1394–403.

119. Copass M, Grandi G, Rappuoli R. Introduction of unmarked mutations in the *Helicobacter pylori* vacA gene with a sucrose sensitivity marker. Infect Immun. 1997;65(5):1949–52.

120. Wang Y, Taylor DE. Chloramphenicol resistance in *Campylobacter coli*: nucleotide sequence, expression, and cloning vector construction. Gene. 1990;94(1):23–8.

121. Höltje J-V. Growth of the Stress-Bearing and Shape-Maintaining Murein Sacculus of *Escherichia coli.* Microbiol Mol Biol Rev. 1998;62(1):181–203.

122. Yang DC, Blair KM, Taylor JA, Petersen TW, Sessler T, Tull CM, et al. A Genome-Wide *Helicobacter pylori* Morphology Screen Uncovers a Membrane-Spanning Helical Cell Shape Complex. J Bacteriol. 2019;201(14):1–16.

123. Taylor JA, Bratton BP, Sichel SR, Blair KM, Jacobs HM, Demeester KE, et al. Distinct cytoskeletal proteins define zones of enhanced cell wall synthesis in

Helicobacter pylori. Elife. 2020;9:1–38.

124. Sycuro LK, Wyckoff TJ, Biboy J, Born P, Pincus Z, Vollmer W, et al. Multiple peptidoglycan modification networks modulate *Helicobacter pylori*'s cell shape, motility, and colonization potential. PLoS Pathog. 2012;8(3).

125. Berg HC, Turner L. Movement of microorganisms in viscous environments. 1979;278(March):349–51.

126. Magariyama Y, Kudo S. A mathematical explanation of an increase in bacterial swimming speed with viscosity in linear-polymer solutions. Biophys J. 2002;83(2):733–9.

127. Constantino MA, Jabbarzadeh M, Fu HC, Bansil R. Helical and rod-shaped bacteria swim in helical trajectories with little additional propulsion from helical shape. Sci Adv. 2016 Nov;2(11):e1601661.

128. Terry K, Williams SM, Connolly L, Ottemann KM. Chemotaxis plays multiple roles during *Helicobacter pylori* animal infection. Infect Immun. 2005;73(2):803–11.

129. Cerda O, Rivas A, Toledo H. *Helicobacter pylori* strain ATCC700392 encodes a methyl-accepting chemotaxis receptor protein (MCP) for arginine and sodium bicarbonate. FEMS Microbiol Lett. 2003;224(2):175–81.

130. Machuca MA, Johnson KS, Liu YC, Steer DL, Ottemann KM, Roujeinikova A. *Helicobacter pylori* chemoreceptor TlpC mediates chemotaxis to lactate. Sci Rep. 2017;7(1):1–15.

131. Croxen MA, Sisson G, Melano R, Hoffman PS. The *Helicobacter pylori* chemotaxis receptor tlpB (HP0103) is required for pH taxis and for colonization of the gastric mucosa. J Bacteriol. 2006;188(7):2656–65.

132. Foynes S, Dorrell N, Ward SJ, Stabler RA, McColm AA, Rycroft AN, et al. *Helicobacter pylori* possesses two CheY response regulators and a histidine kinase sensor, CheA, which are essential for chemotaxis and colonization of the gastric mucosa. Infect Immun. 2000;68(4):2016–23.

133. Wadhams GH, Armitage JP. Making sense of it all: Bacterial chemotaxis. Nat Rev Mol Cell Biol. 2004;5(12):1024–37.

134. Huang JY, Goers Sweeney E, Guillemin K, Amieva MR. Multiple Acid Sensors Control *Helicobacter pylori* Colonization of the Stomach. PLoS Pathog. 2017;13(1):1–28.

135. Dayanand K. Colonization, localization, and inflammation: The roles of *H.pylori* chemotaxis in vivo. Curr Opin Microbiol. 2018;1741(5):139-1451–72.

136. Yang DC, Blair KM, Salama NR. Staying in Shape: the Impact of Cell Shape on Bacterial Survival in Diverse Environments. Microbiol Mol Biol Rev. 2016;80(1):187–203.

137. Girardin SE, Werts C, Mengin-lecreulx D. Role of AmiA in the Morphological Transition of *Helicobacter pylori* and in Immune Escape. PLoS Pathog. 2006;2(9):844–52.

138. Krzyzek P, Gościniak G. Morphology of *Helicobacter pylori* as a result of peptidoglycan and cytoskeleton rearrangements. Prz Gastroenterol. 2018;13(3):182–95.

139. Lowenthal AC, Hill M, Sycuro LK, Mehmood K, Salama NR, Ottemann KM. Functional analysis of the *Helicobacter pylori* flagellar switch proteins. J Bacteriol. 2009;191(23):7147–56.

140. Josenhans C, Eaton KA, Thevenot T, Suerbaum S. Switching of Flagellar Motility in *Helicobater pylori* by Reversible Length Variation of a Short Homopolymeric Sequence Repeat in fliP, a Gene Encoding a Basal Body protein. 2000;68(8):4598–603.

141. Shaevitz JW, Gitai Z. The structure and function of bacterial actin homologs. Cold Spring Harb Perspect Biol. 2010;2(9):1–19.

142. Van den Ent F, Amos LA, Löwe J. Prokaryotic origin of the actin cytoskeleton. Nature. 2001;413(6851):39–44.

143. Hallinger DR, Romero-Gallo J, Peek RM, Forsyth MH. Polymorphisms of the acid sensing histidine kinase gene arsS in *Helicobacter pylori* populations from anatomically distinct gastric sites. Microb Pathog. 2012;53(5–6):227–33.

144. Rolig AS, Shanks J, Carter JE, Ottemann KM. *Helicobacter pylori* Requires TlpD-Driven Chemotaxis To Proliferate in the Antrum. 2012;80(10):3713–20.

145. Jones LJF, Carballido-López R, Errington J. Control of cell shape in bacteria: Helical, actin-like filaments in Bacillus subtilis. Cell. 2001;104(6):913–22.

146. van den Ent F, Izoré T, Bharat TAM, Johnson CM, Löwe J. Bacterial actin MreB forms antiparallel double filaments. Elife. 2014;2014(3):1–22.

147. Justice SS, Hunstad DA, Cegelski L, Hultgren SJ. Morphological plasticity as a bacterial survival strategy. Nat Rev Microbiol. 2008;6(2):162–8.

148. Baltrus DA, Amieva MR, Covacci A, Lowe TM, Merrell DS, Ottemann KM, et al. The complete genome sequence of *Helicobacter pylori* strain G27. J Bacteriol.

2009;91(1):447–8.

149. Pincus Z, Theriot JA. Comparison of quantitative methods for cell-shape analysis. J Microsc. 2007 Aug;227(Pt 2):140–56.

150. Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE. Tools for integrated sequence-structure analysis with UCSF Chimera. BMC Bioinformatics. 2006;7:1–10.

151. Williams SM, Chen Y-T, Andermann TM, Carter JE, McGee DJ, Ottemann KM. *Helicobacter pylori* chemotaxis modulates inflammation and bacterium-gastric epithelium interactions in infected mice. Infect Immun. 2007 Aug;75(8):3747–57.

152. Polk DB, Peek RM. *Helicobacter pylori* : gastric cancer and beyond. 2010;10(jUNE). Available from: http://dx.doi.org/10.1038/nrc2857

153. Correa, Pelayo, Piazuelo, M Blanca, Wilson K. Pathology of Gastric Intestinal Metaplasia Clinical Implications. Am J Gastroenterol. 2010;2009(3):493–8.

154. Song H, Ekheden IG, Zheng Z, Ericsson J, Nyrén O, Ye W. Incidence of gastric cancer among patients with gastric precancerous lesions: Observational cohort study in a low risk Western population. BMJ. 2015;351.

155. Wong BCY, Lam SK, Wong WM, Chen JS, Zheng TT, Feng RE, et al. *Helicobacter pylori* Eradication to Prevent Gastric Cancer in a High-Risk Region of China: A Randomized Controlled Trial. J Am Med Assoc. 2004;291(2):187–94.

156. Fung C, Tan S, Nakajima M, Skoog EC, Camarillo-Guerrero LF, Klein JA, et al. High-resolution mapping reveals that microniches in the gastric glands control *Helicobacter pylori* colonization of the stomach. PLoS Biol. 2019;17(5):1–28.

157. Keilberg D, Zavros Y, Shepherd B, Salama NR, Ottemann KM. Spatial and temporal shifts in bacterial biogeography and gland occupation during the development of a chronic infection. MBio. 2016;7(5):1–13.

158. Schreiber S, Konradt M, Groll C, Scheid P, Hanauer G, Werling HO, et al. The spatial orientation of *Helicobacter pylori* in the gastric mucus. Proc Natl Acad Sci U S A. 2004;101(14):5024–9.

159. Goldenring JR. Expression of Activated Ras in Gastric Chief Cells of Mice Leads tothe Full Spectrum of etaplastic Lineage Transitions. 2017;150(4):918–30.

160. Brien VPO, Koehne AL, Dubrulle J, Rodriguez AE, Leverich CK, Kong VP, et al. Sustained *Helicobacter pylori i*nfection accelerates gastric dysplasia in a mouse model. 2020;4(2):1–19.

161. Lawson DA, Bhakta NR, Kessenbrock K, Prummel KD, Yu Y, Takai K, et al. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. Nature [Internet]. 2015;526(7571):131–5.

162. Sparwasser T, Müller A. Tolerance rather than immunity protects from *Helicobacter*. 2012;140(1):199–209.

163. Asim M, Chikara SK, Ghosh A, Vudathala S, Romero-Gallo J, Krishna US, et al. Draft genome sequence of gerbil-adapted carcinogenic *Helicobacter pylori* strain 7.13. Genome Announc. 2015;3(3):2164.

164. Petersen CP, Mills JC, Goldenring JR. Murine Models of Gastric Corpus Preneoplasia. Cmgh. 2017;3(1):11–26.

165. Jackson L, Potter B, Schneider S, Fitzgibbon M, Blair K, Farah H, et al. *Helicobacter pylori* diversification during chronic infection within a single host generates sub-populations with distinct phenotypes. PLoS Pathog. 2020;1–28.

166. Sun H, Lin S, Ko T, Pan J, Liu C, Lin C, et al. Structure and Mechanism of *Helicobacter pylori* Fucosyltransferase. J Biol Chem. 2007;282(13):9973–82.

167. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. Science (80-). 2009;323(5910):133–8.

168. Mahdavi J, Sondén B, Hurtig M, Olfat FO, Forsberg L, Roche N, et al. *Helicobacter pylori* sabA adhesin in persistent infection and chronic inflammation. Science (80-). 2002;297(5581):573–8.

169. Yamaoka Y. Increasing evidence of the role of *Helicobacter pylori* SabA in the pathogenesis of gastroduodenal disease. J Infect Dev Ctries. 2008;2(3):174–81.

170. Talarico S. Regulation of *Helicobacter pylori* adherence by gene conversion. Mol Microbiol. 2012;84(6):1050–61.

171. Bourzac KM, Botham CM, Guillemin K. *Helicobacter pylori* CagA induces AGS cell elongation through a cell retraction defect that is independent of Cdc42, Rac1, and Arp2/3. Infect Immun. 2007;75(3):1203–13.

172. Jiménez AP, Traum A, Boettger T, Hackstein H, Richter M, Dammann RH. The tumor suppressor RASSF1A induces the YAP1 target gene ANKRD1 that is epigenetically inactivated in human cancers and inhibits tumor growth. Oncotarget. 2017;8(51):1–16.

173. Liu KSH, Wong IOL, Leung WK. *Helicobacter pylori* associated gastric intestinal metaplasia: Treatment and surveillance. World J Gastroenterol. 2016;22(3):1311–20.

174. Xiaodan Mai, PhD, MBBSa, Michael J. LaMonte, PhD, MPHa, Kathleen M. Hovey, MSa, Jo L. Freudenheim, PhDa, Christopher A. Andrews, PhDb, Robert J. Genco, DDS, PhDc, and Jean Wactawski-Wende P. Dynamics of *Helicobacter pylori* infection as a determinant of progression of gastric precancerous lesions: 16-year follow-up of an eradication trial. Physiol Behav. 2017;176(12):139–48.

175. Sáenz JB, Vargas N, Mills JC. Tropism for Spasmolytic Polypeptide-Expressing Metaplasia Allows *Helicobacter pylori* to Expand its Intra-gastric Niche. Gastroenterology. The American Gastroenterological Association; 2018. Available

176. Sakoparnig T, Field C, van Nimwegen E. Whole genome phylogenies reflect the distributions of recombination rates for many bacterial species. Elife. 2021;10:1–61.

177. Schlaermann P, Toelle B, Berger H, Schmidt SC, Glanemann M, Ordemann J, et al. A novel human gastric primary cell culture system for modelling *Helicobacter pylori* infection in vitro. Gut. 2016;65(2):202–13.

178. Wang Z, Shaheen NJ, Whiteman DC, Anderson LA, Vaughan TL, Corley DA, et al. *Helicobacter pylori* Infection Is Associated With Reduced Risk of Barrett's Esophagus: An Analysis of the Barrett's and Esophageal Adenocarcinoma Consortium. Am J Gastroenterol]. 2018;113(8):1148–55.

VITA

Laura Jackson was born and raised in Madison, Connecticut. She graduated with honors from the University of Connecticut with a Bachelor's of Science in Biology with a concentration in Molecular Biology in 2014. She studied and completed her undergraduate thesis on restriction modification systems in the halophilic archeon *Haloferax volcanii* under the direction of Dr. Thane Papke. After graduation, she worked as a research associate for Dr. Bonnie Gould Rothberg at the Yale Cancer Center identifying molecular and clinicopathologic risk factors for cancer progression in early stage melanoma and lung adenocarcinoma. In 2016, she moved to Seattle to join the Molecular and Cellular Biology program at the University of Washington. Laura's research and dissertation was completed in Dr. Nina Salama's lab at the Fred Hutchinson Cancer Research Center where she examined the phenotypic impacts of within host evolution of the bacterial pathogen *Helicobacter pylori*. She received her doctoral degree in 2021.