Genetic basis of host adaptation by the human stomach pathogen *Helicobacter pylori*

Ilana E. Cohen

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Nina Salama, Chair

Brad T Cookson

Jessica A. Hamerman

Program Authorized to Offer Degree:

Molecular and Cellular Biology

Portions of this dissertation were adapted from the following publication:

Infection and Immunity, Vol. 81, No. 1, January 2013, p. 209-215

doi:10.1128/IAI.01042-12

University of Washington

Abstract

Genetic basis of host adaptation by the human stomach pathogen *Helicobacter pylori*

Ilana E. Cohen

Chair of the Supervisory Committee:

Professor Nina Salama

Department of Microbiology, University of Washington

Division of Human Biology, Fred Hutchinson Cancer Research Center

Helicobacter pylori chronically colonizes the stomachs of approximately 50% of the world's population, with outcomes ranging from gastritis to ulcers and gastric cancer. *H. pylori* is genetically diverse both across and within infected individuals. This diversity can be generated by a variety of mechanisms, including mutation and recombination, and strains can exchange DNA through natural competence. The bacterial genetic changes that occur during the course of chronic infection are thought to adapt the organism to its niche at the gastric epithelial surface. Transmission from person to person is primarily among family members or close household contacts. Transmitted strains retain close genetic relatedness, but transmission may also be accompanied by bacterial genetic changes that have not been extensively studied at the whole genome level. Here, *H. pylori* transition to a new host or environment has been modeled by

adaptation of the bacteria to infect mice, as well as to grow under laboratory culture conditions. Whole genome sequencing of *H. pylori* isolates adapted to novel host conditions revealed an overrepresentation of mutations in genes related to the cell envelope, including outer membrane proteins. Isogenic mutants in *H. pylori* genes altered during murine adaptation revealed the dominant role of a loss of function mutation in *imc1*, encoding a previously-uncharacterized protein here denoted Inhibitor of mouse colonization, on murine colonization potential. Imc1 was found to be associated with the *H. pylori* cell envelope and may influence adherence to murine gastric epithelial cells. A loss of function mutation in *cagY*, encoding a component of the *cag* type IV secretion system, abrogated delivery of bacterial factors into host cells by *H. pylori* adapted for persistent murine infection. *H. pylori* selected for interaction with host cells in culture were found to have undergone recombination with a distinct wild type strain, among other mutations. Repeated passaging in liquid culture resulted in changes in genes involved in nutrient uptake and growth. Although many of the identified mutations remain to be characterized, this dissertation lays the groundwork for a more complete understanding of the changes required for *H. pylori* adaptation to new conditions.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Table **Page**

ACKNOWLEDGEMENTS

My deepest gratitude goes to my advisor, Nina Salama, for mentoring and guiding me through this dissertation work, and for the opportunity to pursue experimental strategies that were new to the lab. All of the members of the Salama lab, past and present, have provided an exceptionally collaborative, supportive, and fun working environment, and will be deeply missed. In particular, Jutta Fero, Christina Tull, and Tate H. Sessler provided essential experimental support and assistance with mouse colonization, host cell coculture, and protein experiments. Abigail Mazon assisted in strain construction and characterization of the *cagY* mutation.

Marion S. Dorer, Karen M. Ottemann, Manuel R. Amieva, Karen Guillemin, and David A. Baltrus kindly provided *H. pylori* strains for genome sequencing. I thank Jay Shendure and his laboratory, particularly Bethany Stackhouse, for assistance with sequencing library preparation and the FHCRC Genomics and Computational Biology Shared Resources, particularly Matthew Fitzgibbon, for experimental support and assistance with data analysis. Evgeni Sokurenko and his laboratory, particularly Elena Linardopoulou, provided computational tools and support for processing and assembly of sequence reads. Shumin Tan in the laboratory of Manuel Amieva tested the effects of mutations on bacterial adherence to MDCK cells.

The members of my thesis committee, Brad Cookson, Jessica Hamerman, Evgeni Sokurenko, and Kevin Urdahl, have provided useful experimental suggestions and support during the course of my dissertation research. Brad and Jessica kindly read this dissertation and provided helpful comments. The Molecular and Cellular Biology Graduate Program and its administrators at both the University of Washington and the Fred Hutchinson Cancer Research Center, particularly Michele Karantsavelos, have furnished essential support.

This project was supported by Public Health Service grant AI054423 (to N.R.S.) and NRSA T32 007270 from NIGMS (to I.E.C.). Its contents are the responsibility of the author and do not necessarily represent the official views of the NIH.

DEDICATION

This dissertation is dedicated to my family, for their unfailing love and support.

To my parents, Jonathan and Barbara Cohen,

for instilling in me a love of learning and of science,

and for giving me the confidence to pursue my dreams.

To my husband, Craig Bierle,

for his constant support and encouragement,

and for leading the way in our two PhD family.

To my beloved son, Nolan,

the greatest accomplishment of my years in graduate school.

CHAPTER 1

Background

Helicobacter pylori infection and pathogenesis

Helicobacter pylori is a gram-negative bacterium with a helical cell shape. It infects the human stomach and is found in approximately 50% of the world's population (1), where it persists for decades in a dynamic equilibrium with the host (2, 3). The infection causes chronic gastritis that is usually asymptomatic, but a subset of infected individuals will develop ulcers and gastric cancer (1).

To survive in the challenging environment of the stomach, *H. pylori* specifically localizes to the gastric epithelium and the mucus layer overlying it, where it is found extracellularly but in close proximity to the gastric epithelial cells (4). *H. pylori* strains can exhibit a preference for the corpus (the main body of the stomach) versus the antrum region (closest to the duodenum), which differ in the presence of acid-secreting parietal cells (5). Different patterns of gastritis have different effects on host acid secretion, leading to differences in disease risk, with antral-predominant gastritis (high acid) a risk factor for duodenal ulcer, and corpus-predominant gastritis (low acid) a risk factor for gastric cancer (6).

The origin of *H. pylori*'s association with humans has been dated to approximately 58,000 years ago, predating human migration out of Africa (7, 8). This long association has led to a complex relationship, in which infection has well-documented costs late in life but appears to also provide some benefits to the host (9). All colonized humans develop some degree of histological gastritis, with approximately 10-20% of infected individuals developing gastric or duodenal ulcers, 1-2% gastric adenocarcinoma, and less than 1% mucosa-associated lymphoid

tissue lymphoma (1). Gastric adenocarcinoma is the second leading cause of cancer deaths worldwide and approximately 75% of these cases are attributable to infection with *H. pylori* (10). On the other hand, *H. pylori* may provide a benefit to the host by protecting against pathological sequelae of gastro-esophageal reflux disease (GERD), including esophageal adenocarcinoma and its precursor lesion, Barrett's esophagus, by decreasing acid production in some patients (9). Infection with *H. pylori* may reduce the incidence of asthma and allergies, possibly by inducing regulatory T cells (9), and may protect against active tuberculosis, possibly by promoting strong T_H1 immune responses (11). *H. pylori* has also been proposed to decrease the risk of obesity through effects on the hormones leptin and ghrelin (9).

Genetic diversity, mutation, and recombination

H. pylori was the first bacterial species to have two representatives fully sequenced: strains 26695, isolated from a gastritis patient in the UK, and J99, isolated from a duodenal ulcer patient in the US (12, 13). The complete genome sequences of more than 50 *H. pylori* strains, isolated from patients around the world and with a range of clinical presentations, are available to date (14). These sequences have revealed a conserved core genome of approximately 1200 genes, as well as considerable differences among strains in gene content, sequence, and order (14).

Many mechanisms contribute to the observed genetic diversity in *H. pylori*, including point mutations, inter- and intragenomic recombination, and transcriptional and translational phase variation (15, 16). Estimates of the spontaneous mutation rate in *H. pylori* are variable. The rate of mutation to rifampicin resistance in the strain background used in this work has been measured at approximately 3×10^{-8} per cell per division (17), comparable to the spontaneous mutation rates classically measured for *E. coli* (18). Other studies examining panels of *H. pylori*

clinical isolates have found much higher frequencies of rifampicin resistance, up to 3×10^{-5} in some strains (19, 20). This wide range of mutation rates may reflect differences in the methods used and/or true differences among strains of *H. pylori.*

Recombination between *H. pylori* strains is extremely common. *H. pylori* exhibits natural competence for uptake of DNA from its environment (21). Competence is mediated by the Com apparatus, a type IV secretion system (T4SS) with homology to the prototypical *Agrobacterium tumefaciens* T4SS (22), and homologous recombination can integrate imported DNA into the *H. pylori* genome (23). Recombination-mediated changes in virulence factors and surface exposed proteins, such as deletion of the *cag* pathogenicity island (cagPAI) and allele changes in outer membrane proteins, have been observed in clinical isolates (24) and may benefit *H. pylori* by altering the degree to which it interacts with its host. Recombination is not entirely unlimited in *H. pylori,* however. Strain-variable restriction-modification (R-M) systems can degrade DNA from strains that do not possess the corresponding modification gene (16).

Intragenomic recombination mediated by direct repeats and other regions of sequence similarity within the genome can lead to deletion or duplication of the intervening segments, or to changes in gene order (25). The presence of direct repeats at certain sites may allow some degree of targeting of diversification to particular loci that are best suited for producing adaptive change (26). Both transcriptional and translational phase-variation in the on/off status of expression can be mediated by slipped-strand mispairing of repetitive sequences during DNA replication; these sequences are found in dozens of *H. pylori* genes (15). Changes in the number of repeats in promoter regions can alter transcription levels, while changes in the number of repeats in coding regions can create frameshifts that prevent expression of functional proteins (15).

Adaptation within a host

H. pylori infection is chronic in the absence of specific antimicrobial treatment (27), and bacterial adaptation to the host environment during the course of this chronic infection has been observed in individuals from whom *H. pylori* isolates separated by years have been studied. An important caveat in all studies employing repeated sampling is that any bacterial genetic differences between the two timepoints may have been present, but not detected, at the time of initial sampling. Bacterial isolates separated by 7-10 years underwent genotypic changes, as measured by randomly amplified polymorphic DNA (RAPD) or amplified fragment length polymorphism (AFLP) profiles, but the changes were not mapped to specific genes or sequences (28). A study of multiple isolates obtained from the same patient 6 years after the isolation of J99, one of the original sequenced *H. pylori* strains, revealed changes in gene content by microarray, but no differences in sequence at four loci (29). Genome sequencing of *H. pylori* isolates recovered before and after the development of gastric cancer revealed limited bacterial genetic changes (30).

Paired isolates separated by an average of 2 years from 26 subjects showed point mutations, evidence of recombination, and changes in gene content in some pairs, but no changes in other pairs (31–33). Follow-up of a subset of these chronically-infected subjects, including additional sampling 16 years after initial *H. pylori* isolation, using whole genome sequencing revealed the full spectrum of genetic changes occurring over time (34). These included dozens to hundreds of single nucleotide polymorphisms (SNPs), as well as clustered changes indicative of recombination with co-infecting *H. pylori* strains. However, limited sequence differences were observed during the course of a 3-month experimental infection (34). This is consistent with a

previous study, which found no changes in gene content over a period of weeks to months in four subjects following the same experimental infection protocol (35).

Mixed infection with more than one distinct *H. pylori* strain has been observed (36). Recombination between genetically distinct strains during a mixed infection acts a source of diversity by creating new combinations of alleles (24) and is thought to be a major source of the diversity observed over time in infected humans (34). This intra-host diversity could contribute to adaptation to different niches within the stomach, and may also provide pre-existing genetic variation to facilitate adaptation to changing conditions in the stomach over time.

Exposure to reactive oxygen and nitrogen species, which *H. pylori* is likely to encounter due to host immune responses, leads to transient increases in the rates of point mutations and both inter- and intragenomic recombination (20). This may facilitate the generation of variants that are better adapted to stressful conditions. DNA damage induces natural competence-mediated transformation, which may facilitate diversification in response to stress (37). Naturally competent *H. pylori* adapt more quickly to selection for rapid growth in an in vitro passaging experiment than do isogenic non-competent strains (17), highlighting the importance of DNA exchange for adaptation to selective pressures. Natural competence is also important for establishment of a persistent infection, but not for initial colonization, in a mouse model of *H. pylori,* indicating that diversification may allow the bacteria to respond to long-term immune pressures (38).

H. pylori transmission and adaptation to a new host

H. pylori infection is most commonly acquired during childhood (39) and transmission is thought to occur by an oral-oral, gastro-oral, or fecal-oral route (40). Transmission often occurs within families or among close household contacts. Children have been thought to acquire *H.*

pylori primarily from their mother (41) or siblings (42). This nuclear family transmission pattern has been found to be more common in developed areas with lower *H. pylori* prevalence, while patterns of transmission in higher-prevalence regions are more closely correlated with household residence than with familial relationship (43). A more recent study also found transmission to correlate with both household residence and sibling, but not parental, relationships (44). Household crowding appears to promote transmission (45).

Transmission to the stomach of a new host can present the bacteria with altered conditions, including differences in the host's immune response to the infection, and these differences are likely to impose selective pressures on *H. pylori*. Studies of the bacterial genetic changes accompanying transmission from person to person have largely focused on sequencing of a very limited number of genes (33, 46) or used microarrays to examine the presence or absence of genes (47, 48). These studies have demonstrated the relatedness of strains thought to have been transmitted among family members while indicating the presence of varying levels of diversification. However, the experimental procedures employed in these studies are limited in their ability to detect the full spectrum of genetic changes that may occur following transmission.

Whole genome sequence comparison of transmitted strains has been limited to a single pair of strains believed, due to their similarity, to have been transmitted between adult spouses (49). Comparison of these strains revealed 31 SNPs and 10 instances of intra-genomic recombination, with evidence of purifying selection on essential housekeeping genes and diversifying selection on surface structures potentially exposed to the host immune response (49). In comparison to human-to-human transmission, transmission and adaptation of *H. pylori* to a new species are likely to involve additional selective pressures, due to factors such as anatomical differences in the stomach and more divergent immune responses.

Animal models of *H. pylori* infection

H. pylori naturally infects only humans and some non-human primates, including rhesus macaques (50). The bacteria have been adapted for laboratory study in several small animal model systems (51). Mongolian gerbils infected with *H. pylori* can develop severe pathology that includes gastric cancer (52), but studies of host factors involved in infection are limited by their outbred status and lack of knockout strains.

Mouse models of *H. pylori* are desirable because of their relatively lower cost and the wide variety of genetic tools available. *H. felis,* a closely related species to *H. pylori*, robustly infects the murine stomach and can cause pathology similar to that observed during human chronic infection with *H. pylori* (53). Transgenic mice expressing the human Lewis B antigen in mucous-producing gastric pit cells (54) or lacking the p40 subunit of the cytokine IL-12 (55) are more permissive for infection with clinical isolates of *H. pylori.* Several *H. pylori* strains have been adapted to infect wild type mice by serial passaging (38, 56–59) but the mechanisms underlying the change in host range have not been fully explored.

Most strains of mice do not mimic the full range of pathology observed in *H. pylori*-infected humans, generally developing mild gastritis that does not progress to ulcers and cancer, which limits the ability to study these important outcomes of infection (51). Mice genetically altered to be more prone to gastric cancer can alleviate this limitation (60). Many murine models of *H. pylori* infection result in only transient or low-level colonization, in contrast to the robust, lifelong infection observed in infected humans.

Despite these limitations, the murine model has yielded many important insights into the biology and pathogenesis of *H. pylori*. Many virulence factors required for survival and colonization in the stomach have been identified using the mouse model. These include flagellar

motility and chemotaxis used to reach *H. pylori*'s preferred niche at the gastric epithelial surface (61, 62); outer membrane adhesins used to bind to the gastric epithelium and maintain the bacterium's position in the niche (63); factors involved in maintaining the bacterium's distinctive helical cell shape (64–66); and a urease enzyme that neutralizes the local pH around the bacterium to protect it from the acidic lumen of the stomach (67).

Other *H. pylori* models and selective conditions

In addition to animal models, *H. pylori* has also been extensively studied in laboratory culture systems. Coculture of the bacteria with human gastric epithelial cell lines has allowed investigation of their interactions with cells similar to those encountered during infection of the human stomach (68). An *H. pylori* isolate has been selected for adherence to and growth on the surface of Madin-Darby canine kidney (MDCK) cells, which form a polarized monolayer in culture (69). This model has facilitated investigation of the effects of *H. pylori*, particularly the translocated bacterial effector protein CagA, on cellular polarity. The MDCK-adapted *H. pylori* isolate was found to adhere to the apical cell surface primarily near cell-cell junctions, resulting in disruption of junctions and of barrier function (69). *H. pylori* growth was enhanced by its ability to acquire the essential nutrient iron from the polarized epithelium through disruption of host transferrin recycling, which requires binding of *H. pylori* to the host cells (70).

H. pylori can be grown on both solid and liquid media under laboratory conditions, and repeated passaging in liquid culture for approximately 1000 generations resulted in increased fitness of adapted isolates in competition with the ancestral strain (17). The increase in fitness was reduced when *H. pylori* that were not naturally competent for DNA uptake were passaged, highlighting the importance of genetic exchange for bringing together beneficial mutations into the same cell for optimal fitness (17).

Project Aims

To more fully investigate how the bacteria respond to the selective pressure imposed by transmission, I have sequenced the complete genomes of *H. pylori* strains at several stages of adaptation to mice by serial passaging, as well as strains subjected to other selective pressures (Figs. 1.1, 1.2). Whole genome sequencing allowed identification of a relatively small number of bacterial genetic changes that occurred during selection, described in Chapter 2. As described in Chapter 3, testing the effects of specific isogenic mutations on mouse adaptation revealed that a loss of function mutation in the previously uncharacterized gene *HPG27_792* had the largest single effect on mouse colonization ability. The protein encoded by *HPG27_792*, here named Imc1 (inhibitor of mouse colonization), is associated with the bacterial outer membrane, but does not appear to affect either the murine immune response to *H. pylori* infection or bacterial adherence to the host cell surface. Chapter 4 describes the identification and characterization of a mutation in *cagY* that occurred subsequent to the *imc1* mutation event. CagY is a component of the Cag type 4 secretion system (T4SS), which delivers bacterial products to host cells and is associated with an increased risk of gastric cancer (71, 72). Chapter 5 briefly describes mutations identified in *H. pylori* strains subjected to the selective pressures of adherence to and growth on cultured mammalian cells and replication in vitro. Chapter 6 presents conclusions and proposes future experiments based on the work in this dissertation.

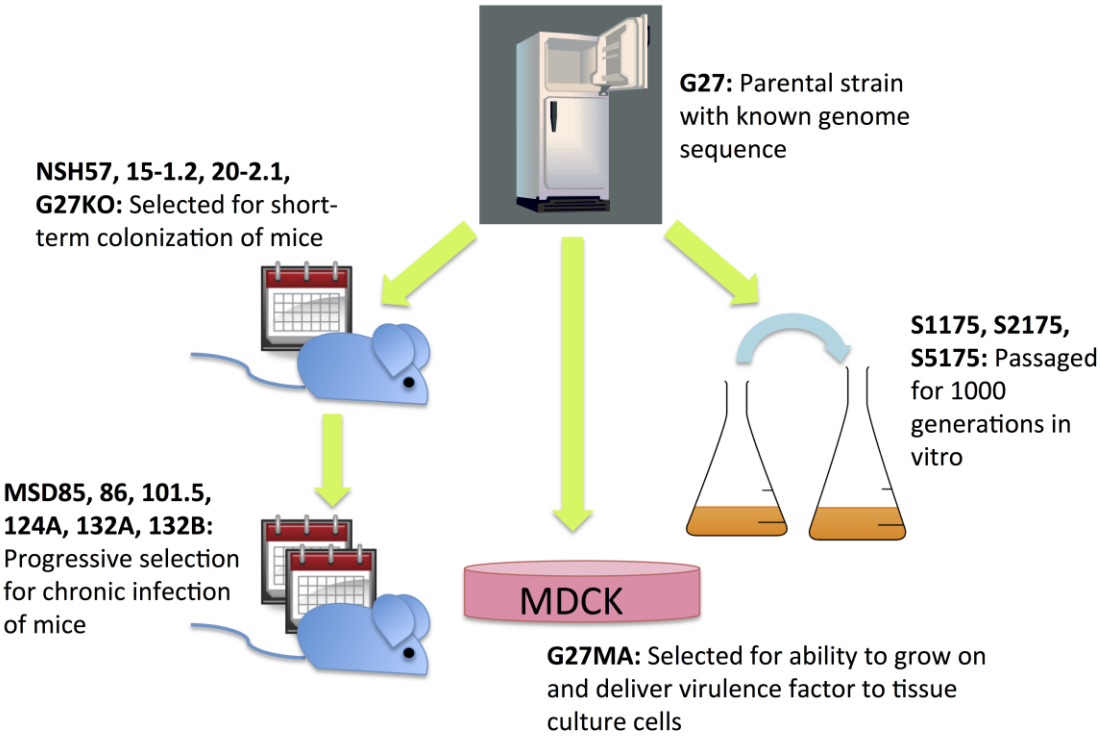

Figure 1.1 Relationship of strains sequenced in this project. G27 was originally isolated from a patient and was used to derive the strains sequenced in this work. MDCK: Madin-Darby Canine Kidney cells.

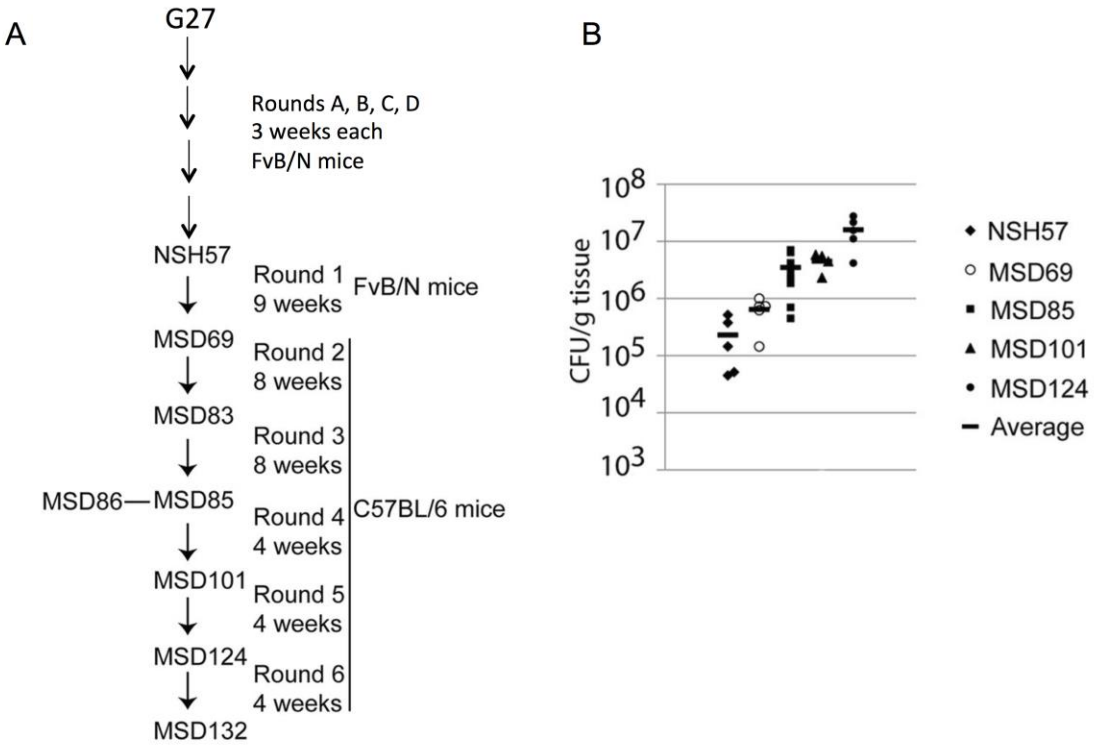

Figure 1.2 Adaptation of *H. pylori* clinical isolate G27 to infect mice by serial passaging. (A) Schematic of mouse passaging strategy. (B) Increasing mouse stomach colonization load of *H. pylori* with successive rounds of serial passaging in mice. Each data point represents the CFU per gram of stomach tissue recovered from a single mouse, after the infection time indicated in (A). Adapted from (38).

CHAPTER 2

Genome sequencing of adapted strains of *H. pylori*

INTRODUCTION

Next-generation sequencing is a powerful tool for many experimental approaches, including identifying mutations, measuring gene expression, and characterizing protein-DNA interactions (73). Next-generation sequencing differs from traditional, Sanger sequencing in its scale and throughput. Although the specifics of both library construction and sequencing chemistry differ among platforms, in general, hundreds of thousands to billions of spatially-separated reaction foci are sequenced simultaneously during a single instrument run, leading to the common description of next-generation sequencing as "massively parallel." Rather than an initial cloning step, next-generation sequencing generally begins with the construction of a library of short DNA fragments, each of which is covalently linked to a platform-specific adaptor. These adaptors are used both for amplification of the library and for initiation of the sequencing reaction. Sequencing and detection are coupled, rather than requiring a separate electrophoresis step, as in traditional sequencing. (74)

Per-base error rates are generally considerably higher for next-generation sequencing platforms than for Sanger sequencing, with each platform having a characteristic error profile; however, the higher per-base error rate can often be counteracted by the increased sequencing depth obtained. Although read lengths continue to increase as the technologies mature, next-generation sequencing platforms generally produce shorter reads than Sanger sequencing. (74, 75) These short reads are often difficult or impossible to accurately assemble de novo, but can be readily aligned to existing, high-quality reference genomes (73, 76), which exist for many model organisms including *H. pylori* (77).

A variety of next-generation sequencing technologies have been developed with varying levels of scientific and commercial success. The two approaches used in this work, 454 pyrosequencing, marketed by Roche Diagnostics, and Illumina sequencing, marketed by Illumina, Inc., have been among the most widely adopted in the early years of next-generation sequencing.

454 pyrosequencing was the first commercially available next-generation sequencing platform. In this approach, adaptor-flanked DNA libraries are amplified on beads by emulsion PCR, and the amplicon-bearing beads are deposited on a plate at single bead per well density. Sequencing primers hybridize to the adaptors, unlabeled nucleotides are added (a single type at a time), and incorporation is detected by the conversion of released pyrophosphate to light via ATP sulfurylase and luciferase. This sequencing chemistry leads to the main drawback of the 454 platform, its relatively high rate of errors in homopolymeric tracts. No mechanism prevents multiple incorporations of the same nucleotide in a given cycle, therefore the length of runs of the same base must be inferred from signal intensity, and this is prone to error, especially for longer homopolymers. The primary advantage of the 454 platform is its several hundred base-pair read length. (75)

The sequencing technology initially developed by Solexa, Inc. and subsequently acquired by Illumina, Inc. also begins with adaptor-flanked DNA libraries. In this approach, libraries are amplified on a solid surface, the flow cell, by a method known as bridge PCR to produce spatially separated clusters. A primer complementary to the adaptor sequence is used to initiate sequencing. A major difference in the Illumina compared to the 454 platform is the use of reversible dye terminator chemistry. All four nucleotides are provided in each cycle, labeled with distinguishable fluorophores and with a blocking group that prevents the incorporation of more

than one base per cycle. Each cycle consists of incorporation of one nucleotide, washing away of unincorporated nucleotides, imaging of the fluorescent signals, and cleavage of both the fluorophore and the blocking terminator to allow the next cycle to begin. (74) This sequencing chemistry makes each base addition an independent event, resulting in low rates of insertion and deletion errors and increasing accuracy for homopolymeric tracts versus other sequencing strategies, particularly 454 pyrosequencing (78).

Read lengths for the Illumina platform have increased from 25 bp to 300 bp as of 2014. Multiplexing allows DNA from distinct samples to be pooled and sequenced together on a single lane of the Illumina flowcell, with a unique sequence tag, or "barcode," allowing each read to be assigned to its source (79). This is particularly useful for organisms with small genomes, such as *H. pylori*, which has an approximately 1.6 Mb genome (77).

The original method for preparing DNA libraries for sequencing on the Illumina platform involved physical fragmentation of genomic DNA, repair of fragment ends, and ligation of sequencing adaptors containing primer binding sites and, optionally, barcodes (80). An alternative strategy, originally marketed as "Nextera" by Epicentre and now adopted by Illumina, uses a transposase enzyme to simultaneously fragment DNA and add tags to the fragment ends. These tags can be used to attach sequencing adaptors (which may be barcoded) to the fragments by PCR (80), (Fig. 2.1). The transposase-based strategy has many advantages over physical fragmentation and ligation. It is faster and involves fewer steps, with reduced opportunity for user error or sample loss. The transposon fragments DNA with high efficiency, allowing reduced DNA input, and produces a more uniform fragment size distribution than physical methods, eliminating the need for gel-based size selection. A bias in transposase insertion site has been

observed, but was found to have limited impact on coverage distribution for bacterial genomes (81). Both methods were used in this work.

The algorithms MAQ (Mapping and Assembly with Qualities) (82) and BWA (Burrows-Wheeler Alignment) (83) align the short reads generated by next-generation sequencing to a reference genome and identify positions that differ from the reference. For single end reads, MAQ produces only ungapped alignments, restricting it to the detection of single nucleotide polymorphisms (SNPs), while BWA can tolerate gaps in alignments, allowing the SAMtools algorithm (84) to detect indels in addition to SNPs. Both algorithms take into account the quality scores of both base calls and read mapping, as well as read coverage, in calling variants. The *breseq* computational platform (85) is optimized for microbial genomes and can detect larger deletions (as missing coverage) and rearrangements (as reads that map to more than one location), in addition to SNPs and small indels.

The application of whole genome sequencing to pre- and post-adaptation bacterial isolates has facilitated the discovery of an array of genetic changes occurring during adaptation to a variety of conditions. Sequencing of isolates from Richard Lenski's classic *Escherichia coli* in vitro passaging experiment revealed the near-linear accumulation of 45 mutations over 20,000 generations, several of which were found in parallel lines and have been shown to have effects on fitness (85). Comparison of the genomes of *Burkholderia dolosa* isolates collected over a 16-year period from the airways of multiple cystic fibrosis patients revealed parallel adaptive evolution to the cystic fibrosis lung in this opportunistic pathogen. 17 bacterial genes were found to have acquired mutations in multiple individuals, suggesting the importance of their pathways, which included LPS biosynthesis, antibiotic resistance, and oxygen sensing, in host adaptation. (86) A rare instance of chronic carriage of the normally acute pathogen *Burkholderia*

pseudomallei over a period of 12 years was accompanied by point mutations, indels, and gene loss, including inactivation of the capsular polysaccharide biosynthetic machinery; this mutation was postulated to contribute to attenuation of virulence in the isolate (87). The early stages of adaptation of *E. coli* to the mouse gut appeared to be driven by loss of function mutations in genes involved in metabolic processes, which were observed in multiple parallel lines (88).

This chapter describes the application of whole genome sequencing to identify mutations occurring during adaptation of *H. pylori* to several conditions, including colonization and persistence in the stomach of mice, growth on polarized mammalian cells, and replication in vitro.

RESULTS

Whole genome sequencing of *H. pylori* reveals adaptation-related mutations and improves a reference sequence

To investigate the genetic basis of *H. pylori* adaptation to new hosts and growth conditions, we sequenced the genomes of strains derived from a single clinical isolate with an existing genome sequence (77) that had been subjected to a variety of selective pressures, including colonization and persistence in the mouse stomach (38), adherence to and growth on tissue culture cells (69), and growth during in vitro passaging (17) (Table 2.1). Two rounds of multiplexed, single end Illumina sequencing resulted in total coverage ranging from 5- to 20-fold per strain (Table 2.2). Because of the differing methods of sequence library preparation employed in the two rounds, illustrated in Fig. 2.1, the numbers of reads mapped to their strain of origin and aligned to the reference genome were variable. In the first round of sequencing, library preparation included an adaptor ligation step (Fig. 2.1 A), which resulted in adaptor self-ligation, observed as reads that could not be aligned to the reference and carried a strong

compositional bias beyond the expected single adaptor sequence (Fig. 2.2 C). This adaptor ligation occurred to varying degrees across samples, contributing to the variability in overall coverage seen in Table 2.2.

In total, 282 putative single base changes (Tables 2.3, 2.4) were identified by alignment and comparison to the reference sequence. Fig. 2.2 depicts the detection of both a single nucleotide polymorphism (SNP, A) and a single base insertion (B) in two strains with varying levels of sequence coverage. Over 90% of the subset of SNPs tested (89 of 94 instances in 13 of 14 genes) and 100% of the subset of indels tested (73 of 73 instances in 7 of 7 genes) by PCR and Sanger capillary resequencing were confirmed to be present, demonstrating the accuracy of this method for identifying mutations. Changes not subjected to resequencing validation were classified as high or low confidence based on the number of reads supporting them and the quality and confidence of the base calling, following visual inspection of the aligned reads.

41 of the identified changes were classified as probable errors in the G27 reference sequence, due to their detection across most strains, including in the resequencing of G27 itself (Table 2.4, first section). 40 of the 41 are single base insertions or deletions, most of which map to homopolymeric tracts (77). These corrections often lead to a gene annotated as "potentially non-functional due to frameshift" being re-annotated as full length, improving the usefulness of the reference sequence.

Complementary sequence analysis methods reveal the spectrum of mutations in adapted *H. pylori* strains

Possible copy number changes were identified by comparing normalized read counts for each strain to those for the resequencing of the G27 reference strain, as depicted in Fig. 2.2 C. Although total coverage limited the identification of meaningful changes in read counts, 11

putative copy number changes were identified using a five standard-deviation threshold (Table 2.5). All of these changes were deletions, detected as a decrease in read coverage in a particular region; no significant increases in read counts suggestive of amplification were identified. The highest confidence deletions were found in the strains selected for growth in vitro and are described below. None of these putative deletions have been validated by PCR and therefore all must be treated with caution.

Possible rearrangements in the genome due to recombination were detected using *breseq* (85) as reads for which the ends of a single read map to two distinct locations in the reference genome, as depicted in Fig. 2.2 D. Thirteen of these potential new junctions were identified by *breseq* and are listed in Table 2.5. The majority of these putative rearrangements are between genes that are widely separated in the genome and do not have obvious functional similarities. Attempts to verify the presence of these new junctions by PCR spanning the site were unsuccessful, but further optimization of the PCR protocol would be required to draw definitive conclusions regarding their true presence or absence. In addition to rearrangements, *breseq* also calls SNPs and indels, and this program detected many, but not all, of the same single base variants previously identified using MAQ and BWA/SAMtools. A relatively small number of previously undetected single base variants were called by *breseq*, but only two of these were classified as high confidence based on their coverage and quality scores. Variants identified using *breseq* are included in Table 2.4, as well as in the totals in Table 2.3.

De novo assembly of the Illumina sequence reads for each strain resulted in a single, continuous contig for each strain when the coverage threshold was set to only a single read. Contig length ranged from 1,377,833 bp to 1,649,238 bp across strains, in comparison to the G27 reference sequence length of 1,652,982. When a more conservative coverage threshold of two

reads was applied, each strain was divided into a very large number of predominantly short contigs. Contig numbers ranged from 2,408 to 22,634 by strain, with average contig length of 73 to 686 bp. The very large number of contigs produced when requiring 2-fold coverage, as well as the sub-genome length contigs generated with 1-fold coverage, together indicate that the sequence coverage was insufficient to robustly assemble genomes de novo and obtain meaningful information about genome rearrangements or deletions.

454 sequencing was performed on strain MSD85, chosen because it was relatively poorly covered by the two rounds of Illumina sequencing, with total coverage of 5.4-fold (average for all 17 sequenced strains: 11.7-fold). 454 sequencing produced approximately 20-fold coverage of strain MSD85, with average read length of 425 bases. Analysis focused on the gene *cagY* in an attempt to confirm the possible deletions in mouse-adapted strains that were suggested by the copy number analysis (Table 2.5). However, despite the longer read lengths of 454 as compared to Illumina sequencing, *cagY* remained poorly covered due to difficulties in both sequencing and mapping reads in the repetitive regions of this gene. Although a single base insertion in *cagY* that was detected in the Illumina sequence data was also detected by 454 sequencing, no new variants in this gene were detected with confidence. 454 sequencing also identified a putative 63 bp deletion in the intergenic regions between genes *HPG27_52* and *HPG27_53*, which was supported by an absence of Illumina reads mapped to this region (Table 2.5).

Mutations specific to strains grown in vitro or with cultured mammalian cells

The gene *sabB* appears to have been deleted in strain G27 DB1, as evidenced by an absence of confidently mapped reads across the gene in G27 DB1 as well as its derivatives S1175, S2175, and S5175, passaged for approximately 1000 generations in vitro. Additionally, the nitrite extrusion gene *narK* was partially deleted in strain S5175. Evidence of three distinct

recombination events was found in strain G27 MA, selected for growth on and adherence to cultured mammalian cells. Regions of 455, 416, and 877 base pairs in genes *HPG27_910*, *hofH*, and *HPG27_1112*, respectively, had numerous differences from the G27 reference sequence but were an exact match to the homologous regions in another *H. pylori* wild-type strain, J99. Changes in these strains are described further in Chapter 5.

Mouse adaptation enriches for mutations in cell envelope genes

I chose to focus on the strains selected for the ability to colonize and persist in mice, a non-natural host for this human pathogen. I identified 27 amino acid-altering mutations with high confidence (either verified by Sanger sequencing or supported by a minimum of three reads with high quality scores and unambiguous base calls) in the lineage of mouse-adapted strains (Fig. 2.3), as well as 4 putative copy number changes and 12 putative rearrangements in these strains (Table 2.5). Of the 27 amino-acid altering changes, 18 were non-synonymous single nucleotide polymorphisms (SNPs) and 9 were one to two base insertions or deletions (4 insertions, 5 deletions). Only 2 synonymous SNPs were present in the validated and high confidence data. The elevated ratio of nonsynonymous to synonymous mutations in the mouse-adapted lineage (27 : 2) is consistent with strong selective pressure exerted by the change in host.

The cell envelope, including outer membrane proteins (OMPs), was a frequent target of mutation in the mouse-adapted strains (Fig. 2.4). OMPs are a large family that includes adhesins, porins, transporters, and other proteins found in the *H. pylori* outer membrane (89). Of the 27 genes with amino acid-altering mutations in mouse-adapted strains, 5 (19%) were OMPs and a further 5 (19%) were other genes associated with the cell envelope, including peptidoglycan synthesis and recycling, lipopolysaccharide and lipid A synthesis, and the Cag type IV secretion system. A strain selected for growth on cultured mammalian cells (G27 MA) had an even greater

enrichment for mutations in cell envelope-associated genes, with 4 of 16 (25%) mutated genes encoding OMPs and a further 4 of 16 (25%) encoding other cell envelope factors. In contrast, OMPs make up only 4% of the *H. pylori* genome, with other cell envelope genes representing a further 6% of the genome. This overrepresentation of mutations in genes involved in the outer membrane and cell envelope compared to their proportions in the genome was statistically significant for both mouse adapted ($p < 0.0001$) and tissue culture-adapted ($p < 0.0001$) strains, while strains selected for growth in vitro were not enriched in mutations in these gene classes (4 of 15 genes, $p = 0.0862$).

During the selective process separating G27 (the original clinical isolate) and NSH57 (mouse colonizing for up to 4 weeks, (59)), five protein-altering mutations were identified: a previously-characterized SNP in *fliM* (90) that leads to a bias towards counter-clockwise flagellar rotation directional switching, resulting in a smooth swimming pattern; truncation of *oppD*, encoding an oligopeptide permease ATPase; a SNP in *HPG27_557*, encoding the homolog of penicillin binding protein 1A, the sole bifunctional peptidoglycan synthase in *H. pylori*; truncation of *HPG27_792*, encoding a hypothetical secreted protein; and truncation of *HPG27_1188*, encoding a hypothetical protein. Strain G27 KO, created by an independent replication of the mouse passaging conditions used to select for NSH57 (91), shared no mutations in common with NSH57. Three mutations were identified in G27 KO: truncation of *HPG27_255*, a encoding a tryptophan biosynthetic protein; a SNP in *flgH*, encoding a flagellar basal body protein; and a SNP in *dnaA*, encoding a chromosomal replication initiator protein.

Additional protein-altering mutations occurred during further selection of the mouse-colonizing strain NSH57 to generate strains that persist for weeks to months in mice. Between NSH57 (mouse colonizing) and MSD85 and 86 (clones isolated from the same round of selection

in mice that differ in function of the Cag type IV secretion system (T4SS), (38)), there were three changes common to both strains, two exclusive to MSD85, and four exclusive to MSD86. The shared changes were a SNP in the OMP *hopC*; a SNP in *dnaE*, encoding the alpha subunit of DNA polymerase III; and truncation of *HPG27_676*, encoding a hypothetical protein that may be a zinc transporter. Changes unique to MSD85 were a frameshift of *hopH*, encoding an OMP; and truncation of *cagY*, encoding a protein required for T4SS function. Changes unique to MSD86 were a SNP in *HPG27_610*, encoding a UDP-N-acetylglucosamine 1-carboxyvinyltransferase involved in peptidoglycan synthesis; a SNP in *HPG27_1024*/*homC*, encoding an OMP; a SNP in *HPG27_1431*, encoding a hypothetical protein; and a SNP in *HPG27_1442*, encoding a restriction enzyme. Further selection of strain MSD85 for mouse persistence, resulting in strains MSD132A and B (clones from the same round of selection in mice), resulted in five changes: a SNP in *fliI*, encoding a flagellar ATP synthase; truncation of *HPG27_1047*, encoding a hypothetical protein; a frameshift resulting in fusion of the DNA methylase *HPG27_1444* to the downstream gene, also annotated as a DNA methylase; a SNP in the DNA repair factor *mutY*; and a SNP in *HPG27_537*, encoding a hypothetical protein.

There was a relative concentration of mutations during early stages of mouse adaptation, between the original strain G27 and the strain NSH57, and relatively fewer mutations per step at later stages of the adaptation process. The early stage mutations are further characterized in Chapter 3, and later mutations are described in Chapter 4. Mutations in the other sequenced strains are described in Chapter 5.

DISCUSSION

Whole genome sequencing is a powerful tool for identifying mutations in bacterial isolates. The small genome size of most bacteria, including *H. pylori* (approximately 1.6 Mb),

allows relatively high coverage to be obtained with relatively low costs. The addition of "barcode" tag sequences to target DNA allows multiplexing of multiple strains onto a single Illumina flow cell. In the case of isolates that are closely related to a reference strain with an existing complete genome, as was the case for the strains sequenced here, alignment of sequence reads to the reference facilitates the identification of one to two base changes, including substitutions and small insertions and deletions. The identification of larger deletions, amplifications, and rearrangements is possible using computational tools optimized for microbial genomes, such as *breseq*, or from de novo sequence assembly. However, it appears that the coverage obtained in this study, which averaged approximately 10-fold, was insufficient for confident detection of these classes of changes. It is also likely that single base changes were missed in this sequence analysis due to insufficient coverage.

The identification of 41 apparent errors in the G27 reference sequence (77) will improve the annotation of the reference sequence, with potential utility to the *H. pylori* field. The majority of the errors are in homopolymeric tracts and are likely to be derived from the use of 454 sequencing in creation of the reference sequence (77). 454 sequencing technology has a higher rate of error in homopolymeric tracts than the cleavable terminator sequencing chemistry of Illumina sequencing, in which each base addition is an independent event (78). The corrections identified here often lead to a gene annotated as "potentially non-functional due to frameshift" being re-annotated as full length.

This work identified a significant overrepresentation of point mutations in genes encoding cell-envelope-associated proteins in *H. pylori* strains selected for both colonization of mice and growth with cultured eukaryotic cells, but not during passaging in vitro. One major class of cell-envelope-associated proteins in *H. pylori* is the outer membrane proteins (OMPs),

which include porins, transporters, adhesins, and other surface-exposed proteins (89). Allele changes in *H. pylori* outer membrane proteins have been reported in clinical isolates (24) and elevated rates of interstrain recombination were observed for genes of the *hop* family of OMPs in paired isolates separated by 3 to 16 years (34). OMP genes were also frequently mutated after transmission of *H. pylori* between spouses (49).

Changes in the bacterial cell surface may benefit *H. pylori* by altering the degree to which it interacts with its host. Some members of the OMP family are adhesins that bind host molecules, including BabA, which binds the Lewis b blood group antigen (92), and SabA, which binds the sialyl-Lewis x blood group antigen (93), while other OMPs may mediate adherence to unknown host molecules. Adherence to the host's epithelial cells may help *H. pylori* remain in its preferred niche in the stomach, away from the acidic lumen, while avoiding loss from the stomach upon emptying. Adherence also facilitates translocation of the bacterial effector CagA, which is delivered to host cells by the contact-dependent Cag T4SS, leading to disruption of host cell polarity and/or barrier functions and increasing bacterial access to nutrients (69).

However, interaction with the host can be costly for *H. pylori*. Close proximity to host cells places the bacteria in a position vulnerable to immune responses that could result in bacterial killing. Bacterial peptidoglycan entering cells via the Cag T4SS induces host inflammatory responses, such as release of the chemokine IL-8 and subsequent recruitment of neutrophils (94). Induction of host T_H1 responses with production of the pro-inflammatory cytokine IFNγ can limit bacterial replication and has the potential to completely eliminate the bacteria in mice (95). In addition to its effects on the bacteria, inflammation is also damaging to host tissues and plays a major role in pathological sequelae (95).

This dynamic creates a need to tightly regulate the amount of inflammation induced to create an environment in which *H. pylori* can persist in its host. Disruption of BabA-mediated binding to host Lewis b has been observed during adaptation of *H. pylori* in multiple animal models. In rhesus macaques both point mutation of *babA* and gene conversion by the paralog *babB* have been observed (96), as well as changes in expression of Lewis x and Lewis y (97). Phase variation to "off" status, point mutations, and replacement of the Lewis b-binding allele of *babA* with a non-binding allele of *babA* have all been observed during passage of *H. pylori* in gerbils and mice (98).

Five of the 27 (19%) genes in which mutations were detected in mouse-adapted strains were OMPs, while this gene family makes up only 4% of the *H. pylori* genome. Furthermore, five additional genes mutated during mouse adaptation fall into the broader category of "cell envelope associated." Even higher rates of mutations in OMPs and other cell envelope genes were observed for the strain selected for growth on polarized cells, with 25% of mutated genes falling into each of these two categories. These rates are remarkably similar to those reported following transmission of *H. pylori* between human spouses, where 7 of 28 genes with SNPs (25%) were OMPs (49). These results highlight the importance of modulating cell surface interactions and suggest a key role for alterations in bacterial/host interaction during adaptation to novel hosts, whether across or within species boundaries.

MATERIALS AND METHODS

Strains

Strains sequenced are listed in Table 2.1. All strains are derivatives of the previously sequenced human clinical isolate G27 (77, 99, 100). Mouse adapted strains NSH57 (59), MSD85, MSD86, MSD101, MSD124, and MSD132 (38) have been previously described. Additional mouse

adapted strains G27 KO (91), G27 15-1.2, and G27 20-2.1 were generated using the same procedure used to generate NSH57 (59). Briefly, G27 bacteria were used to inoculate FvB/N mice by oral gavage for three passages of three weeks each, in which pooled bacterial colonies from the previous round were used to inoculate the next round of mice. The fourth passage of three weeks took place in C57BL/6 mice and single colony isolates obtained after this passage were designated G27 15-1.2, G27 20-2.1 (from the same experiment that generated NSH57) and G27 KO (from a separate experiment following the same procedure). Strain G27 MA, selected for adhesion and CagA delivery to Madin-Darby canine kidney (MDCK) cells, has been previously described (69). Strains G27DB1, S1175, S2175, and S5175 are the ancestral G27 clone and three isolates obtained after approximately 1000 generations of in vitro passage, as described (17).

Genome Sequencing

DNA Extraction. Genomic DNA was prepared from *H. pylori* strains as described (99) with slight modifications. Briefly, bacterial growth from two plates was suspended in 0.1 M NaCl / 10 mM Tris HCl / 1 mM EDTA, pH 8, treated with lysozyme (340 μg ml^{-1} for 15 min at 37°C), heated for 15 min at 65°C with 1% SDS, and treated with proteinase K (100 μg ml^{-1} for 2 h at 50°C). DNA was purified by CsCl gradient centrifugation with ethidium bromide and precipitated with 77% EtOH as described (99) and resuspended in 100 mM Tris / 5 mM EDTA, pH 8 (TE). For strains MSD86, MSD101, and MSD132B, genomic DNA was prepared using the Wizard genomic DNA kit (Promega) according to the manufacturer's protocol, except that cells were heated in nuclei lysis buffer at 80°C for 5 min. DNA prepared using the Wizard kit was further purified by two extractions with an equal volume of a 25 : 21 : 1 mixture of phenol,

chloroform, and isoamyl alcohol, pH 8, followed by extraction with an equal volume of chloroform and precipitation.

Library preparation and Illumina sequencing. Barcoded sequencing libraries were prepared by two methods for each *H. pylori* strain (Fig. 2.1): 1) mechanical shearing of 5 μg genomic DNA followed by end repair, A-tailing, ligation of barcoded adaptors (Table 2.6), size selection, gel purification, PCR amplification, and additional size selection as described (80, 81); 2) Nextera (Illumina) transposase-based tagmentation starting with 50 ng genomic DNA followed by PCR addition of barcoded adaptors (Table 2.6) as described (81). For method 1, samples were sub-pooled in groups of two to three post-barcoding but prior to size selection. Prepared libraries were analyzed for concentration and size distribution using a Bioanalyzer (Agilent) and all 17 strains pooled at a final concentration of 10 nM (method 1) or 20 ng μl^{-1} (method 2). Cluster generation and sequencing were performed using Illumina reagents according to the manufacturer's standard procedures. Single end reads of 60 or 36 base pairs, respectively, were sequenced on an Illumina Genome Analyzer at the Fred Hutchinson Cancer Research Center's Genomics Shared Resource, or on an Illumina Genome Analyzer IIx in the laboratory of Jay Shendure, University of Washington.

Sequence analysis

Reference alignment. Initial quality control, base calling, and quality score assignment were performed by the Illumina pipeline, resulting in 9.4 million and 11.3 million reads for the two runs, respectively. The barcode sequence, contained in the initial 13 nucleotides of the read (method 1) or in a separate, linked 9 nucleotide index read (method 2) was used to assign each read to its strain of origin; 91% and 58% of reads, respectively, could be assigned for the two runs. Assigned reads were mapped to the existing G27 reference sequence (77) using the aligners

MAQ (Mapping and Assembly with Qualities (82)) and BWA (Burroughs-Wheeler Aligner (83)). 27% (round 1) and 94% of reads (round 2) were confidently aligned to the reference genome, with alignment rates ranging from 3-73% (round 1) and 93-94% (round 2) (Table 2.2). Differences in the efficiency of barcode assignment and read alignment are likely to be due to the differing methods of library preparation for the two rounds of sequencing. In round 1, prepared by physical shearing of DNA followed by adaptor ligation, a large number of reads were observed with "chains" of multiple adaptors that seem to have ligated to each other, leaving very little alignable bacterial DNA sequence (Fig. 2.1, C).

Single base variant calling. Differences from the reference sequence were identified by MAQ (82) and SAMtools (84). For single end reads, MAQ produces only ungapped alignments, restricting it to the detection of single nucleotide polymorphisms (SNPs), while BWA + SAMtools can tolerate gaps in alignments, allowing the detection of indels in addition to SNPs. Both algorithms take into account the quality scores of both base calls and read mapping, as well as read coverage, in calling variants. Additional sequence analysis was performed using *breseq* (85), which is optimized for microbial genomes and uses Bowtie2 to map reads to a reference sequence, then identifies differences from the reference, including SNPs, indels, larger deletions, and reads mapping to more than one location.

Detection of copy number variants and rearrangements. To detect larger deletions as areas of missing coverage, read counts in a 50 base pair sliding window were normalized to the total number of reads for the strain and compared to the normalized read count for G27_Original. Log_2 ratios exceeding a five standard deviation threshold were scored as positive (Table 2.5). Possible rearrangements in the genome due to recombination were detected using *breseq* (85) as reads for which the ends map to two distinct locations in the reference genome (Table 2.5). As a

complementary approach, sequence reads for each strain were assembled de novo using coverage thresholds of both 1 and 2 with the Perl script Pileuptofasta3, provided by the lab of Evgeni Sokurenko.

454 sequencing

454 library preparation and sequencing were performed by the FHCRC Genomics Shared Resource using a GS Junior Instrument (Roche). Approximately 93,000 reads with an average length of 425 bp passed initial quality filters. Reads were mapped to the G27 reference genome using *gsMapper v2*.6, and assembled de novo using *gsassembler*.

Sequence confirmation

Read alignments in the vicinity of identified mutations were visually inspected using the Integrative Genomics Viewer (www.broadinstitute.org/igv). A subset of mutations identified by whole genome sequencing was confirmed by PCR amplification and capillary sequencing of the region surrounding the mutation of interest, using the primers indicated in Table 2.6. Capillary sequencing was performed by the FHCRC Genomics Shared Resource using ABI's BigDye terminator cycle sequencing reagent on an ABI 3730xl DNA analyzer (Applied Biosystems).

Statistical analysis

Statistical analysis was performed using GraphPad Prism version 6.02 for Windows (GraphPad Software, La Jolla, CA). The distribution of genes with mutations was compared to their representation in the *H. pylori* genome using a two-tailed Chi-square test.

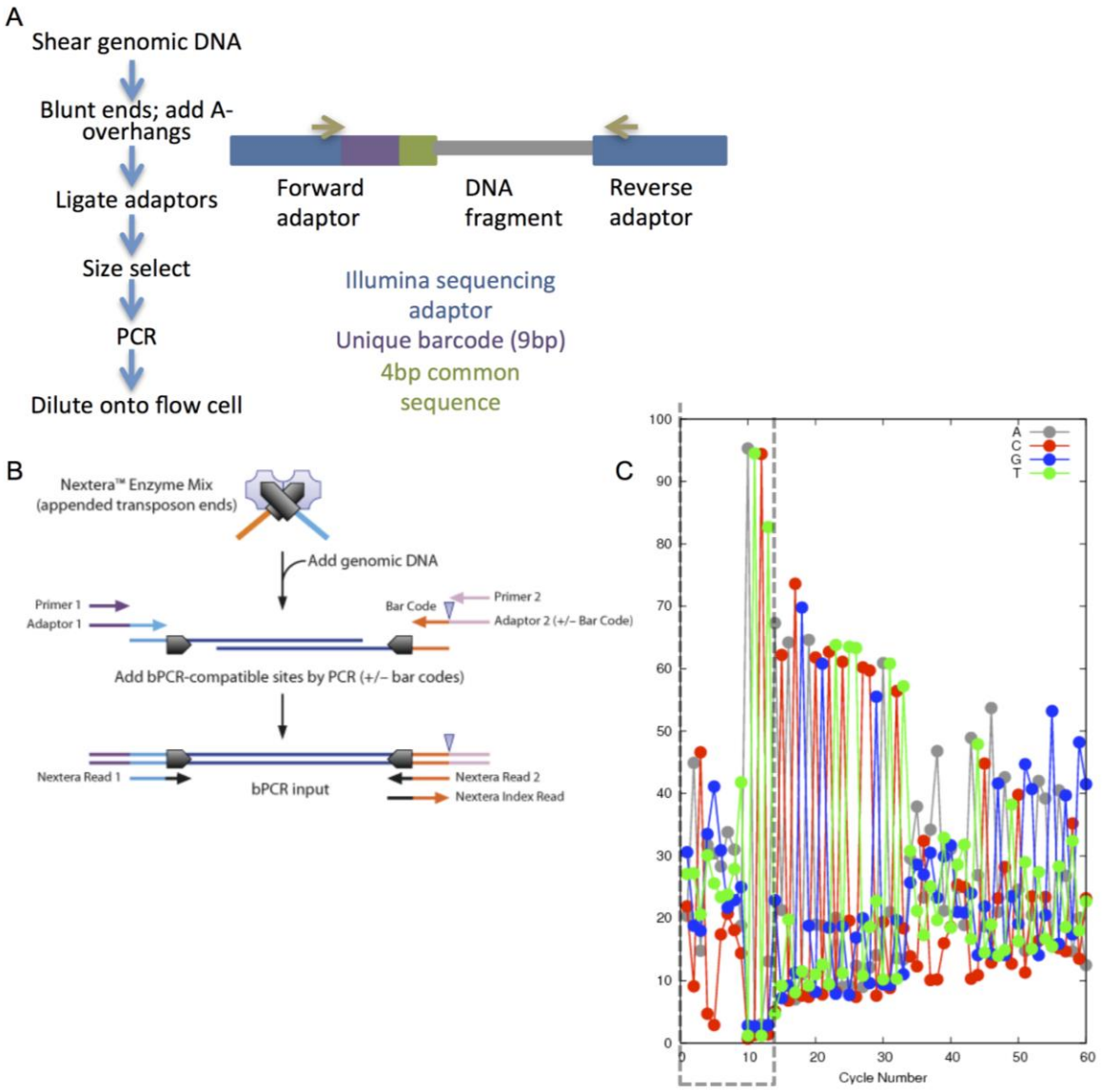

Figure 2.1 Preparation of sequencing libraries leads to differences in ability to align reads to the reference. (A) Traditional ligation-based library preparation. (B) Nextera tagmentation-based library preparation. (C) Ligation-based library preparation resulted in frequent adaptor-adaptor ligation, generating reads that do not align to the reference genome. The percentage of basecalls of a given nucleotide remained strongly biased beyond the expected 13 bp barcode and adaptor sequence (demarcated with a grey dashed line), indicating the presence of additional adaptors. Panel (A) is adapted from Emily Turner, personal communication; (B) is adapted from Epicentre Biotechnologies; (C) is adapted from Matthew Fitzgibbon, personal communication.

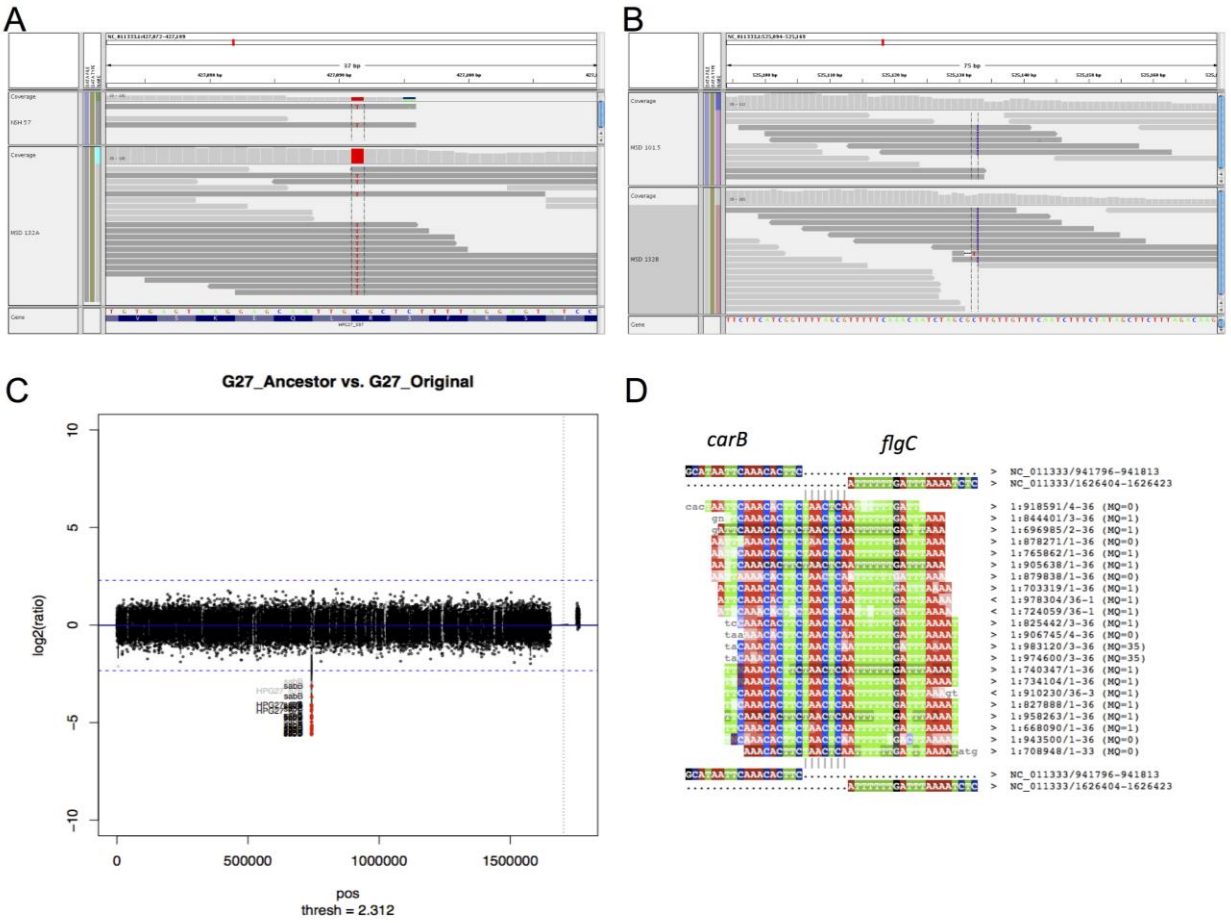

Figure 2.2 Identification of mutations from next-generation sequence data. (A, B) Reads aligned to the G27 reference sequence are visualized with the Integrative Genomics Viewer (IGV). (A) Calling a single nucleotide polymorphism (SNP) in *fliM*. (B) Calling a single base insertion in *cagY*. (C) Calling a deletion in *sabB* using the ratio of normalized read counts to the read counts for resequencing of G27. (D) Calling a rearrangement between *carB* and *flgC* using *breseq*.

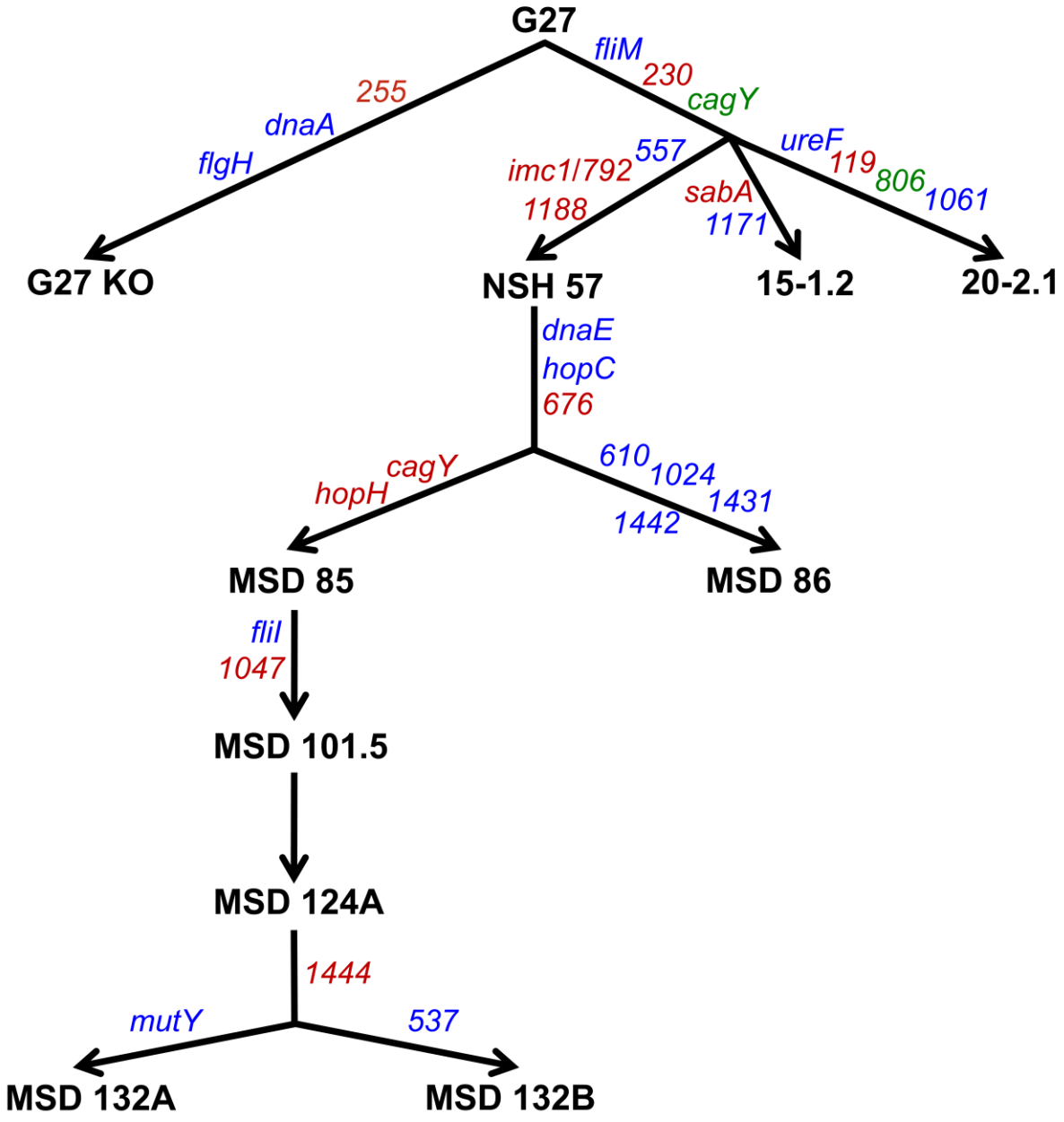

Figure 2.3 Mutations occurring during adaptation of *H. pylori*. Lineage of mouse adapted strains derived from clinical isolate G27. Genes with coding changes identified by whole genome sequencing are indicated at the point of first observation. Red, truncation caused by frameshift or nonsense mutation; blue, nonsynonymous SNP; green, synonymous SNP.

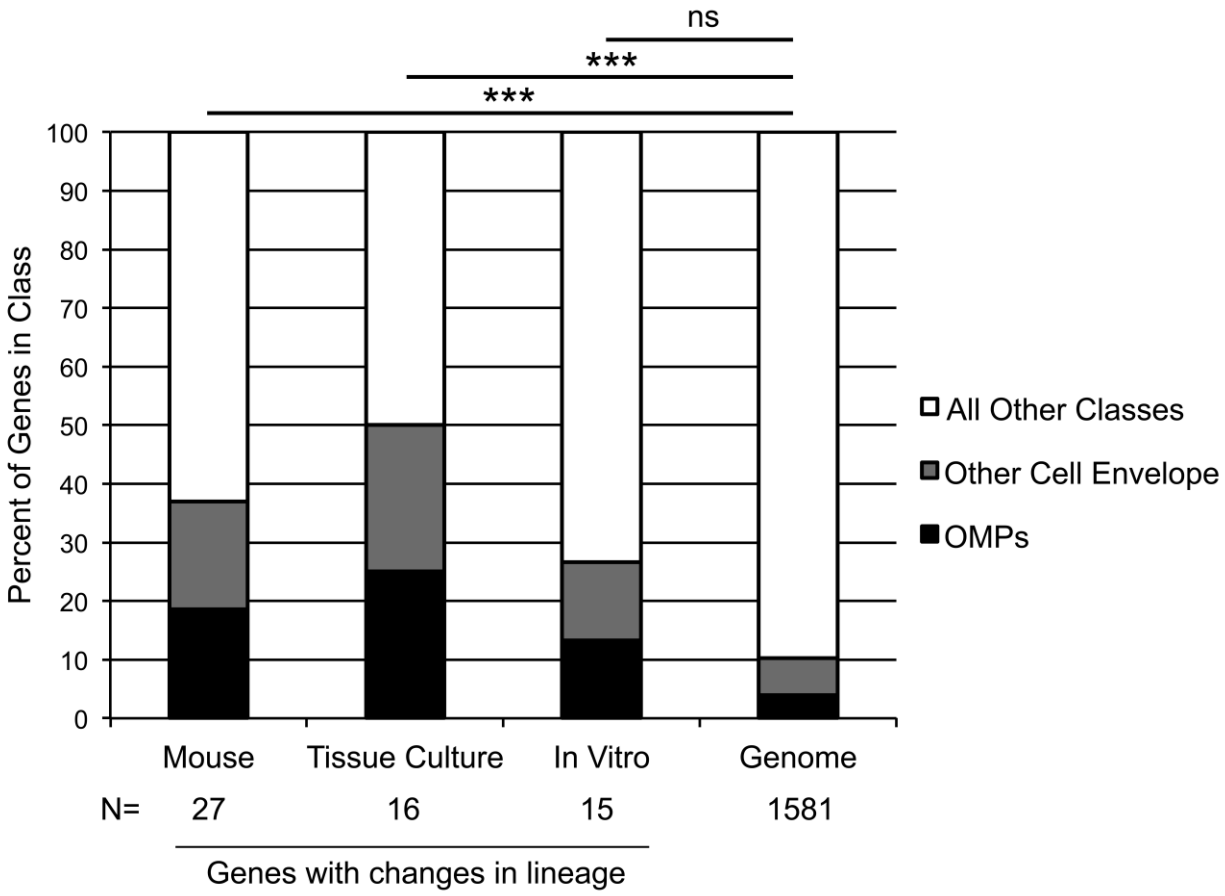

Figure 2.4 Enrichment of mutations in outer membrane and cell envelope proteins. The percentage of genes that have a coding change among strains selected for mouse adaptation, tissue culture, and in vitro growth that encode outer membrane proteins (OMPs) and other cell envelope components. The total number of genes with coding changes in each lineage is indicated. Right, representation of these gene classes in the G27 genome. ***, $p \leq 0.0001$; ns, $p > 0.01$ for comparison of the numbers of mutated genes in each class to their representation in the genome as determined by Chi-square test.

Table 2.1 Strains sequenced.

Strain Name	Genotype	Reference	Derived from
G27 Original	Wild-type G27	(99)	Endoscopy patient, Italy
G27 Working	Wild-type G27, working stock	(99)	Endoscopy patient, Italy
G27 KO	Mouse adapted from G27	(91)	G27
G27 15-1.2	Mouse adapted from G27	This work	G27
G27 20-2.1	Mouse adapted from G27	This work	G27
NSH57	Mouse adapted from G27	(59)	G27
MSD86	Mouse adapted from NSH57	(38)	NSH57
MSD85	Mouse adapted from NSH57	(38)	NSH57
MSD101	Mouse adapted from MSD85	(38)	MSD85
MSD124	Mouse adapted from MSD101	(38)	MSD101
MSD132A	Mouse adapted from MSD124	(38)	MSD124
MSD132B	Mouse adapted from MSD124	(38)	MSD124
G27 MA	Adapted to MDCK cells	(69)	G27
G27 DB1	Wild-type G27, working stock	(17)	G27
G27 S1175	Adapted to in vitro passaging	(17)	G27
G27 S2175	Adapted to in vitro passaging	(17)	G27
G27 S5175	Adapted to in vitro passaging	(17)	G27

Table 2.2 Assignment of single end sequencing reads to strains by barcode and alignment to the G27 reference genome.

Sequencing Round 1							
	Map reads to barcodes		**Align mapped reads to genome**				
Strain name	**Allow 1 mismatch (not used)**	**Exact match (used by BWA, Maq)**	**BWA Aligned Reads**	**BWA % Aligned**	**Maq Aligned Reads**	**Maq % Aligned**	**Fold Coverage**
G27 original	547955	531681	176786	33	178183	34	4.9
G27 working	604021	586363	63647	11	64459	11	1.8
G27 KO	501733	484105	241164	50	243398	50	6.7
G27 15-1.2	1585542	1542364	51347	3	51860	3	1.4
G27 20-2.1	267541	257344	83243	32	83957	33	2.3
NSH57	115492	111021	43611	39	43953	40	1.2
MSD86	547307	521546	112942	22	114916	22	3.1
MSD85	31322	29888	19266	64	19481	65	0.5
MSD101	517009	500222	221968	44	224234	45	6.2
MSD124	531830	506516	60107	12	60947	12	1.7
MSD132A	683848	654785	229021	35	231003	35	6.4
MSD132B	519891	504735	267938	53	270323	54	7.5
G27 MA	332690	321488	233832	73	235884	73	6.5
G27 DB1	453072	433839	142739	33	144129	33	4.0
G27 S1175	186378	180133	59368	33	59921	33	1.7
G27 S2175	368006	352106	166452	47	168015	48	4.6
G27 S5175	1125008	1087058	163627	15	165325	15	4.6
Not mapped to a barcode	527943	841394					
Total mapped reads	8918645	**8605194**					
Total reads	**9446588**	**9446588**					
% Reads mapped to barcode	94	**91**					
Total reads aligned			**2337058**		**2359988**		
% Mapped reads aligned			**27**		**27**		

Sequencing Round 2

	Map reads to barcodes		Align mapped reads to genome		
Strain name	**Allow 1 mismatch (BWA)**	**Exact match (used by BreSeq)**	**BWA Aligned Reads**	**BWA % Aligned**	**Fold Coverage**
G27 original	285289	157126	267741	94	5.8
G27 working	227340	126315	212205	93	4.6
G27 KO	672357	366489	629800	94	13.7
G27 15-1.2	455490	265494	427263	94	9.3
G27 20-2.1	381614	219475	359326	94	7.8
NSH57	763654	432284	713557	93	15.5
MSD86	440227	241824	412498	94	9.0
MSD85	235922	122475	221098	94	4.8
MSD101	222703	127079	208649	94	4.5
MSD124	271009	157180	254341	94	5.5
MSD132A	561205	329578	526771	94	11.5
MSD132B	449971	254911	424127	94	9.2
G27 MA	533341	296652	502042	94	10.9
G27 DB1	373851	207413	352783	94	7.7
G27 S1175	181673	96444	171166	94	3.7
G27 S2175	322047	181822	303566	94	6.6
G27 S5175	202599	114829	190718	94	4.2
Not mapped to a barcode	4761802	7644704			
Total mapped reads	**6580292**	**3697390**			
Total reads	**11342094**	**11342094**			
% Reads mapped to barcode	**58**	33			
Total reads aligned			**6177651**		
% Mapped reads aligned			**94**		

Sequencing Combined
Round 1 + Round 2

Strain name	Total BWA Aligned Reads	BreSeq Input Reads	BreSeq Aligned Reads	BreSeq % Aligned	Fold Coverage
G27 original	444527	688581	318813	46	10.8
G27 working	275852	712477	175269	25	6.4
G27 KO	870964	850396	556158	65	20.4
G27 15-1.2	478610	1807299	290975	16	10.7
G27 20-2.1	442569	476721	276498	58	10.1
NSH57	757168	543256	441123	81	16.8
MSD86	525440	763164	335792	44	12.1
MSD85	240364	152352	130260	85	5.4
MSD101	430617	627117	335507	53	10.7
MSD124	314448	663498	202366	30	7.2
MSD132A	755792	984133	518638	53	17.8
MSD132B	692065	759438	497431	65	16.7
G27 MA	735874				17.4
G27 DB1	495522				11.7
G27 S1175	230534				5.4
G27 S2175	470018				11.2
G27 S5175	354345				8.7
Total reads aligned	**8514709**				
% Mapped reads aligned	**56**				

Table 2.3 Mutations found in sequenced strains.

	Total	**In Genes**		**Intergenic**
		Nonsynonymous	Synonymous	
Errors in Reference Sequence	41	27 (in 24 genes)	-	14
Verified Mutations	81	37 (in 17 genes)	44 (in 4 genes)	-
High Confidence Mutations	57	41 (in 38 genes)	4 (in 4 genes)	12
Low Confidence Mutations	103	70 (in 69 genes)	26 (in 25 genes)	7
Total	**282**	**173**	**74**	**33**

Table 2.4 All single base variants identified in sequenced strains using MAQ, BWA/SAMtools, and *breseq*.

					Strain																		
					Reference		Mouse-Adapted										In vitro-adapted				TC-adapted		
Gene Name	Strand	Position	Reference Base	New Base	G27_Original	G27_Working	G27_KO	G27_15-1.2	G27_20-2.1	NSH57	MSD86	MSD85	MSD101	MSD124	MSD132A	MSD132B	G27_DB1	G27_S1175	G27_S5175	G27_S2175	G27_MA	Amino acid change	Gene function
Errors in Reference																							
ureF	neg	74846	*	+A	+		+	+	+	+	+				+	+	+		+	+	+	fixes "frameshift" in reference	Urease accessory protein
HPG27_137	pos	156870	*	+T			+	+	+	+			+	+	+	+			+	+	+	fixes "frameshift" in reference	hypothetical protein
intergenic		204054	*	+G	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	intergenic between 185/186	185= 3-acyl-oxyacyl carrier protein synthase, 186= hypothetical protein
intergenic		232310	*	+A			+	+		+	+	+		+	+	+	+				+	intergenic between 207/208	207= OMP hopM, 208=putative sulfate permease
HPG27_251	pos	277049	*	+T	+		+	+	+	+	+				+	+			+	+	+	changes alternate start codon	hypothetical protein
HPG27_252	pos	277872	*	+A	+	+	+	+	+	+	+			+	+	+	+	+		+	+	adds 43AA to C terminus	hypothetical protein
intergenic		334159	*	-T	+	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+	intergenic 5bp up from 301	301= hypothetical protein
HPG27_371	pos	401117	*	+G	+	+	+		+	+	+	+	+	+	+	+	+	+	+	+	+	fixes "frameshift" in reference	hypothetical protein
HPG27_392	pos	423288	T	G			+	+		+	+			+	+		+		+	+		S50R	7,8-dihydro-6-hydroxy methylpterin-pyrophosphokinase (folK)
HPG27_403	neg	431781	*	+T	+		+	+	+	+	+				+	+	+		+	+	+	adds 15AA to C terminus	putative heat shock protein
HPG27_405	pos	433755	*	+G	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	fixes "frameshift" in reference	hypothetical protein
HPG27_412	neg	441037	*	+T	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	adds 49AA to C terminus	phosphatidyl glycerophosphate synthase
HPG27_459	pos	494157	*	+G						+		+		+	+							fixes "frameshift" in reference	phospholipase A1
intergenic		507732	*	+T	+	+	+	+	+	+	+	+			+			+	+	+	+	intergenic between glnA/472	glnA=glutamine synthetase, 472=hypothetical protein
intergenic		508421	*	+C	+		+	+	+	+	+		+	+	+	+	+			+	+	intergenic between 472/rplI	472= hypothetical protein, rplI= 50S ribosomal protein L9
cagA	pos	551897	*	+T	+	+	+	+	+	+	+		+	+	+	+	+			+	+	removes 14 AA from C terminus	cytotoxin-associated protein A
HPG27_512	pos	555760	*	+A		+	+			+	+		+	+	+	+	+	+		+	+	fixes "frameshift" in reference	hypothetical protein
intergenic		566958	*	+T	+				+	+	+	+	+		+	+	+		+		+	intergenic between dapF/526	dapF= Diaminopimelate epimerase, 526= membrane protein
HPG27_541	neg	581541	*	-T	+		+	+		+				+	+	+		+			+	still frameshifted	putative siderophore-mediated iron transport protein
HPG27_541	neg	581570	*	-T					+		+				+		+				+	still frameshifted	putative siderophore-mediated iron transport protein
HPG27_589	pos	643278	*	-T			+		+	+					+		+					L34stop, F32L, D33I	hypothetical protein
HPG27_662	neg	725944	*	+A	+		+	+	+	+	+	+	+		+	+	+			+	+	fixes "frameshift" in reference	excinuclease ABC subunit A
dnaX	neg	739196	*	-G	+	+	+	+	+	+	+		+	+	+	+	+		+	+	+	fixes "frameshift" in reference	DNA polymerase III gamma and tau subunits
hopL	pos	1208733	*	-A	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	fixes "frameshift" in reference	outer membrane protein
intergenic		1228406	*	-A	+	+	+	+	+	+	+	+	+	+	+	+		+		+	+	intergenic between 1119/hopQ	1119=hypothetical protein, hopQ= OMP
intergenic		1228626	*	+T			+	+		+			+	+	+		+					intergenic between 1119/hopQ	1119=hypothetical protein, hopQ= OMP
HPG27_1143	neg	1265037	*	+A	+	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+	1143 fuses to 1142	DNA-directed RNA polymerase, beta subunit
intergenic		1304662	*	+T	+	+	+	+	+	+	+	+	+		+	+	+		+	+	+	intergenic 11 bp up from alaS	alaS= alanyl-tRNA synthetase
HPG27_1197	neg	1316093	*	-A		+	+	+	+	+	+		+	+	+	+	+			+		fixes "frameshift" in reference	oligopeptide ABC transporter
nuoM	pos	1333420	*	-A		+	+	+	+	+	+		+	+	+	+	+	+	+	+	+	fixes "frameshift" in reference	NADH-ubiquinone oxidoreductase chain M (respiratory chain)
nuoM	pos	1333463	*	-C	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	fixes "frameshift" in reference	NADH-ubiquinone oxidoreductase chain M (respiratory chain)
nuoM	pos	1333524	*	-T	+	+	+	+	+	+	+	+	+		+	+	+	+	+	+	+	fixes "frameshift" in reference	NADH-ubiquinone oxidoreductase chain M (respiratory chain)
rpoA	neg	1359849	*	+T	+		+			+	+		+		+	+	+		+	+	+	adds 8AA to C terminus	DNA-directed RNA polymerase subunit alpha
intergenic		1367919	*	-T	+	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+	intergenic between rplN/rplX	rplN= 50S ribosomoal protein L14 , rplX= 50S ribosomal protein L24
ubiA	neg	1410928	*	-A	+		+	+	+		+		+		+	+	+	+	+	+	+	adds 6AA to C terminus	UbiA prenyltransferase
intergenic		1428286	*	+A	+		+	+	+	+	+	+	+	+	+	+	+		+	+	+	intergenic between 1322/comL	1322= flagellar assembly protein fliW, comL=competence lipoprotein
HPG27_1369	pos	1480681	*	+T	+		+	+	+	+	+			+	+	+	+	+	+	+	+	fixes "frameshift" in reference	ribonuclease P, protein component
intergenic		1491244	*	+A		+	+	+	+	+	+	+	+	+	+	+	+		+	+	+	intergenic between 1380/1381	1380=hypothetical protein, 1381=thioredoxin
intergenic		1491283	*	-A		+	+										+					intergenic between 1380/1381	1380=hypothetical protein, 1381=thioredoxin
HPG27_1427	neg	1541768	*	+A	+		+	+	+	+	+		+		+	+	+	+	+	+	+	fixes "frameshift" in reference	putative cation transporting P-type ATPase
intergenic		1568738	*	+A	+		+	+	+	+	+	+	+		+	+		+		+	+	intergenic between 1446/1447	1446= hypothetical protein, 1447= hypothetical protein
Verified Changes																							
HPG27_230	neg	254865	*	+T				+	+	+	+	+	+	+	+	+						K231stop +9 AA substitutions	oligopeptide permease ATPase protein, oppD
flgH	pos	338514	G	T		+	+										+	+	+	+	+	G178W	Flagellar basal-body L-ring protein
fliM	pos	427091	C	T				+	+	+	+	+	+	+	+	+						R34C	flagellar motor switch
cagY	neg	524524	C	T						+	+		+	+	+	+						L656L (syn)	cagA secretion system "sheath", essential for cagA translocation
cagY	neg	525132	*	+T								+	+		+	+						K462stop + 7 AA substitutions	cagA secretion system "sheath", essential for cagA translocation
HPG27_557	neg	599023	C	T						+	+		+	+	+	+						V623I	penicillin-binding protein 1A
hopH	pos	652935	*	-CT									+		+	+						- 1 CT repeat, leading to frameshift	Outer membrane protein
HPG27_676	pos	741019	C	T							+	+	+	+	+	+						Q68stop	periplasmic zinc transporter (znuA)
HPG27_792	neg	856793	*	+A						+	+	+	+	+	+	+						K78stop + 16 AA substitutions	hypothetical protein
hopC	pos	932492	G	A							+		+		+	+						G470R	porin, outer membrane protein
HPG27_910	neg	989061 to 989515	HPG27_910 sequence	jhp0897 (J99) sequence																	+	AA 36 to 187 replaced with corresponding AA sequence from strain J99	Putative uncharacterized protein
HPG27_1024	neg	1128112	C	A							+											G452V	Outer membrane protein (homC)
hofH	pos	1218753 to 1219168	hofH sequence	jhp1094 (J99) sequence																	+	AA 358 to 471 (C-terminus) of hofH replaced with corresponding AA sequence from strain J99	Outer membrane protein

					Strain																			
					Reference		Mouse-Adapted									In vitro-adapted				TC-adapted				
Gene Name	Strand	Position	Reference Base	New Base	G27_Original	G27_Working	G27_KO	G27_15-1.2	G27_20-2.1	NSH57	MSD86	MSD85	MSD101	MSD124	MSD132A	MSD132B	G27_DB1	G27_S1175	G27_S5175	G27_S2175	G27_MA	Amino acid change	Gene function
HPG27_1112	neg	1220108 to 1220984	HPG27_1112 sequence	jhp1095 (J99) sequence																	+	AA 160 to 452 of HPG27_1112 replaced with corresponding AA sequence from strain J99	Carbon starvation protein
HPG27_1188	pos	1307990	*	-T						+	+				+							L31stop + 14 AA substitutions	hypothetical protein
fliI	neg	1454469	G	A											+							A413V	Flagellum-specific ATP synthase
dnaE	neg	1495661	G	A							+	+	+	+	+	+						R156C	DNA polymerase III subunit alpha
dnaA	neg	1572496	C	G			+															M269I	Chromosomal replication initiator protein
High Confidence Changes																							
intergenic		7141	T	G											+							intergenic between hopZ/groEL	hopZ= OMP, groEL= 60kDa chaperonin
intergenic		63484	A	C							+				+	+						intergenic between 52/53	52=hypothetical protein, 53= hypothetical protein
intergenic		63487	C	T							+					+						intergenic between 52/53	52=hypothetical protein, 53= hypothetical protein
intergenic		63536	A	G											+							intergenic between 52/53	52= hypothetical protein, 53= hypothetical protein
ureF	neg	75116	T	G					+													K5T	Urease accessory protein
rpoD	neg	93613	T	C														+				D254G	
flaB	neg	123005	*	-A																	+	L238stop + 3 AA substitutions	flagellin B, minor flagellar subunit
HPG27_119	pos	136609	*	+T					+													N25stop + 16 AA substitutions	hypothetical protein
mutY	neg	150182	G	R											+							P165S	A/G-specific adenine glycosylase involved in base-excision repair
intergenic		210636	*	-CT																	+	intergenic between rfaJ-2/hofA	rfaJ-2= Lipopolysaccharide 1,2-glycosyltransferase, hofA= OMP
HPG27_255	pos	280452	*	-A			+															I181stop	Indole-3-glycerol phosphate synthase, tryptophan biosynthesis
hbpA	pos	311275	*	-T																	+	L20stop + 5 AA substitutions	Periplasmic dipeptide-binding protein
narK	neg	327078	*	-G														+				L225stop + 24 AA substitutions	Nitrite extrusion protein
HPG27_418	neg	446963	T	C																+		S406G	Polyphosphate kinase
HPG27_418	neg	447443	G	C														+				Q246E	Polyphosphate kinase
HPG27_436	pos	470465	*	-A																	+	fuses 436 to 437	Type II adenine specific methyltransferase
HPG27_450	neg	485787	G	A																+		R135C	Putative potassium channel protein
cagS	neg	533008	C	G															+			D62H	cag PAI protein, cag13 homolog
HPG27_537	neg	577540	G	T												+						T282K	hypothetical protein
HPG27_553	pos	594823	G	A																	+	A200A (syn)	Type III R-M system modification enzyme
HPG27_571	neg	625306	C	T																+		A59T	ABC transporter, permease
HPG27_610	pos	663148	C	A							+											A188D	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
HPG27_613	neg	667159	G	A																	+	P64P (syn)	Alpha1,3-fucosyl transferase
HPG27_654	pos	717490	G	A																	+	A13T	hypothetical protein
sabB	pos	742455	G	A																	+	E287K	Outer membrane protein
sabA	neg	747826	G	A				+														Q107stop	OMP, binds sialyl-lewis X blood group antigen
HPG27_689	neg	758083	A	C														+				F85V	hypothetical protein. Predicted Ras-like GTPase superfamily domain.
intergenic		758373	*	+A			+															intergenic between 689/690	689= hypothetical protein, 690= hypothetical protein
HPG27_698	pos	766449	C	T															+			R202C	Ribose-phosphate pyrophosphokinase (nucleotide biosynthesis)
horG	neg	822670	A	T																+		V104D	Outer membrane protein
HPG27_806	neg	872455	G	A					+													C469C (syn)	Type I restriction enzyme M protein
HPG27_812	neg	879923	G	A															+			A283V	bifunctional ADP-heptose synthase, hldE
HPG27_812	neg	879924	C	T														+				A283T	bifunctional ADP-heptose synthase, hldE
intergenic		963717	A	G															+			intergenic between 885/886	885= proline and betaine transporter, 886= hypothetical protein
intergenic		995536	*	+AG													+			+		intergenic between 914/915	914= hypothetical protein, 915= hypothetical protein
HPG27_961	pos	1048525	T	G														+				S32A	hypothetical protein
HPG27_1047	pos	1152613	*	-A									+	+		+						L24stop + 11 AA substitutions	hypothetical protein
HPG27_1061	neg	1167133	C	A					+													E110D	cysteine-rich protein X
HPG27_1063	neg	1168693	*	+T															+			K253stop + 5 AA substitutions	gamma-glutamyl transpeptidase
HPG27_1063	neg	1169416	*	+A																	+	K33stop + 27 AA substitutions	gamma-glutamyl transpeptidase
hopL	pos	1207755	*	-A																	+	L524stop + 22 AA substitutions	outer membrane protein
intergenic		1234345	*	+A																+		intergenic between 1123/tRNA28	1123= pyrimidine nucleoside transport protein (nupC), tRNA28= tRNA Met
HPG27_1139	neg	1255384	G	A														+				G46G (syn)	fusA, Elongation factor G
HPG27_1142	neg	1262106	A	C																	+	L942V	DNA-directed RNA polymerase, beta subunit
HPG27_1171	pos	1290824	C	T				+														A24V	cytochrome c 553
HPG27_1197	neg	1316285	C	T																	+	W274stop	oligopeptide ABC transporter
pyrE	pos	1320606	G	A															+			R150H	Orotate phosphoribosyltransferase, pyrimidine biosynthesis
HPG27_1236	neg	1354619	*	+A																+		K14stop	hypothetical protein
HPG27_1314	pos	1418758	C	T																	+	R297C	Type III R-M system modification enzyme
intergenic		1491284	A	G					+													intergenic between 1380/1381	1380=hypothetical protein, 1381=thioredoxin
HPG27_1427	neg	1540731	C	T																	+	A768T	putative cation transporting P-type ATPase
HPG27_1431	pos	1546981	C	T							+											P274S	hypothetical protien

					Strain																		
					Reference		Mouse-Adapted									In vitro-adapted				TC-adapted			
Gene Name	Strand	Position	Reference Base	New Base	G27_Original	G27_Working	G27_KO	G27_15-1.2	G27_20-2.1	NSH57	MSD86	MSD85	MSD101	MSD124	MSD132A	MSD132B	G27_DB1	G27_S1175	G27_S5175	G27_S2175	G27_MA	Amino acid change	Gene function
HPG27_1442	neg	1563783	C	T							+											S196N (if ref not frameshifted)	type III R-M system restriction enzyme
HPG27_1444	neg	1564973	*	-C											+	+						fuses 1444 to 1443	type III R-M system modification enzyme
intergenic		1586573	*	-T			+															intergenic between 1459/rocF	1459= Iron (III) dicitrate transport protein (membrane protein), rocF= arginase
intergenic		1623495	G	A							+											intergenic between tsf/1494	tsf= elongation factor Ts (protein biosynthesis), 1494= cell division protein (ftsI)
HPG27_1518	neg	1644782	C	T																	+	G44E	Undecaprenylphosphate N-acetylglucosaminyl transferase (peptidoglycan)
Low Confidence Changes																							
HPG27_21	neg	21246	T	Y										+								I418V	lipid A phosphoethanolamine transferase
comB10-1	pos	39799	C	Y								+										Q183stop	Competence protein
HPG27_51	neg	61622	T	W						+												N291I	Proline/delta1-pyrroline-5-carboxylate dehydrogenase
HPG27_88	pos	100172	G	A		+																A88T	Putative D-2-hydroxyacid dehydrogenase, L-serine biosynthetic process
HPG27_89	neg	101496	G	R								+										S101S (syn)	hypothetical protein
thrS	pos	132073	C	T	+																	G265G (syn)	Threonyl-tRNA synthetase
eno	pos	160352	G	A		+																A96T	Enolase, carbohydrate degradation in glycolysis
rfaJ-1	neg	163867	T	W													+					E311V	Lipopolysaccharide 1,2-glycosyltransferase
HPG27_157	neg	175232	C	G								+										D197H	Molybdopterin biosynthesis protein moeA
lysS	pos	185904	T	C								+										I382I (syn)	Lysyl-tRNA synthetase
HPG27_188	neg	208235	C	Y														+				G46S	hypothetical protein
HPG27_198	pos	221937	G	R				+														L354L (syn)	1-deoxy-D-xylulose 5-phosphate reductoisomerase, dxr, isoprenoid biosynthesis
intergenic		236258	G	R													+					intergenic between 210/211	210= kdsB, 211= disulphide isomerase
HPG27_236	pos	261900	C	A		+																S127Y	Conserved hypothetical secreted protein
HPG27_236	pos	261970	C	T		+																P150P (syn)	Conserved hypothetical secreted protein
ftsH-1	pos	291126	A	R									+									K39E	Cell division protein, belongs to the AAA ATPase family
HPG27_267	pos	293453	G	R						+												V109I	hypothetical protein
HPG27_272	pos	306179	A	R										+								K249K (syn)	Para-amino benzoate synthetase
HPG27_286	pos	320842	T	C								+										W55R	hypothetical protein
babA	neg	329282	G	T								+										N684K	Adhesin, binds Lewis B blood group antigen
babA	neg	329855	A	R						+		+			+	+					+	T493T (syn)	Adhesin, binds Lewis B blood group antigen
babA	neg	329864	C	T								+										L490L (syn)	Adhesin, binds Lewis B blood group antigen
intergenic		331419	*	+T									+									intergenic between babA/299	babA= OMP,, 299= hypothetical protein
intergenic		348646	G	R																	+	intergenic between 320/321	320= hypothetical protein, 321= hypothetical protein
fliH	pos	359341	A	W							+											D150V	Flagellar assembly protein
HPG27_342	neg	375465	C	M								+										S166I	Transketolase A tktA
HPG27_404	neg	432215	T	Y							+											K255K (syn)	Co-chaperone-curved DNA binding protein A (cpbA)
HPG27_414	neg	442522	T	W								+										I98F	7 alpha-hydroxy steroid dehydrogenase
HPG27_428	neg	460838	A	R							+											I224I (syn)	Oligoendopeptidase F (pepF)
HPG27_450	neg	485422	A	M																	+	D256E	Putative potassium channel protein
mraY	pos	488013	G	A								+										S151N	Phospho-N-acetylmuramoyl-pentapeptide-transferase,peptidoglycan biosynthesis
gyrB	pos	497563	C	Y						+												L610F	DNA gyrase subunit B
rplI	pos	508620	C	Y		+																G13G (syn)	50S ribosomal protein L9, binds to the 23S rRNA
cagZ	neg	520238	A	R										+								L136S	cag PAI protein, cag6 homolog
cagY	neg	525861	*	-T		+																V217stop + 7 AA substitutions	cagA secretion system "sheath", essential for cagA translocation
cagX	neg	528007	C	Y			+															S22N	virB9 homolog, part of T4SS outer membrane pore
cagU	pos	531448	G	R							+											G187S	cag PAI protein, cag11 homolong
cagM	pos	537316	G	K							+											D73Y	cag PAI protein, cag16 homolog
rpmE	pos	555114	A	T								+										K29M	50S ribosomal protein L31, binds 23S rRNA
HPG27_553	pos	595048	C	Y						+												G275G (syn)	Type III R-M system modification enzyme
HPG27_573	neg	625779	T	G	+																	K220T	ABC transporter. ATP-binding protein
HPG27_613	neg	666303	G	C	+																	P350A	Alpha1,3-fucosyl transferase, protein glycoslyation, membrane protein
gatB	pos	675835	T	C								+										Y240Y	Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B
HPG27_627	pos	681673	C	T	+																	S21S (syn)	Oxygen-independent coproporphyrinogen III oxidase, porphyrin biosynthesis
HPG27_627	pos	682468	G	A										+								M286I	Oxygen-independent coproporphyrinogen III oxidase, porphyrin biosynthesis
HPG27_659	pos	722262	T	K														+				Y70D	Transcriptional activator of flagella proteins
HPG27_668	pos	734470	G	R							+											G371D	hypothetical protein
HPG27_681	neg	748979	G	K								+										H84N	Outer membrane protein
HPG27_729	neg	793884	C	M							+											G216V	N-acetylmuramoyl-L-alanine amidase (amiA), peptidoglycan
HPG27_742	pos	810712	G	R								+										A654T	secA, protein translocase subunit
HPG27_747	neg	817326	C	G								+										V223V (syn)	Cadmium-transporting ATPase cadA
horG	neg	822271	T	K								+										K237T	outer membrane protein
HPG27_760	neg	827367	C	Y									+									G313R	DHBP synthase, involved in riboflavin biosynthesis
HPG27_776	neg	840484	A	G		+																C229C (syn)	DNA methylase (on Cs)

					Strain																			
					Reference		Mouse-Adapted										In vitro-adapted				TC-adapted			
Gene Name	Strand	Position	Reference Base	New Base	G27_Original	G27_Working	G27_KO	G27_15-1.2	G27_20-2.1	NSH57	MSD86	MSD85	MSD101	MSD124	MSD132A	MSD132B	G27_DB1	G27_S1175	G27_S5175	G27_S2175	G27_MA	Amino acid change	Gene function
uvrC	pos	846045	G	R							+											K558K (syn)	UvrABC system protein C, DNA damage repair
engA	pos	858144	A	W								+										E364D	GTP-binding protein engA
HPG27_794	pos	858791	A	R														+				T72A	DNA-binding protein HU
HPG27_804	neg	870763	C	Y		+																S23N	Type I restriction enzyme R protein
HPG27_817	neg	884053	A	R										+								V434A	hypothetical protein
HPG27_830	pos	896070	T	Y						+												V48A	Iron-regulated outer membrane protein (frpB-1)
intergenic		900369	C	Y							+											intergenic between 834/835	834= hypothetical protein, 835= hypothetical protein
HPG27_857	pos	922650	C	Y			+															L99L	hypothetical protein
HPG27_902	neg	981014	C	Y				+				+										L188L (syn)	Oxygen-insensitive NADPH nitroreductase (rdxA)
HPG27_903	neg	982041	G	A										+								R130C	Prolipoprotein diacylglyceryl transferase
HPG27_910	neg	988905	G	R	+																	A239V	Putative uncharacterized protein
HPG27_911	neg	989748	C	S							+											E144Q	hypothetical protein
HPG27_941	neg	1019627	C	T								+										G720E	hypothetical protein
HPG27_941	neg	1020005	C	T								+										R594K	hypothetical protein
HPG27_974	pos	1061233	A	R													+					Y253C	DNA transfer protein (VirD4-like)
HPG27_980	neg	1075999	C	T								+										K64K (syn)	Adenine specific DNA methyltransferase
HPG27_987	neg	1086299	A	T								+										V134E	Putative neuraminyllactose-binding hemagglutinin-like protein hpaA
HPG27_1010	neg	1113134	A	R															+			L475L (syn)	Primosomal protein replication factor (priA), involved in DNA replication
HPG27_1017	pos	1120019	G	R								+										A375T	Glutamate dehydrogenase
HPG27_1018	neg	1121439	T	K																	+	K54T	Alpha1,3-fucosyl transferase, protein glycoslyation, membrane protein
truA	pos	1138886	G	A										+								P84P (syn)	tRNA pseudouridine synthase A
HPG27_1067	neg	1173104	C	Y					+													G65E	flgM, anti sigma factor for fliA
murG	pos	1203031	C	Y							+											A17V	N-acetylglucosaminyl transferase, PG synthesis
hisS	neg	1249221	G	R									+									Q256stop	Histidyl-tRNA synthetase
HPG27_1161	neg	1280976	G	R															+			T230T (syn)	Organic solvent tolerance protein
HPG27_1161	neg	1281393	C	M														+				L91F	Organic solvent tolerance protein
HPG27_1166	neg	1285352	G	R		+																A899V	D-lactate dehydrogenase (dld)
HPG27_1192	neg	1310200	C	Y							+											G249R	DNA polymerase III holoenzyme delta subunit
aroE	neg	1313417	G	R														+				C86C (syn)	Shikimate 5-dehydrogenase, chorismate (precursor of AA) biosynthesis
secG	pos	1319477	T	K						+												S158R	preprotein translocase subunit secG
HPG27_1204	pos	1321340	A	G								+										K43E	NAD-dependent deacetylase
HPG27_1224	neg	1342796	A	R								+										I79T	Tryptophan synthase, beta subunit
intergenic		1350561	C	Y				+														intergenic between 1231/1232	1231= Type II adenine specific DNA methyltransferase, 1232= hypothetical protei
HPG27_1235	neg	1352535	G	A								+										Y329Y (syn)	hypothetical protein
intergenic		1356642	A	C								+										intergenic between 1238/1239	1238= Conserved hyp. secreted protein, 1239= Putative transcriptional regulator
fumC	neg	1376775	G	R					+													S269S (syn)	fumerase (fumerate hydratase), TCA cycle
HPG27_1316	pos	1422974	C	Y														+				P878L	Type III restriction enzyme R protein
mreB	neg	1424106	G	T		+																A327A (syn)	Rod shape-determining protein
intergenic		1453301	T	A		+																intergenic between 1340/murB	1340= hypothetical, murB= UDP-N-acetylenolpyruvoyl glucosamine reductase
fliI	neg	1454469	G	R												+						A413V	Flagellum-specific ATP synthase
HPG27_1348	neg	1460487	C	M								+										G64V	hypothetical protein
HPG27_1388	neg	1499744	C	M										+								D212Y	ABC transporter ATP-binding protein
HPG27_1391	neg	1502907	G	S								+										P154A	Branched-chain-amino-acid aminotransferase
polA	neg	1505520	G	A		+																R493C	DNA polymerase I
HPG27_1416	pos	1530416	G	R								+										G99R	hypothetical protein
HPG27_1474	pos	1600089	T	G								+										I123M	Putative type II methylase protein
HPG27_1513	pos	1640904	A	R							+											I286V	metN. DL-methionine transporter, ATP binding subunit
HPG27_1518	neg	1644235	A	G								+										N226N (syn)	Undecaprenylphosphate N-acetylglucosaminyl transferase (peptidoglycan)
HPG27_1518	neg	1644852	*	-T	+																	L53stop + 28 AA substitutions	Undecaprenylphosphate N-acetylglucosaminyl transferase (peptidoglycan)

Table notes: Errors in Reference are defined as mutations found in multiple gene classes that are either verified or high confidence; Verified Changes have been validated by Sanger sequencing in at least one strain; High Confidence Changes are supported by a minimum of 3 reads with unambiguous base calling and high quality scores, and are consistent with strain relationships; Low Confidence Changes may have conflicting or ambiguous base calls, low quality scores, or be inconsistent with strain relationships (i.e., found in a parental strain but not its derivatives). + indicates Illumina reads support a mutation in the indicated strain; yellow shading indicates a mutation confirmed present and grey shading a mutation confirmed absent by resequencing in the indicated strain. Y: C or T; W: A or T; R: A or G; M: A or C; K: G or T; S: G or C.

Table 2.5 Copy number variation and recombination.

Gene(s) Involved	Approximate Region Affected	Type of Change	Strains Affected	Confidence
52/53	63 bp, intergenic	Deletion[a]	MSD85	High: supported by both 454 and Illumina data.
209/210	~100 bp, intergenic	Deletion[b]	G27 working	Moderate: good coverage flanking
294 (narK)	~500 bp, 5' end	Deletion[b]	S5175	High: good coverage flanking
342	~75 bp, central	Deletion[b]	MSD85	Low: poor coverage flanking
407	~50 bp, 3' end	Deletion[b]	G27 working	Low: poor coverage flanking
486 (cagY)	~100 bp, central	Deletion[b]	15-1.2, 20-2.1, NSH57, MSD85, MSD86, MSD101, MSD124, MSD132A, MSD132B	Low: poor mapping of reads in repeat regions
613	~500 bp, near 5' end	Deletion[b]	G27 MA	Moderate: good coverage flanking but poor read mapping
677 (sabB)	~2 kb, whole gene	Deletion[b]	G27 DB1, S1175, S2175, S5175	High: very good coverage flanking
958/959	~50 bp, intergenic	Deletion[b]	S5175	Moderate: good coverage flanking
1061	~100 bp, central	Deletion[b]	20-2.1	Moderate: good coverage flanking
1267	~50 bp, near 5' end	Deletion[b]	S1175	Low: poor coverage flanking
1479	~50 bp, central	Deletion[b]	S1175	Moderate: good coverage flanking
88 + panB	100463/400051	New Junction[c]	NSH57	Untested
intergenic (130/131) + cagH	152243/541360	New Junction[c]	G27 KO	Untested
ftsK + 1327	370277/1435396	New Junction[c]	G27 Original	Untested
infB + 896	409539/977519	New Junction[c]	15-1.2	Untested
gyrB + 651	495915/713354	New Junction[c]	KO, 15-1.2	Tested and not confirmed by PCR
547 + 695	587483/763710	New Junction[c]	KO	Untested
590 + aspA	645248/665018	New Junction[c]	KO	Untested
668 + 1423	733888/1536865	New Junction[c]	KO	Untested

706 + intergenic (gpsA/910)	773479/987972	New Junction[c]	KO	Untested
hofE + 1517	806725/1643540	New Junction[c]	15-1.2	Tested and not confirmed by PCR
carB + flgC	941813/1626404	New Junction[c]	20-2.1, MSD85, MSD132A	Tested and not confirmed by PCR
cheA + 1228	1109332/1346470	New Junction[c]	NSH57	Untested
intergenic (1102/1103) + 1163	1210736/1283463	New Junction[c]	MSD132A	Tested and not confirmed by PCR

[a] Identified from 454 contig assembly.
[b] Identified by comparison of normalized read counts for each strain to those for G27 Original.
[c] Identified from *breseq* as reads with ends mapping to two distinct sites.

Table 2.6 Oligonucleotides used in this chapter.

Oligo ID	Oligo Name	Oligo Sequence	Targeted Gene (if applicable)	Purpose of Oligo
BCHI-1	BCHI-1	TACACGACGCTCTTCCGATCTTACGAAGTCATC*T		Barcoded adaptor; used for NSH57
BCHI-2	BCHI-2	TACACGACGCTCTTCCGATCTGACGAGATTATC*T		Barcoded adaptor; used for G27 15-1.2
BCHI-3	BCHI-3	TACACGACGCTCTTCCGATCTACCGTAAGAATC*T		Barcoded adaptor; used for G27 20-2.1
BCHI-4	BCHI-4	TACACGACGCTCTTCCGATCTTAGTGGCAAATC*T		Barcoded adaptor; used for G27 KO
BCHI-5	BCHI-5	TACACGACGCTCTTCCGATCTCATTAACGCATC*T		Barcoded adaptor; used for G27 MA
BCHI-6	BCHI-6	TACACGACGCTCTTCCGATCTTCGTTGAAGATC*T		Barcoded adaptor; used for G27 original
BCHI-7	BCHI-7	TACACGACGCTCTTCCGATCTTAGTACGCTATC*T		Barcoded adaptor; used for G27 working
BCHI-8	BCHI-8	TACACGACGCTCTTCCGATCTCTCAGATCAATC*T		Barcoded adaptor; used for MSD85
BCHI-13	BCHI-13	TACACGACGCTCTTCCGATCTGGATGTTCTATC*T		Barcoded adaptor; used for MSD86
BCHI-14	BCHI-14	TACACGACGCTCTTCCGATCTCTTATCCAGATC*T		Barcoded adaptor; used for MSD101
BCHI-15	BCHI-15	TACACGACGCTCTTCCGATCTGTAAGTCACATC*T		Barcoded adaptor; used for MSD124
BCHI-16	BCHI-16	TACACGACGCTCTTCCGATCTTTCAGTGAGATC*T		Barcoded adaptor; used for MSD132A
BCHI-17	BCHI-17	TACACGACGCTCTTCCGATCTCTCGTAATGATC*T		Barcoded adaptor; used for MSD132B
BCHI-18	BCHI-18	TACACGACGCTCTTCCGATCTCATGTCTCAATC*T		Barcoded adaptor; used for G27 DB1

BCHI-19	BCHI-19	TACACGACGCTCTTCCGATCTAATCGTGGAATC*T	Barcoded adaptor; used for G27 S2175
BCHI-20	BCHI-20	TACACGACGCTCTTCCGATCTGTATCAGTCATC*T	Barcoded adaptor; used for G27 S1175
BCHI-21	BCHI-21	TACACGACGCTCTTCCGATCTAGCAGATGTATC*T	Barcoded adaptor; used for G27 S5175
oIEC1	Nextera 1	CAAGCAGAAGACGGCATACGAGATTACGAAGTCC GGTCTGCCTTGCCAGCCCGCTCAG	Barcoded adaptor for Nextera sequencing library prep; used for NSH57
oIEC2	Nextera 2	CAAGCAGAAGACGGCATACGAGATGACGAGATTC GGTCTGCCTTGCCAGCCCGCTCAG	Barcoded adaptor for Nextera sequencing library prep; used for G27 15-1.2
oIEC3	Nextera 3	CAAGCAGAAGACGGCATACGAGATACCGTAAGAC GGTCTGCCTTGCCAGCCCGCTCAG	Barcoded adaptor for Nextera sequencing library prep; used for G27 20-2.1
oIEC4	Nextera 4	CAAGCAGAAGACGGCATACGAGATTAGTGGCAAC GGTCTGCCTTGCCAGCCCGCTCAG	Barcoded adaptor for Nextera sequencing library prep; used for G27 KO
oIEC5	Nextera 5	CAAGCAGAAGACGGCATACGAGATCATTAACGCC GGTCTGCCTTGCCAGCCCGCTCAG	Barcoded adaptor for Nextera sequencing library prep; used for G27 MA
oIEC6	Nextera 6	CAAGCAGAAGACGGCATACGAGATTCGTTGAAGC GGTCTGCCTTGCCAGCCCGCTCAG	Barcoded adaptor for Nextera sequencing library prep; used for G27 original
oIEC7	Nextera 7	CAAGCAGAAGACGGCATACGAGATTAGTACGCTC GGTCTGCCTTGCCAGCCCGCTCAG	Barcoded adaptor for Nextera sequencing library prep; used for G27 working
oIEC8	Nextera 8	CAAGCAGAAGACGGCATACGAGATCTCAGATCAC GGTCTGCCTTGCCAGCCCGCTCAG	Barcoded adaptor for Nextera sequencing library prep; used for MSD85
oIEC9	Nextera 9	CAAGCAGAAGACGGCATACGAGATTTCACCGTAC GGTCTGCCTTGCCAGCCCGCTCAG	Barcoded adaptor for Nextera sequencing library prep; used for MSD86

oIEC10	Nextera 10	CAAGCAGAAGACGGCATACGAGATGTCATGCATC GGTCTGCCTTGCCAGCCCGCTCAG		Barcoded adaptor for Nextera sequencing library prep; used for MSD101
oIEC11	Nextera 11	CAAGCAGAAGACGGCATACGAGATAGGACAGTTC GGTCTGCCTTGCCAGCCCGCTCAG		Barcoded adaptor for Nextera sequencing library prep; used for MSD124
oIEC12	Nextera 12	CAAGCAGAAGACGGCATACGAGATATGGTGTCTC GGTCTGCCTTGCCAGCCCGCTCAG		Barcoded adaptor for Nextera sequencing library prep; used for MSD132A
oIEC13	Nextera 13	CAAGCAGAAGACGGCATACGAGATGGATGTTCTC GGTCTGCCTTGCCAGCCCGCTCAG		Barcoded adaptor for Nextera sequencing library prep; used for MSD132B
oIEC14	Nextera 14	CAAGCAGAAGACGGCATACGAGATCTTATCCAGC GGTCTGCCTTGCCAGCCCGCTCAG		Barcoded adaptor for Nextera sequencing library prep; used for G27 DB1
oIEC15	Nextera 15	CAAGCAGAAGACGGCATACGAGATGTAAGTCACC GGTCTGCCTTGCCAGCCCGCTCAG		Barcoded adaptor for Nextera sequencing library prep; used for S2175
oIEC16	Nextera 16	CAAGCAGAAGACGGCATACGAGATTTCAGTGAGC GGTCTGCCTTGCCAGCCCGCTCAG		Barcoded adaptor for Nextera sequencing library prep; used for G27 S1175
oIEC17	Nextera 17	CAAGCAGAAGACGGCATACGAGATCTCGTAATGC GGTCTGCCTTGCCAGCCCGCTCAG		Barcoded adaptor for Nextera sequencing library prep; used for G27 S5175
oIEC25	flgHFor	GGCGTGTGTTAGCAAGCTTGTGC	flgH	Resequencing primers for SNPs/indels
oIEC26	fliMFor	GGCATTACTGAATCGCGCATTTCTCA	fliM	Resequencing primers for SNPs/indels
oIEC27	557For	TAGTGCGCGCGCGTGGTTTT	G27_557	Resequencing primers for SNPs/indels
oIEC28	hopCFor	AGGCTAATCCCCCTATGAAATCCCG	hopC	Resequencing primers for SNPs/indels
oIEC30	676For	AAGGATTGCGCCCCCACAGC	G27_676	Resequencing primers for

				SNPs/indels
oIEC31	dnaEFor	GCCGGCGCTGGCAAAAATCC	dnaE	Resequencing primers for SNPs/indels
oIEC32	babAFor	AGCGAAAGCGGTGGGGATTGG	babA	Resequencing primers for SNPs/indels
oIEC33	976For	AGTGCGTGTTTTAGCCACCTGT	G27_976	Resequencing primers for SNPs/indels
oIEC34	flgHRev	GCGATCGCTCTCATGCTAGGCT	flgH	Resequencing primers for SNPs/indels
oIEC35	fliMRev	CCCACGCTTGGAGCCTTACCC	fliM	Resequencing primers for SNPs/indels
oIEC36	557Rev	GCTTGCGAGCTATGGGGCGT	G27_557	Resequencing primers for SNPs/indels
oIEC37	hopCRev	TCAGCGCCCACTGGATTGAAGC	hopC	Resequencing primers for SNPs/indels
oIEC39	676Rev	ACGCTCAAAGCGTTTAGGATTCTGC	G27_676	Resequencing primers for SNPs/indels
oIEC40	dnaERev	GCATGAGCGCAAGAAAAGGCCA	dnaE	Resequencing primers for SNPs/indels
oIEC41	babARev	GGGGCAAACGCCCCAAAAGAGT	babA	Resequencing primers for SNPs/indels
oIEC42	976Rev	AGAGTTGCACCCAATCGCAAGC	G27_976	Resequencing primers for SNPs/indels
oIEC43	flgHSeq	CGAATTTTAAAGGCGGTGGCTCGC	flgH	Resequencing primers for SNPs/indels
oIEC44	fliMSeq	AATCCTTAGGAGTGGATCATGGCTGA	fliM	Resequencing primers for SNPs/indels
oIEC45	557Seq	AGCGGTGAGTATCCTTGTGGCT	G27_557	Resequencing primers for SNPs/indels
oIEC46	hopCSeq	CGGAGCGAGCGTGGGCTTTA	hopC	Resequencing primers for SNPs/indels
oIEC48	676Seq	TGAGCGTTTGCAGCGGCGAT	G27_676	Resequencing primers for SNPs/indels

oIEC49	dnaESeq	AGTGGGTGTCGTTGGTGGCA	dnaE	Resequencing primers for SNPs/indels
oIEC50	babASeq	AGGGGTTACGCCCGAGTTCT	babA	Resequencing primers for SNPs/indels
oIEC51	976Seq	ACAGCCGTTACACAAGACACGCA	G27_976	Resequencing primers for SNPs/indels
oIEC54	cagAFor	GGCGTTACGAATGGCGTTTCCC	cagA	Resequencing primers for SNPs/indels
oIEC55	hopHFor	TACCAGCTAGGGGGCGCGAA	hopH	Resequencing primers for SNPs/indels
oIEC56	910For	CCGCAAAACTCGGACCCGCT	G27_910	Resequencing primers for SNPs/indels
oIEC57	1024For	AAGCTAGTGGGCGCGCGATT	G27_1024	Resequencing primers for SNPs/indels
oIEC58	hofHFor	ATGGGGCGTGGAGCTTGGGA	hofH	Resequencing primers for SNPs/indels
oIEC59	1112For	GTTTCAGCCAGCGGCGGGTA	G27_1112	Resequencing primers for SNPs/indels
oIEC60	nuoMFor	ATCCCGTCCCGCTCGCTCTT	nuoM	Resequencing primers for SNPs/indels
oIEC61	fliIFor	TTGGCCGCTCCCCCTACCTC	fliI	Resequencing primers for SNPs/indels
oIEC62	1276For	CCCGGCATGCCAGCGATGAA	G27_1276	Resequencing primers for SNPs/indels
oIEC63	1407For	AGCACCGGGGCTGAGACGAT	G27_1407	Resequencing primers for SNPs/indels
oIEC64	dnaAFor	GCATTTGGTGCGCGTGTGCT	dnaA	Resequencing primers for SNPs/indels
oIEC66	cagARev	TCATGCGAGCGGCGATGTGA	cagA	Resequencing primers for SNPs/indels
oIEC67	hopHRev	GCGTCTAGCGTTCTGCCGGT	hopH	Resequencing primers for SNPs/indels
oIEC68	910Rev	GCGAGATAGGAGCAAAGACGGGC	G27_910	Resequencing primers for

				SNPs/indels
oIEC69	1024Rev	ACAGCCACCCCACAAAGCGT	G27_1024	Resequencing primers for SNPs/indels
oIEC70	hofHRev	TGGATGCCGGCACACGAACC	hofH	Resequencing primers for SNPs/indels
oIEC71	1112Rev	CGATAGCAGAGCCACGCCCG	G27_1112	Resequencing primers for SNPs/indels
oIEC72	nuoMRev	ACTGCTGGCGCGTTCTTCCA	nuoM	Resequencing primers for SNPs/indels
oIEC73	fliIRev	GCGATGGCGCTGAATGCGTG	fliI	Resequencing primers for SNPs/indels
oIEC74	1276Rev	GCACGACCACTTGAGCGGGG	G27_1276	Resequencing primers for SNPs/indels
oIEC75	1407Rev	GCACGCTTTTAGGGGTGGGGG	G27_1407	Resequencing primers for SNPs/indels
oIEC76	dnaARev	GGCTAAACACGGCGCGTTGC	dnaA	Resequencing primers for SNPs/indels
oIEC78	cagASeq	CCCTGAGTGGCTCAAGCTCGT	cagA	Resequencing primers for SNPs/indels
oIEC79	hopHSeq	GGCACATTCGCCCCACAAGC	hopH	Resequencing primers for SNPs/indels
oIEC80	910Seq	TGTGCGGCTATCAGCCCTACTCA	G27_910	Resequencing primers for SNPs/indels
oIEC81	1024Seq	TTGCGCATTCAACCCAGCGT	G27_1024	Resequencing primers for SNPs/indels
oIEC82	hofHSeq	GCGGCACAGGCTTTCGCTCT	hofH	Resequencing primers for SNPs/indels
oIEC83	1112Seq1	AACTGCCGGCAAGGCGCTTA	G27_1112	Resequencing primers for SNPs/indels
oIEC84	1112Seq2	TCCCTGCACACACCAACGCC	G27_1112	Resequencing primers for SNPs/indels
oIEC85	nuoMSeq	CGATCAAGCGAGCAGGGCGT	nuoM	Resequencing primers for SNPs/indels

oIEC86	fliISeq	GCCCCACCACTAAGCCCGC	fliI	Resequencing primers for SNPs/indels
oIEC87	1276Seq	TTCGCCCAAAGCCGGCACTA	G27_1276	Resequencing primers for SNPs/indels
oIEC88	1407Seq	GGGGTCTTTTTCCAACTCGCGCA	G27_1407	Resequencing primers for SNPs/indels
oIEC89	dnaASeq	CGCGCCTTCCATTTGGCGGA	dnaA	Resequencing primers for SNPs/indels
oIEC90	cagYFor2	TTTCTCGCTTCAGGGGTGAGCAAC	cagY	Resequencing primers for SNPs/indels
oIEC91	cagYRev2	ATCAAGCAAAGGCGCTAGAGACCC	cagY	Resequencing primers for SNPs/indels
oIEC92	cagYSeq2	GCGCAGTCTTTATAAGCTTTCAGGC	cagY	Resequencing primers for SNPs/indels
oIEC121	230_seq	GGCCCCACGCTTTGGCCTAA	G27_230	Resequencing primers for SNPs/indels
oIEC122	792_seq	GGATCGTTTTGGTATAGGGGCGGG	G27_792	Resequencing primers for SNPs/indels
oIEC241	primerA-1	TTGCGATACGATTAACATCACT	gyrB	putative new junction
oIEC242	primerA-2	TTTGTGCGCATTGATCCACA	G27_651	putative new junction
oIEC243	primerA-3	CGCTTCTGACCCTGATATG	G27_651	putative new junction
oIEC244	primerA-4	CCTGGAATGTATATTGGCGA	gyrB	putative new junction
oIEC245	primerA-5	TTGAAATTAGTGATGTTAATCGTATC	gyrB/651 putative junction	putative new junction
oIEC246	primerB-1	AGTAACAGCATTTATGGTTTGGT	hofE	putative new junction
oIEC247	primerB-2	CGATCACAGGAAGAGCC	G27_1517	putative new junction
oIEC248	primerB-3	AAGCAAACCCCCATGCG	G27_1517	putative new junction
oIEC249	primerB-4	CGAATTTAACTTTGGCTTTGGTT	hofE	putative new junction
oIEC250	primerB-5	TTGATGTTCTACCAAACCATAAAT	hofE/1517 putative junction	putative new junction
oIEC251	primerC-1	ATCTGGCATAATTCAAACACTTC	carB	putative new junction

oIEC252	primerC-2	GATTTCAACGAGATTTTAAATCAAAAAAT	flgC	putative new junction
oIEC253	primerC-3	TTTGATTTCTTCTAATATCGCTAAC	flgC	putative new junction
oIEC254	primerC-4	GTCCTCTTAATTTTTTGGCATC	carB	putative new junction
oIEC255	primerC-5	TGAGTTAGAAGTGTTTGAATTATG	carB/flgC putative junction	putative new junction
oIEC256	primerD-1	GAAAAAAATGCACCTCTAAAAATTAC	G27_1102	putative new junction
oIEC257	primerD-2	CCACAATTAAAACATTATAGTTATTGT	G27_1163	putative new junction
oIEC258	primerD-3	CACCATTACCCAAACAAAAATAC	G27_1163	putative new junction
oIEC259	primerD-4	CTAAAGGCGCAACGATTGAA	G27_1102	putative new junction
oIEC260	primerD-5	TATTTCCGTAATTTTTAGAGGTGC	1102/1163 putative junction	putative new junction

CHAPTER 3

Characterization of early stages of mouse adaptation

INTRODUCTION

Mouse-adapted strains of *H. pylori* capable of both colonizing and persisting for several months in wild-type C57BL/6 mice have been previously generated from the clinical isolate G27 by serial passaging (38, 59), see Fig. 1.2. Initial passaging took place in FvB/N mice, which are more permissive for *H. pylori* colonization, and involved inoculation of mice with the pooled bacteria recovered from the previous round to capture the diversity present (59). Further serial passaging was performed in C57BL/6 mice, using single colony isolates from mice with the highest loads to infect the next round of mice (38), and resulting in clonal isolates from each of these rounds of passaging. Genetic changes occurring at each round of serial passaging in mice were identified by whole genome sequencing and are described in Chapter 2.

To examine the phenotypic contribution of the changes identified by sequencing to mouse adaptation, defined isogenic mutants of *H. pylori* can be created using a two-step process. *H. pylori'*s natural competence for DNA transformation facilitates genetic manipulation (101). Double cross-over homologous recombination can be used to replace a gene of interest with a counter-selectable marker conferring chloramphenicol resistance and sucrose sensitivity, which can then be replaced with the desired allele of the gene, creating an unmarked, allele-switched strain (102). These isogenic mutants can then be compared to both their parental strain background and the adapted strain from which their mutant allele was derived for a variety of phenotypes, including infection of mice, induction of immune responses, and interactions with cultured cells.

This chapter includes the creation and characterization of isogenic mutants for the five genes identified in Chapter 2 as subject to mutation during the early stage of *H. pylori* murine adaptation, taking place between strains G27 and NSH57.

RESULTS

Mutation of *HPG27_792*, denoted *imc1*, can account for increased mouse colonization of NSH57 relative to G27

Because of the relative concentration of mutations during early stages of adaptation to infect mice, between the original strain G27 and the strain NSH57, and relatively fewer mutations per step at later stages of the adaptation process, we focused on characterization of the early mutations, particularly in the previously uncharacterized gene *HPG27_792*. On the basis of experiments described below, we have designated HPG27_792 as Imc1 (inhibitor of mouse colonization).

Isogenic mutant strains were used to determine the contribution of changes identified through genome sequencing to murine colonization potential. The adapted (NSH57) allele of a single gene or a defined combination of genes was introduced into the parental genetic background (G27) at the native locus by natural transformation and double-crossover homologous recombination using a counter-selectable marker system (Fig. 3.1).

To determine the relative ability of these strains to colonize the murine stomach, mice were infected by oral gavage with the parental strain (G27), the adapted strain (NSH57), or a pool of isogenic mutants containing equal proportions of four clones, each with the NSH57 allele of a single gene in the G27 background (indicated as *fliM**, *230**, *557**, or *imc1**) for one week. The mouse-adapted strain NSH57 colonized mice to levels that were slightly, but non-significantly ($p = 0.1575$), greater than those of the parental strain, G27. Infection with the pool

of isogenic mutants resulted in a greater proportion of mice with recoverable bacteria after one week (15 of 20 mice), compared to the parental strain G27 (12 of 20 mice), although average bacterial loads were not significantly higher ($p = 0.4820$). (Fig. 3.2 A)

An allele-specific PCR assay was used to genotype the colonies recovered from mouse stomachs after infection. The strain with the adapted allele of gene *imc1* (*imc1**), encoding a hypothetical protein, was recovered most frequently (53% of colonies) and was found in 11 of the 15 mice from which bacteria were recovered. Strains with the adapted alleles of two genes involved in cell wall synthesis and repair were recovered in approximately equal numbers (25% of colonies *557**, 23% of colonies *230**), and the strain with the adapted allele of the flagellar gene *fliM* was never recovered. (Fig. 3.2 B, Table 3.1) The adapted allele of *fliM* has been previously described to have no impact on murine colonization (103), which is consistent with the lack of selective advantage for this mutant observed here. In no case was more than one derived allele detected in a single clone recovered after murine infection, indicating that despite *H. pylori*'s natural competence and propensity for recombination (31) there was no recombination between the four single mutants introduced in the pool.

Because bacteria carrying the adapted allele of *imc1* (*imc1**) were recovered most commonly from mice infected with the pool of four mutants, we tested the *imc1** mutant's colonization potential in a single strain infection. Single infection with the *imc1** mutant resulted in loads similar to those observed with the pool of mutants ($p = 0.9699$) as well as recovery of bacteria from 9 of 10 inoculated mice (Fig. 3.2 A, right). All clones recovered from mice infected with the *imc1** mutant carried the adapted allele of *imc1* but not the other three mutations tested (Fig. 3.2 B, bottom), indicating that this mutation is stable during infection.

We further evaluated the role of the *imc1* mutation in mouse adaptation by comparing the relative colonization of a mutant with all five adapted alleles found in NSH57 (see Fig. 2.3) in the parental G27 background (5X Mutant; *fliM*/230*/557*/imc1*/1188**) versus wild type G27 and that of the *imc1** single adapted allele versus G27, as shown in Fig. 3.3. Both the 5X mutant and the *imc1** mutant outcompeted the parental strain G27, as measured by a statistically significant increase in bacterial load (5X versus G27*cat*, $p = 0.0002$; *imc1** versus G27*cat*, $p = 0.0001$), despite having only five or one adapted allele, respectively, in the parental genetic background. There was no significant difference in bacterial loads between the 5X mutant and the single *imc1** mutant ($p > 0.9999$). A control competition between unmarked G27 and G27 with a chloramphenicol resistance marker (G^{cat}) showed no significant marker effect ($p = 0.9948$); therefore the mutant strains, which are unmarked, were competed against chloramphenicol-resistant G27 for differentiation of the two strains on selective media.

As an additional test of the isogenic mutant *H. pylori* strains' ability to colonize mice, both the 5X mutant and the *imc1** single isogenic mutant were competed against the adapted strain, NSH57, as shown in Fig. 3.4. The 5X mutant, with all five adapted alleles in the G27 background, colonized mice to the same level as the adapted strain NSH57 ($p = 0.8143$). This suggests that these five mutations are sufficient to allow *H. pylori* to colonize mice, and that no mutations important for this phenotype were missed in sequencing. Additionally, the adapted allele of *imc1* alone, in the parental background (*imc1**) was also not significantly outcompeted by the adapted strain ($p = 0.3378$), indicating that the change in this single gene accounts for the majority of the adapted strain's ability to colonize mice. Once again, there was no significant difference in bacterial loads between the 5X mutant and the single *imc1** mutant ($p = 0.1941$). A control competition between unmarked NSH57 and NSH57 with a chloramphenicol resistance

marker (NSH57^{cat}) showed no significant marker effect (p = 0.6804), and the unmarked mutant strains were competed against the chloramphenicol-resistant NSH57 for differentiation on selective media.

The combination of four additional mutations found in NSH57 also increases colonization potential in the G27 background

In addition to the single base insertion in *imc1*, four additional mutations were identified through whole genome sequencing of mouse-adapted strain NSH57. A strain with the NSH57 alleles of these four genes (*HPG27_230*, *HPG27_557*, *HPG27_1188*, and *fliM*) in the G27 background (4X Mutant) was competed for one week against the adapted strain, NSH57, as shown in Fig. 3.5. The 4X mutant colonized mice to the same level as the adapted strain NSH57 (p > 0.999). Confirming the results shown in Fig. 3.4, the strain with the adapted allele of *imc1* alone, in the parental background, (*imc1**) was once again not significantly outcompeted by the adapted strain (p = 0.7379). Additionally, there was no significant difference in bacterial loads between the 4X mutant and the single *imc1** mutant (p = 0.3140). As above, a control competition between unmarked NSH57 and NSH57 with a chloramphenicol resistance marker (NSH57^{cat}) showed no significant marker effect (p = 0.9288), and the unmarked mutant strains were competed against the chloramphenicol-resistant NSH57 for differentiation on selective media. These results suggest that a combination of one or more of the non-*imc1* mutations identified in NSH57 also contributes to murine colonization potential.

***imc1* encodes an outer membrane-associated protein**

imc1 is an 879 bp gene encoding a 292 amino acid open reading frame (Fig. 3.6 A). Homologs are found in all fully sequenced genomes of *H. pylori* and exhibit at least 90% identity at the nucleotide level over at least 98% of the gene's length, with the exception of three *H.*

pylori strains in which the homolog has 87—88% identity over 98—100% of the length. The most closely related species to *H. pylori, H. acinonychus*, originally isolated from a cheetah (104), contains a homolog of *imc1* that is 86% identical over 100% of its length. The next closest relative is *H. cetorum*, found in dolphins and whales (105). Two *H. cetorum* genomes contain *imc1* homologs with 73—77% nucleotide identity over 53—63% of their length. No other sequences with significant homology were found, either within or outside the Helicobacteraceae (Fig. 3.6 B). *imc1* is not predicted to be in an operon (106) (Fig. 3.6 A), but the gene order is conserved for at least 7000 bp flanking *imc1* across the fully sequenced strains.

Imc1 has no predicted sequence or structural homology to other proteins, making predictions of function challenging. A full-length (parental allele) copy of Imc1 with a C-terminal FLAG epitope tag (Fig. 3.7 A) was used to probe the expression and localization of this uncharacterized protein. A single band of 45-50 kDa was detected by immunoblotting in whole cell lysates (Fig. 3.7 B). Although the SignalP algorithm weakly predicts an N-terminal signal sequence for Imc1, the FLAG-tagged protein was detected only in the cell pellet from cultures, not in the supernatant (Fig. 3.7 B), suggesting that it is not secreted into the extracellular environment. Separation of *H. pylori* lysates into cytoplasmic, periplasmic, inner membrane, and outer membrane fractions followed by immunoblotting revealed that Imc1 was present in the outer membrane fraction, co-fractionating with the outer membrane protein SabA (93) (Fig. 3.8).

To further investigate Imc1 function and localization, we used anti-FLAG affinity chromatography followed by mass spectrometry to identify proteins interacting with Imc1-FLAG. Extracts from wild type and Imc1-FLAG-expressing *H. pylori* were loaded on an anti-FLAG affinity column and peptides from proteins co-eluting specifically with Imc1-FLAG were identified by mass spectrometry. In addition to Imc1 itself, 11 proteins were identified as

interacting with Imc1 by a substantially greater number of peptides in Imc1-FLAG lysates than in lysates from wild type G27 bacteria (Table 3.2). These include several proteins found in the outer membrane or periplasm, further suggesting that Imc1 is located in the bacterial cell envelope. A previous study used affinity purification of Cag3 to isolate the Cag T4SS outer membrane complex, followed by mass spectrometry, and did not detect an interaction of Imc1 with the Cag T4SS (107). Additionally, we did not detect any Cag T4SS components by affinity purification of Imc1 followed by mass spectrometry (Table 3.2). These data suggest that Imc1 is not associated with *H. pylori*'s Cag T4SS.

The identified frameshift mutation in *imc1* is a null allele

The identified mouse-adapting mutation in *imc1* is predicted to cause loss-of-function due to a frameshift and resulting truncation. Insertion of a single A at genomic position 856,793 in mouse-adapted strains causes a frameshift starting from amino acid 72 and a premature stop at amino acid 78 of 293, resulting in loss of approximately 75% of the coding sequence (Fig. 3.6 A). *H. pylori* with the parental genetic background and either the single base insertion or a *catsacB* insertion-deletion in *imc1* (20% of the coding sequence replaced with *catsacB*) infected mice to similar levels ($p = 0.9961$ and $p = 0.8748$ for the comparison between *imc1** and Δ*imc1* in Figs. 3.3 and 3.4, respectively). A FLAG-tagged truncated allele of Imc1 was not detected in bacterial lysates, providing further evidence that the mutation causes loss of Imc1 expression (Fig. 3.7 B).

Imc1 affects *H. pylori* adherence to cultured murine, but not human, cells

The association of Imc1 with the *H. pylori* outer membrane suggests a possible role in interaction with host tissues and cells. To investigate this, *H. pylori* were co-cultured with human AGS gastric epithelial cells or with a conditionally immortalized mouse gastric epithelial cell

line (MGEC (108, 109)) at a multiplicity of infection (MOI) of 10:1 or 100:1 for AGS and MGEC, respectively, to identify differences between parental and mouse-adapted strains in adherence to cultured host cells. Mouse-adapted strain NSH57 displayed a modest but reproducible increase in adherence to the host cells of both human and murine origin. Mutation of *imc1* did not account for the increased adherence to human cells, as the *imc1** mutant adhered to AGS cells at levels comparable to the parental strain, G27 (Fig. 3.9 A). However, the *imc1** mutant adhered to MGEC cells at levels comparable to adapted strain NSH57 (Fig. 3.9 B), suggesting that the loss of Imc1 protein expression during murine adaptation may enhance adherence of *H. pylori* to the murine gastric epithelium.

Imc1 does not induce an IL-8/MIP-2 immune response

To investigate possible roles for Imc1 in activation of an immune response, which might provide a selective pressure against its expression during adaptation to mice, we measured cytokine responses to *H. pylori* by both cultured cells and infected mice. One of the hallmarks of *H. pylori*-induced gastritis is the infiltration of neutrophils and other inflammatory cells into the gastric tissue, which is initiated by production of the neutrophil chemoattractant IL-8 by gastric epithelial cells (110).

To investigate whether Imc1 contributes to sensing of *H. pylori* by gastric epithelial cells, production of IL-8 by the human gastric epithelial cell line AGS during co-culture with *H. pylori* was measured by ELISA on co-culture supernatants. Although coculture with the mouse adapted strain NSH57 resulted in slightly reduced IL-8 production by AGS cells compared to the parental strain G27, coculture with the isogenic *imc1** mutant resulted in production of IL-8 at levels similar to those seen with G27 (Fig. 3.10 A), indicating that this mutation is not responsible for reduced IL-8 induction by mouse-adapted bacteria. No differences were detected in the

production of a murine homolog of IL-8, MIP-2 (CXCL2), by the murine gastric epithelial cell line MGEC during co-culture with *H. pylori* with parental or adapted alleles of *imc1* (Fig. 3.10 B).

To investigate a possible immune-stimulatory role for Imc1 in vivo, cytokine gene induction was measured in the murine stomach after infection with the parental *H. pylori* strain G27, which has an intact *imc1* gene, and the isogenic mutant with the loss-of-function allele of *imc1* (*imc1**). Mice were infected with G27, *imc1**, or mock-infected and sacrificed after 24 hours, 1 week, or 2 weeks. Gastric mRNA levels of the murine IL-8 homologs MIP-2 and KC, the inflammatory mediator TNF-α, and additional cytokines representative of T_H1- (IFNγ), T_H2- (IL-4, IL-10), and T_H17- (IL-17) polarized responses were measured by qRT-PCR. After 24 hours of infection, MIP-2, IL-4, IFNγ, and TNF-α were upregulated and IL-10 was downregulated compared to their levels in mock-infected mice, but levels were not different in G27-infected compared to *imc1**-infected animals (Fig. 3.11, left). After 1 and 2 weeks of infection, none of the measured cytokines was significantly up- or down-regulated compared to levels in uninfected control animals, nor were mRNA levels significantly different in G27-infected compared to *imc1**-infected animals (Fig. 3.11, center and right). The results were not substantially altered by restricting the analysis to mice from which *H. pylori* colonies were recovered after sacrifice (data not shown).

Change in helical cell shape during mouse adaptation is not explained by identified mutations

Several periplasmic proteins have been found to influence cell shape in *H. pylori* (64–66), and we observed a change in the shape of our strains during mouse adaptation, where the mouse adapted strain NSH57 exhibits increased helical side curvature in comparison to the

parental strain G27 (103). This change in shape is illustrated in Fig. 3.12 A, and can be quantified using CellTool software (64). Measurement of the shape parameters of several hundred isolated bacterial cells enables comparison of the cell shape on a population level, illustrated by the smoothed histograms in Fig. 3.12. In addition to the increased side curvature of mouse-adapted strain NSH57, two additional mouse-adapted strains generated using the same mouse passaging protocol, 15-1.2 and 20-2.1, also display increased side curvature (Fig. 3.12 B).

The similar shape profiles of NSH57, 15-1.2, and 20-2.1 suggest that a genetic change common to all three strains may be responsible for their altered shape. Of the two genes with mutations in all three strains, *fliM* and *HPG27_230*, the latter appeared more likely due to its annotation as encoding OppD, an ATP-binding component of the Opp peptidoglycan recycling system. Bacteria with the NSH57 allele of *HPG27_230* in the G27 strain background (*230**) displayed side curvature profiles similar to those of G27, suggesting that the mutation in this gene is not responsible for the increased curvature of the mouse-adapted strains (Fig. 3.12 C). Because Imc1 appears to be associated with the periplasm, it could potentially affect bacterial cell shape. *H. pylori* with the NSH57 allele of *imc1* in the G27 background (*imc1**) displayed cell curvature intermediate between that of G27 and NSH57, as did bacteria with the adapted alleles of all five genes observed to have mutated in NSH57 (Fig. 3.12 D). No mutations were identified in the mouse-adapted strains in genes known to play a role in generation of *H. pylori's* helical cell shape (*csd1-6, ccmA, mviN, slt, HPG27_782, HPG27_1298;* data not shown). These results suggest that although Imc1 may play a role in cell shape, there are likely to be additional, unidentified genes involved.

DISCUSSION

A single mutation causing loss of expression of a previously uncharacterized protein, HPG27_792 or Imc1, was found to increase colonization efficiency of *H. pylori* strain G27 in mice. Bacteria carrying the adapted allele of *imc1* in the parental strain background outcompeted parental bacteria and were not outcompeted by adapted bacteria in a one-week competitive infection of mice. Imc1 is associated with the bacterial outer membrane and may influence adherence to murine gastric epithelial cells during *H. pylori* exposure.

The identified mutation in *imc1* was a single base insertion leading to a frameshift and appeared to cause loss of Imc1 function. Adaptive loss of function has been recently appreciated as a relatively common and important phenomenon in several bacterial systems (111), including adaptation to the murine gut by *E. coli* (88). Loss of function can be beneficial by altering the regulation of other genes or by removing a factor that is detrimental in the current environment (111). Since Imc1 does not appear to have any features of a regulatory protein, the latter case is more likely to apply here.

imc1 is conserved across the range of *H. pylori* strains sequenced to date, all of which infect the human stomach. It is also found in a subset of the gastric non-*pylori* species, but is absent from all *Helicobacter* species isolated from non-gastric sites, such as the intestine and liver. This suggests a possible, as yet unidentified, role for *imc1* in the stomach environment.

Although the mutation in *imc1* had the largest single effect on murine infection, the combination of the other four mutations identified in the adapted strain NSH57, when placed in the parental G27 background, also facilitated murine colonization. This suggests that one of these four genes, or some combination of them, also contribute to the ability of *H. pylori* to colonize mice, although the mechanisms remain unexplored.

Fractionation and proteomics approaches indicated that Imc1 appears to be localized to the periplasm or outer membrane of *H. pylori.* Mutations in surface exposed proteins have been observed in clinical isolates (24, 32) and may benefit *H. pylori* by altering the degree to which it interacts with its host. Alterations in the gene presence, sequence, and expression status of outer membrane adhesins, such as BabA, have been observed during adaptation of *H. pylori* to mice (98), as well as to gerbils (98), and macaques (96). The mouse-adapted strain NSH57 exhibits increased adherence to gastric epithelial cells of both human and murine origin. The identified mutation in *imc1* appears to mediate this increase in bacterial adherence to murine, but not to human cells, by a mechanism that remains to be determined. One or more of the other identified mutations, which affect genes involved in cell wall synthesis and an uncharacterized protein, may be responsible for the increased adherence to human-derived cells.

Previous studies have reported diminished induction of host inflammatory responses after adaptation of *H. pylori* to mice. Genes of the *cag* pathogenicity island (PAI), a 40 kb region encoding a type IV secretion system (Cag T4SS) that delivers immune-stimulatory bacterial products to host cells (67), are often deleted during adaptation of *H. pylori* to mice (112–114). *H. pylori* murine adaptation has also been associated with decreased capacity to induce inflammatory responses in the absence of *cag* PAI gene loss (113). These lines of evidence suggest that some feature of the murine immune system exerts a differential selective pressure from that of the human host, at least with respect to cytokine induction.

Fragments of *H. pylori* peptidoglycan enter the gastric epithelial cell cytosol via the bacterial Cag T4SS and are sensed by Nod1, which recognizes the GM-tripeptide signature of Gram-negative peptidoglycan, leading to NF-κB activation and upregulation of IL-8 (94). IL-8 and its murine functional homologs MIP-2 (CXCL2) and KC (CXCL1) act as chemoattractants

for neutrophils (115, 116). Infiltrating immune cells produce additional cytokines, such as interferon gamma, leading to a predominantly T_H1-polarized immune response to *H. pylori* colonization (117). This inflammatory response has been shown to be important for clearance of the related bacterium *H. felis* from mice (95) and reducing the induction of inflammatory mediators may help *H. pylori* avoid clearance from the murine stomach and establish a chronic infection.

Although we did not detect substantial differences in the host cytokine response to parental or mouse-adapted *H. pylori*, these responses may have been present at levels below our detection threshold or have involved components of the immune response that we did not measure. It remains possible that pressure from a murine immune response to Imc1 may have provided selective pressure to mutate *imc1* during murine adaptation, enabling the bacteria to survive in the new host. Although Imc1 appears to be associated with the bacterial outer envelope, it remains to be determined whether Imc1 can be directly sensed by either gastric epithelial cells or dedicated immune cells. Elucidation of the specific effectors of the murine immune system that may have provided the selective pressure for mutation of *imc1*, as well as other genes mutated during mouse adaptation, could be explored though the use of mice deficient in specific immune components.

A change in *H. pylori*'s cell shape during was observed during adaptation to colonize mice, with increased helical curvature of strains NSH57, 15-1.2, and 20-2.1 as compared to their parental strain G27. Mutations in *H. pylori* cell shape determining genes that result in loss of helical twist and/or curvature have been demonstrated to negatively affect murine colonization potential (64, 65). The mechanism by which cell shape contributes to infectivity is currently under investigation, but may be related to the ability to penetrate the gastric mucus and reach the

epithelial cell surface and gastric glands (Laura Martinez, unpublished). The increased curvature of the mouse-adapted strains seen here has not been fully recapitulated by isogenic mutants in any of the genes identified as differing between these strains. This raises the possibility that other mutations, one or more of which contributes to the change in cell shape, occurred but were missed due to insufficient sequence coverage.

MATERIALS AND METHODS

Media and antibiotics

H. pylori strains were grown on solid horse blood agar (HB) plates containing 4% Columbia agar base (BD Bioscience), 5% defibrinated horse blood (HemoStat Laboratories), 0.2% β-cyclodextrin (Sigma), vancomycin (10 μg ml^{-1}; Sigma), polymyxin B (2.5 U ml^{-1}; Sigma), and amphotericin B (8 μg ml^{-1}; Sigma) at 37°C in a trigas incubator equilibrated with 10% oxygen, 10% carbon dioxide, and 80% nitrogen. For antibiotic resistance marker selection (chloramphenicol acetyltransferase gene [*cat*]), bacterial media were additionally supplemented with chloramphenicol (15 μg ml^{-1}; Sigma), and for sucrose sensitivity counter-selection (levansucrase gene [*sacB*]), bacterial media were additionally supplemented with sucrose (60 mg ml^{-1}; Fisher). When bacteria were cultured from mouse stomachs, cefsulodin (5 μg ml^{-1}; Sigma), trimethoprim (5 μg ml^{-1}; Sigma), and bacitracin (200 μg ml^{-1}; Fisher) were added. For liquid culture, *H. pylori* was grown in Brucella broth (BD Biosciences) containing 10% fetal bovine serum (BB10; HyClone) with shaking in the trigas incubator.

Strains

Strains used are listed in Table 3.3. Isogenic mutant strains were constructed by allelic exchange (102). A *cat sacB* insertion cassette, conferring chloramphenicol resistance and sucrose sensitivity and flanked by regions of homology to the site of interest, was constructed using PCR

with splicing by overlapping extension (SOEing) (118) using the primers indicated in Tables 3.3 and 3.4 for each gene of interest. This insertion cassette was introduced into parental strain G27 by natural transformation and replaced the portion of the targeted gene indicated for each strain (Table 3.3). The resulting gene disruption mutants were transformed with DNA amplified from adapted strain NSH57 using the primers indicated in Table 3.4 or with genomic DNA isolated from strain NSH57, and clones in which the *cat sacB* cassette had been replaced with the adapted allele of the gene of interest were selected based on sucrose resistance and chloramphenicol sensitivity. PCR and Sanger sequencing were used to confirm the allelic exchange. Strain IEC67 contains a FLAG-tagged allele of gene *HPG27_792/imc1* at the native locus as the sole copy of this gene and was generated by PCR SOEing of the products of primers oIEC183, 185, 186, and 187 (Table 3.4) to place two copies of an *H. pylori* codon-optimized FLAG epitope tag followed by a *cat* selection cassette at the C-terminus of Imc1 as a translational fusion.

Mouse colonization

Female C57BL/6 mice four to six weeks old were obtained from Jackson Laboratories and certified free of endogenous *Helicobacter* infection by the vendor. The mice were housed in an Association for the Assessment and Accreditation of Laboratory Animal Care-accredited facility in sterilized microisolator caging and provided with irradiated PMI 5053 rodent chow and acidified, reverse-osmosis-purified water ad libitum. The Institutional Animal Care and Use Committee of the Fred Hutchinson Cancer Research Center approved all manipulations. *H. pylori* infection by oral gavage of 5×10^7 CFU in a volume of 0.1 ml BB10 and recovery from the stomach were performed as described previously (119). To determine bacterial loads, dilutions of mouse stomach homogenates were plated on the appropriate selective medium. CFU

counts were normalized to the weight of the stomach tissue, and 10 colonies were considered the limit of detection.

Genotyping

Colonies recovered from the stomachs of infected mice were genotyped for selected mutations using tetra-primer PCR (120). A multiplexed reaction containing 5 pmol each of two outer, non-genotype-specific primers and 10 pmol each of two inner, genotype-specific primers (Table 3.4) allowed differentiation of wild-type and mutant alleles differing by a single nucleotide. Specificity is conferred by a mismatch at the 3' end in one inner primer, and a second mismatch 2 base pairs upstream.

Bioinformatics

Pathosystems Resource Integration Center (PATRIC, (121)) databases were used for identification of *imc1* homologs. Phylogenetic trees were generated from whole genome amino acid sequences using PATRIC, and visualized and trimmed using TreeGraph 2 (122). *H. pylori* operon structure was previously defined (106). SignalP (123) was used to identify the presence of signal peptides in amino acid sequences.

Immunodetection of Imc1

For detection of FLAG-tagged Imc1, *H. pylori* strains were grown in liquid medium to mid-log phase. Cells were pelleted at 3,220 × g for 5 min, washed with PBS, and resuspended at 10 OD_{600} units ml^{-1} in Laemmli buffer (50 mM Tris pH 6.8, 2% SDS, 10% glycerol, 0.075% bromophenol blue, 5% β-mercaptoethanol). The culture supernatant was filtered through a 50 kDa molecular weight cut-off membrane (Amicon) to remove large, serum-derived proteins or left unfiltered, then concentrated by precipitation with 20% trichloroacetic acid (TCA) followed by two washes in cold acetone, and resuspended in 375 µl Laemmli buffer. Samples were run on

15% SDS-PAGE gels and transferred to PVDF membranes using an iBlot system (Invitrogen). Mouse M2 anti-FLAG primary antibody (Sigma) was used at a dilution of 1:1000 to detect FLAG-tagged proteins, followed by goat anti-mouse secondary antibody conjugated to horseradish peroxidase (Jackson ImmunoResearch Laboratories) at a dilution of 1:10,000. SabA was detected without stripping membranes using rabbit polyclonal antiserum (124) at a dilution of 1:10,000 followed by donkey anti-rabbit secondary antibody conjugated to horseradish peroxidase (Santa Cruz Biotechnology).

Fractionation of cellular compartments

H. pylori grown on fresh HB plates were collected in cold PBS, pelleted (10 min at 3,220 × g), and resuspended in 0.2 M Tris pH 8, 1 M sucrose, 1 mM EDTA, and 1 mg ml^{-1} lysozyme. After incubation for 5 min at room temperature, the suspension was diluted 5-fold with water and placed on ice. Suspensions were spun in a SW 41 Ti rotor (Beckman) at 150,000 × g at 4°C for 45 min, and the supernatant (periplasmic fraction) concentrated 6-fold by TCA precipitation. The pellet was resuspended in 5 ml of 10 mM Tris pH 7.5, 5 mM EDTA, 0.2 mM DTT and 10 µl Benzonase endonuclease (EMD Millipore) and sonicated on ice for 4 pulses of 20 sec each (setting 5.0 on a Branson Sonifier 250). The sonicate was spun at 1600 × g at 4°C for 10 min to remove unbroken cells and the resulting supernatant spun at 250,000 × g at 4°C for 2 hours in a Beckman VTi 65.2 rotor. The cytoplasmic fraction (supernatant) was concentrated 5-fold by TCA precipitation. The crude membrane pellet was further fractionated using a sucrose density gradient. Membranes were resuspended in 2 ml of 10 mM Tris pH 7.5, 15% sucrose, 5 mM EDTA, and 0.2 mM DTT. The membrane suspension was layered on a sucrose gradient consisting of 1 ml 55%, 2.25 ml 50%, 2.25 ml 45%, 2.25 ml 35%, and 2.25 ml 30% sucrose in 10 mM Tris pH 7.5 and 5 mM EDTA and spun in an SW-41 rotor at 250,000 × g at 4°C for 15

hours with slow brake. Fractions were collected, diluted 3-fold in water, spun for 2 hours in 30 kDa molecular weight cut off filters (Amicon) to remove sucrose, and resuspended in 1 ml TBS (50 mM Tris pH 8, 150 mM NaCl). All samples were precipitated with TCA and resuspended in 100 µl Laemmli buffer. SDS-PAGE and detection of FLAG were performed as described above. Membranes were reprobed with anti-SabA as described above to detect a known outer membrane component.

Affinity chromatography and mass spectrometry

Approximately 125 OD_{600} units of *H. pylori* were harvested from HB plates and suspended in TBS. Bacteria were pelleted, frozen, and resuspended in lysis buffer (50 mM Tris pH 8, 150 mM NaCl, 1% NP-40, 1X cOmplete protease inhibitor (Roche)). Lysates were sonicated for 3 pulses of 20 sec each (setting 5.0 on a Branson Sonifier 250) and unbroken cells pelleted. The supernatant was loaded on an Anti-FLAG M2 agarose affinity column (Sigma) prepared according to the manufacturer's instructions and equilibrated with TBS. Columns were sealed and incubated at 4°C for 70 min with rotation, then washed with TBS to remove unbound proteins. Proteins bound to the column via direct or indirect interaction with FLAG were eluted in 6 fractions of 1.4 ml each with 0.1 M glycine, pH 3.5 into 80 µl of 1 M Tris, pH 8. Samples were submitted for mass spectrometry analysis by the FHCRC Proteomics Shared Resource. Data were analyzed using the FHCRC Computational Proteomics Analysis System (CPAS) software and filtered using a Peptide Prophet score ≥ 0.9 and % ionization ≥ 30. Proteins identified from Imc1-FLAG samples, but not from wild type samples, in two independent experiments were scored as hits.

Coculture of *H. pylori* with host cells

Human AGS cells (ATCC CRL-1739), derived from a patient with gastric adenocarcinoma, were grown in DMEM + 10% FBS at 37°C and 10% CO_2 and seeded at 5×10^4 cells/well in 24-well plates for 16 h before infection. Conditionally immortalized mouse gastric epithelial cells (MGEC, (108, 109)) were grown in DMEM + 10% FBS + 300 pg ml^{-1} recombinant mouse interferon gamma (R&D Systems) at 33°C and 10% CO_2 (immortalized growth condition). MGEC were seeded at 5×10^4 cells/well in 24-well plates in medium without interferon gamma at 37°C and 10% CO_2 for 24 h before infection to induce a primary-like state. *H. pylori* cells from mid-log-phase liquid cultures were added at a multiplicity of infection of 10:1 or 100:1 in the appropriate cell culture medium with the addition of 10% BB10 to support bacterial growth. Infections were synchronized by brief centrifugation (1000 rpm for 5 min). Supernatants were collected at 1, 6, and 24 h, and release of human IL-8 from AGS and mouse MIP-2 from MGEC was assayed using an enzyme-linked immunosorbent assay (ELISA) following the manufacturer's protocol (IL-8: BioLegend; MIP-2: R&D Systems). Growth of non-adherent bacteria was measured by plating the appropriate dilution of the co-culture supernatant on HB plates. Adherence of *H. pylori* to host cells and growth on the cell surface were assayed by washing wells 6 times with cold PBS to remove non-adherent cells, treatment with 500 µl of 0.1% saponin in PBS for 5 minutes at room temperature and mechanical disruption with a cell scraper to release cells from the wells. Appropriate dilutions of the cell suspension were plated on HB plates to enumerate adherent bacteria.

qRT-PCR measurement of cytokines in mouse stomach tissue

Mouse stomach tissue was stabilized in RNA*later* (Ambion) immediately after sacrifice and disrupted using a PowerGen 125 rotor-stator homogenizer (Fisher Scientific). Total RNA was

isolated using the Qiagen RNeasy Mini or Midi kit according to the manufacturer's protocol for isolation of total RNA from animal tissues, including the on-column DNase digestion step. cDNA synthesis and PCR were performed in a single step using the SuperScript III Platinum One-Step Quantitative RT-PCR System with ROX (Invitrogen) on an ABI PRISM 7900HT Sequence Detection System. Murine MIP-2 (125), KC (125), IL-4 (126), IL-10 (127), IL-17 (126), TNF-α (128) and IFNγ (126) were detected using published TaqMan primers and FAM-labeled probes. Beta-2 microglobulin served as an endogenous control for normalization and was detected using published TaqMan primers and HEX-labeled probe (129). Each cytokine was measured in triplicate technical replicates in each sample, normalized to beta-2 microglobulin, and expressed as fold change relative to the levels in uninfected control mice (ΔΔCt method).

Shape measurement

Bacterial cell shape was measured as previously described (64, 103). Briefly, *H. pylori* cells were grown in liquid culture to mid-log phase and fixed with a 4% paraformaldehyde/PBS solution containing 10% glycerol. Fixed cells were mounted on glass coverslips and phase contrast images were collected using a Nikon Eclipse TE 200 inverted microscope with a 100 X Plan Apo (1.40 NA) oil-immersion objective and Nikon CoolSNAP HQ CCD camera controlled by MetaMorph software (MDS Analytical Technologies). Images were thresholded and stray pixels removed using ImageJ software (imagej.nih.gov). A minimum of 100 bacterial cells was analyzed per strain. CellTool software was used to capture the polygonal outline of each cell and calculate shape parameters, including side curvature (64). Data was displayed as a smooth histogram of the population distribution.

Statistical analysis

Statistical analysis was performed using GraphPad Prism version 6.02 for Windows (GraphPad Software, La Jolla, CA). Murine colonization loads of single strain *H. pylori* infections were compared using one-way analysis of variance (ANOVA) on log-transformed data with Tukey's test for multiple comparisons. Murine colonization loads for competitive infections were compared using one-way ANOVA on log-transformed data with Šidák's test for multiple comparisons.

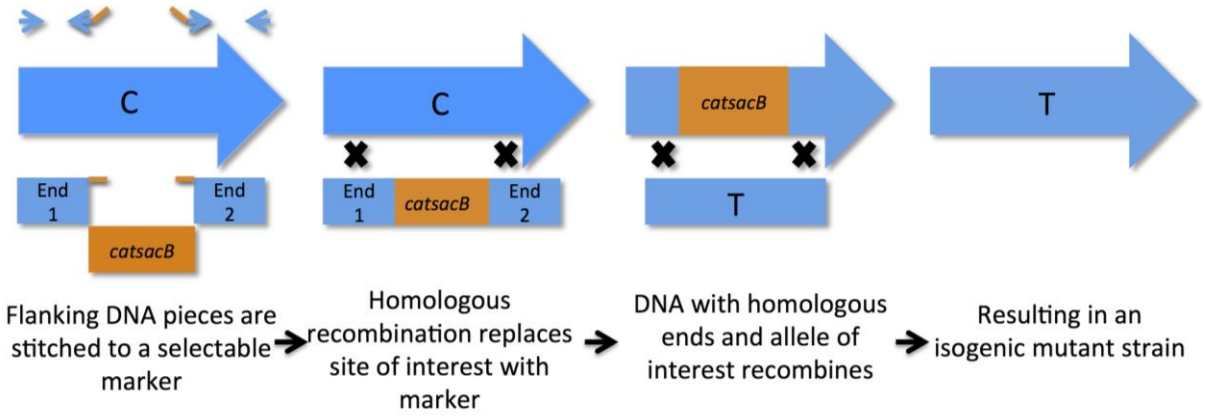

Figure 3.1 Schematic of allelic exchange strategy for strain construction. The *catsacB* construct is a counter-selectable marker conferring chloramphenicol resistance and sucrose sensitivity. "C" and "T" represent different alleles of the gene of interest.

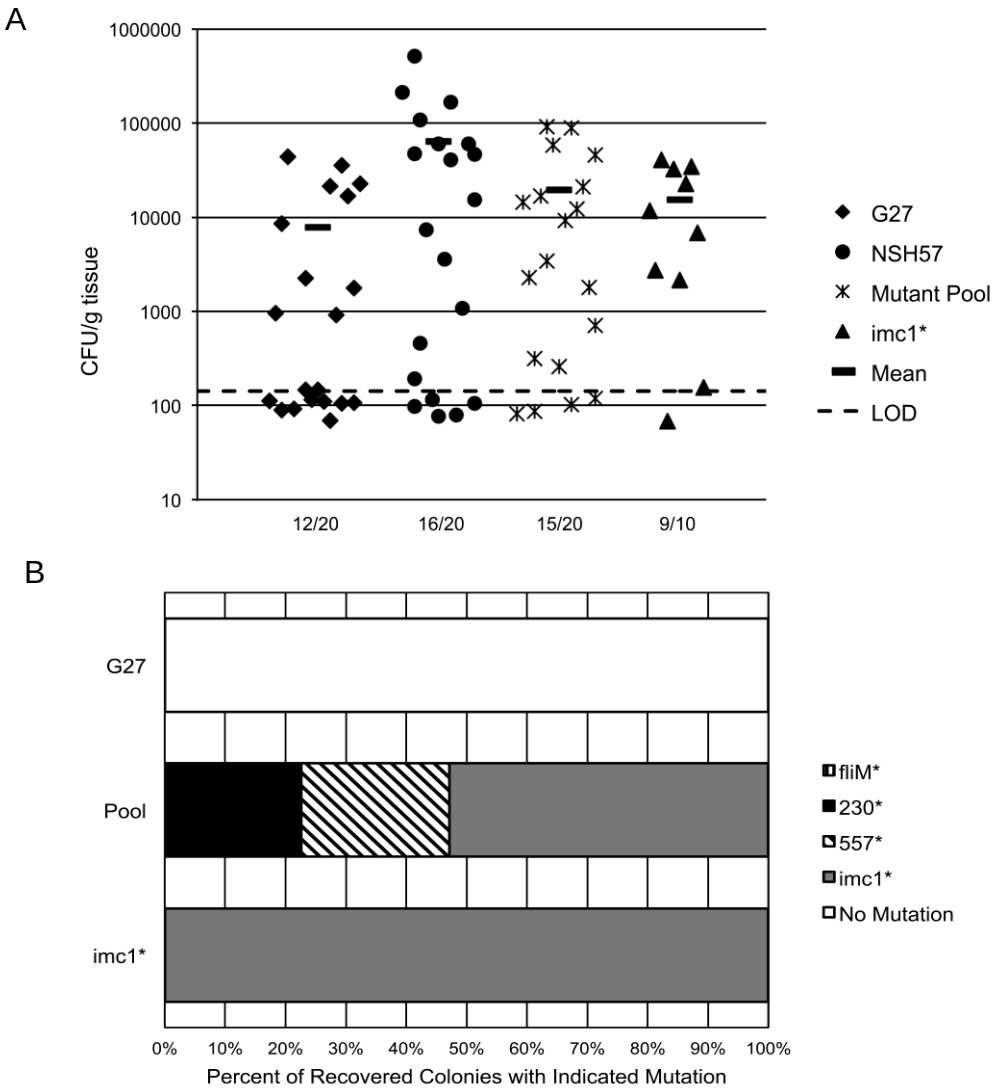

Figure 3.2 Mutation in *imc1* is a major driver of *H. pylori* mouse adaptation. (A) Mice were infected with parental strain G27, adapted strain NSH57, a pool containing equal proportions of four strains with the adapted allele of *fliM*, *HPG27_230*, *HPG27_557*, or *HPG27_792*/*imc1* in the parental genetic background (Mutant Pool), or a stain with the adapted allele of *imc1* in the parental background (*imc1**). Bacterial loads recovered from individual mouse stomachs after a one week infection with 5×10^7 CFU of the indicated *H. pylori* strain are plotted as CFU g^{-1} stomach tissue. Mean loads are indicated by bars and the limit of detection (LOD) by a dashed line. The proportion of mice from which bacteria were recovered is indicated below the load data. Data are compiled from 3 independent experiments, except *imc1**, which is a single experiment; N = 20 mice (G27), 20 mice (NSH57), 19 mice (Mutant Pool), 10 mice (*imc1**). All comparisons of bacterial loads between groups are not statistically significant ($p > 0.01$) as determined by one-way ANOVA on log-transformed data with Tukey's test for multiple comparisons. (B) Colonies recovered from mice infected in (A) were genotyped using a PCR assay to determine whether they carried the parental or adapted allele of *HPG27_230*, *HPG27_557*, *imc1*, and *fliM* and the proportion of colonies with the adapted genotype of each gene is indicated. N = 19 colonies recovered from 9 mice (G27), 53 colonies from 15 mice (Pool), and 24 colonies from 8 mice (*imc1**).

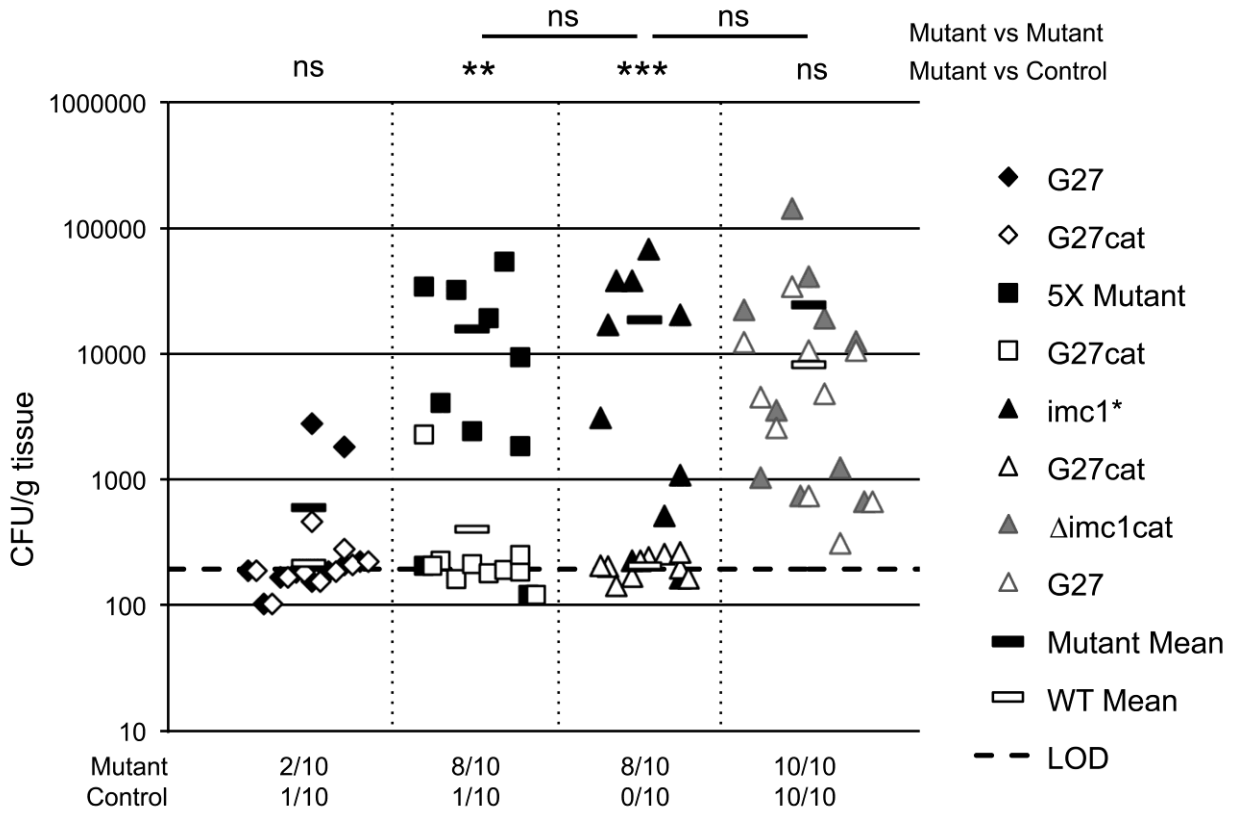

Figure 3.3 Mutation in *imc1* allows *H. pylori* to outcompete the parental strain in mouse colonization. Mice were infected with parental strain G27, a strain with the adapted alleles of *HPG27_230*, *HPG27_557*, *imc1*, *HPG27_1188*, and *fliM* in the parental genetic background (5X Mutant), a strain with the adapted allele of *imc1* in the parental background (*imc1**), or a strain with *imc1* replaced by an antibiotic resistance cassette (Δ*imc1*) in the parental background in 1:1 competition with G27^{cat} or G27 (for Δ*imc1*). The relative bacterial loads were determined by plating on media with and without chloramphenicol. Bacterial loads recovered from individual mouse stomachs after a one week infection with 5×10^7 CFU of the indicated *H. pylori* strain are plotted as CFU g^{-1} stomach tissue. Mutant bacterial loads are represented by filled symbols and control bacterial loads by open symbols, and data points aligned vertically represent bacteria recovered from the same mouse. The mean bacterial load for the mutant strain and control strains are denoted by filled and open bars, respectively and the limit of detection (LOD) by a dashed line. The proportion of mice from which bacteria were recovered is indicated below the load data. *, $p \leq 0.01$; **, $p \leq .001$; ***, $p \leq .0001$; ns, $p > 0.01$ for comparison of bacterial loads as determined by one-way ANOVA on log-transformed data with Šidák's test for multiple comparisons. Data are compiled from 2 independent experiments; N = 10 mice per group.

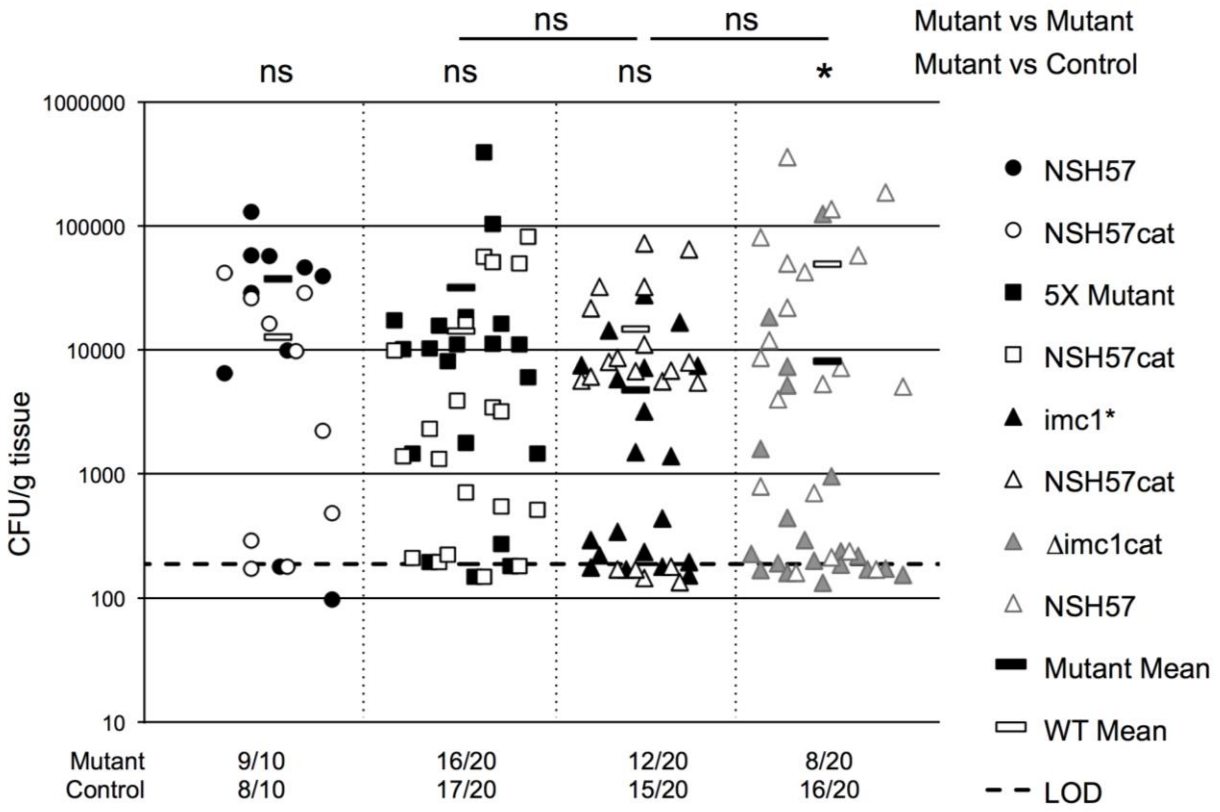

Figure 3.4 Mutation in *imc1* gives *H. pylori* equal mouse colonization ability to the adapted strain. Mice were infected with adapted strain NSH57, a strain with the adapted alleles of *HPG27_230*, *HPG27_557*, *imc1*, *HPG27_1188*, and *fliM* in the parental genetic background (5X Mutant), a strain with the adapted allele of *imc1* in the parental background (*imc1**), or a strain with *imc1* replaced by an antibiotic resistance cassette (Δ*imc1*) in the parental background in 1:1 competition with NSH57^{cat} or NSH57 (for Δ*imc1*). The relative bacterial loads were determined by plating on media with and without chloramphenicol. Bacterial loads recovered from individual mouse stomachs after a one week infection with 5×10^7 CFU of the indicated *H. pylori* strain are plotted as CFU g^{-1} stomach tissue. Mutant bacterial loads are represented by filled symbols and control bacterial loads by open symbols, and data points aligned vertically represent bacteria recovered from the same mouse. The mean bacterial load for the mutant strain and control strains are denoted by filled and open bars, respectively and the limit of detection (LOD) by a dashed line. The proportion of mice from which bacteria were recovered is indicated below the load data. *, $p \leq 0.01$; **, $p \leq .001$; ***, $p \leq .0001$; ns, $p > 0.01$ for comparison of bacterial loads as determined by one-way ANOVA on log-transformed data with Šidák's test for multiple comparisons. Data are compiled from 2 independent experiments; N = 10 mice for NSH57 vs NSH57^{cat}, 20 mice per group for all others.

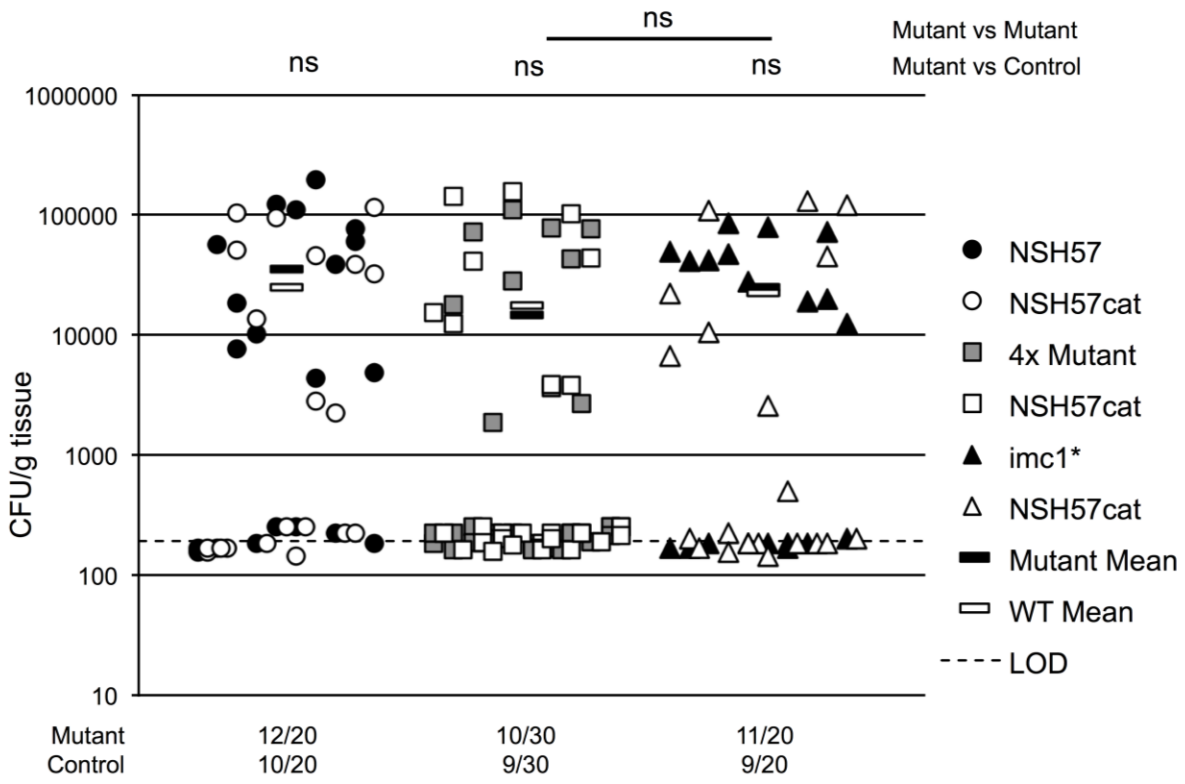

Figure 3.5 Four additional mutations in combination allow *H. pylori* to infect mice to similar loads as the adapted strain. Bacterial loads recovered from individual mouse stomachs after a one week infection with 5×10^7 CFU of the indicated *H. pylori* strains are plotted as CFU g^{-1} stomach tissue. Mice were infected with parental strain G27, a strain with the adapted alleles of *HPG27_230*, *HPG27_557*, *HPG27_1188*, and *fliM* in the parental genetic background (4X Mutant), or a strain with the adapted allele of *imc1* in the parental background (*imc1**) in a 1:1 competition with NSH57^{cat}. Relative loads were determined by plating on media with and without chloramphenicol. Mutant bacterial loads are represented by filled symbols and control bacterial loads by open symbols, and data points aligned vertically represent bacteria recovered from the same mouse. The mean bacterial load for the mutant strain and control strains are denoted by filled and open bars, respectively, and the limit of detection (LOD) by a dashed line. The proportion of mice from which bacteria were recovered is indicated below the load data. Ns indicates $p > 0.01$ for comparison of bacterial loads as determined by one-way ANOVA on log-transformed data with Šidák's test for multiple comparisons. Data are compiled from 3 independent experiments; N = 20 mice per group for NSH57 vs NSH57^{cat} and *imc1** vs NSH57^{cat}; 30 mice for 4X Mutant vs NSH57^{cat}.

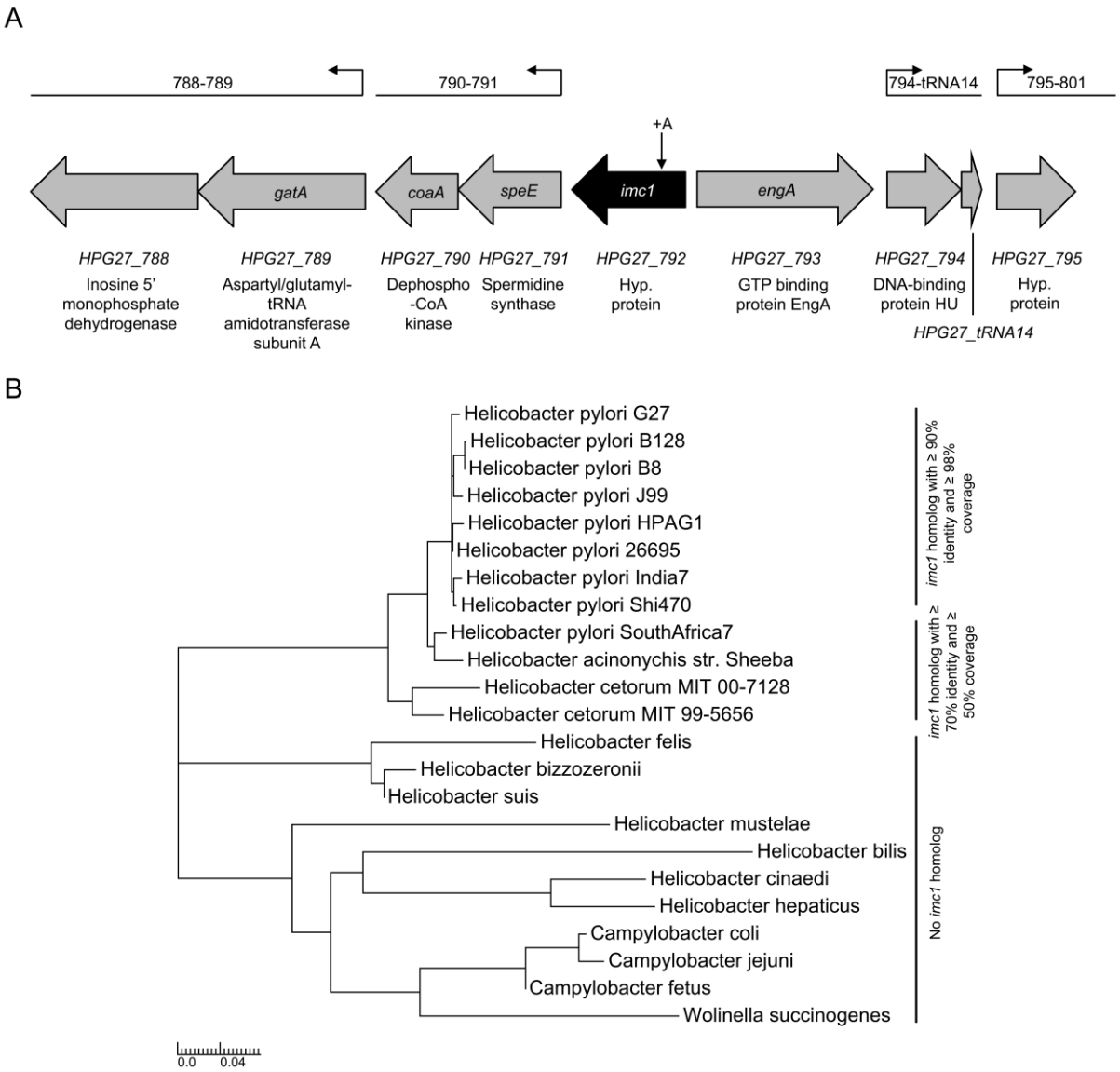

Figure 3.6 *imc1* encodes a protein of unknown function and is conserved across *H. pylori* strains. (A) Chromosomal location of *imc1* (black) and flanking genes. Predicted operon structure in the region is indicated above. The approximate location of the insertion mutation identified in *imc1* is indicated. (B) Presence of *imc1* homologs in selected *H. pylori* strains and related bacterial species. Phylogram based on complete genome sequence data; scale bar indicates amino acid substitutions per site.

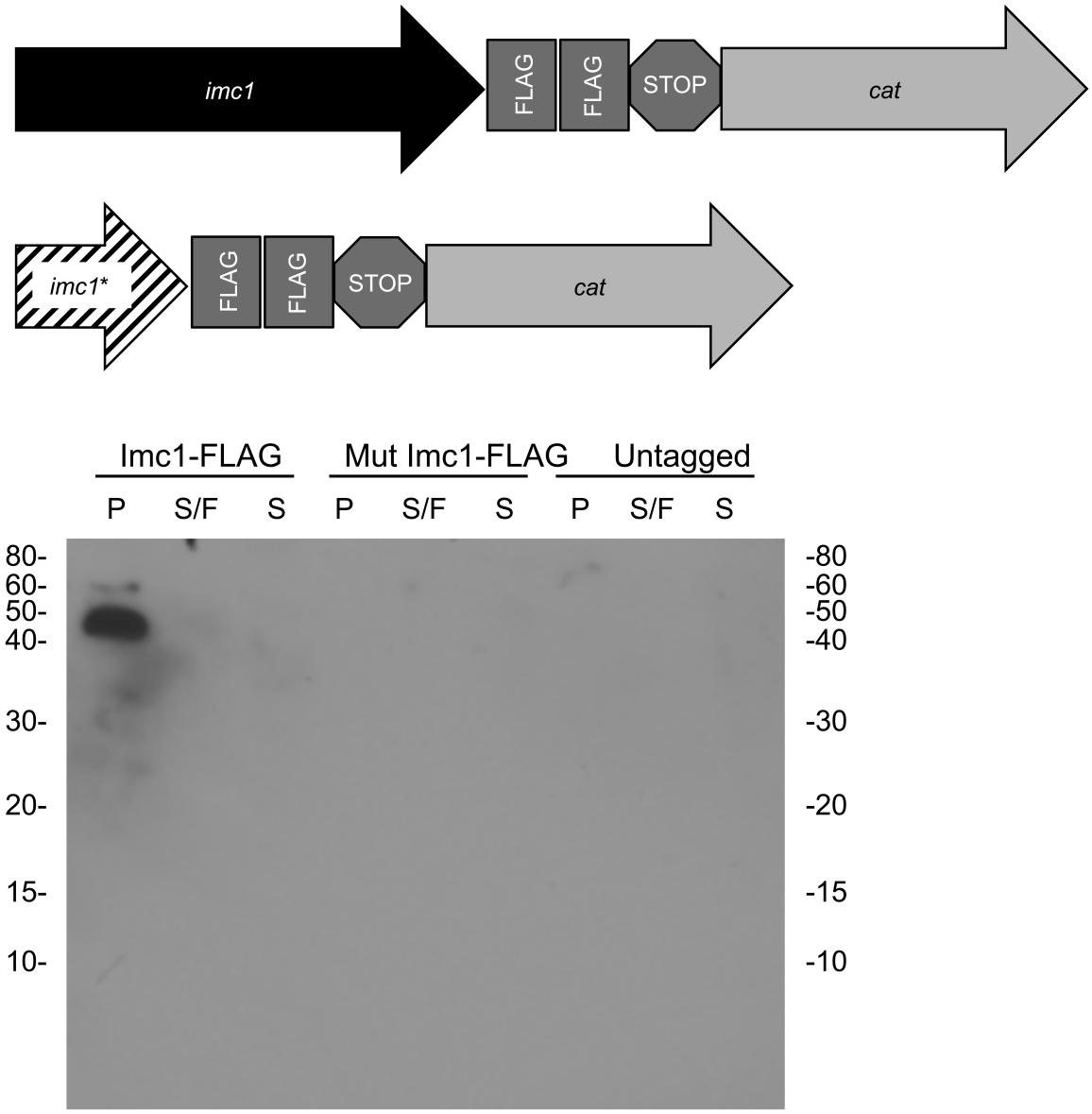

Figure 3.7 Imc1 is expressed in *H. pylori*. (A) Full length (parental allele) and truncated (adapted allele) Imc1 was expressed at the native locus with two copies of an *H. pylori* codon-optimized FLAG epitope tag as a C-terminal translational fusion. (B) Full length Imc1-FLAG was detected in bacterial lysates but not in culture supernatants, and truncated Imc1-FLAG was not detected. P: pellet; S/F: supernatant, filtered; S: supernatant.

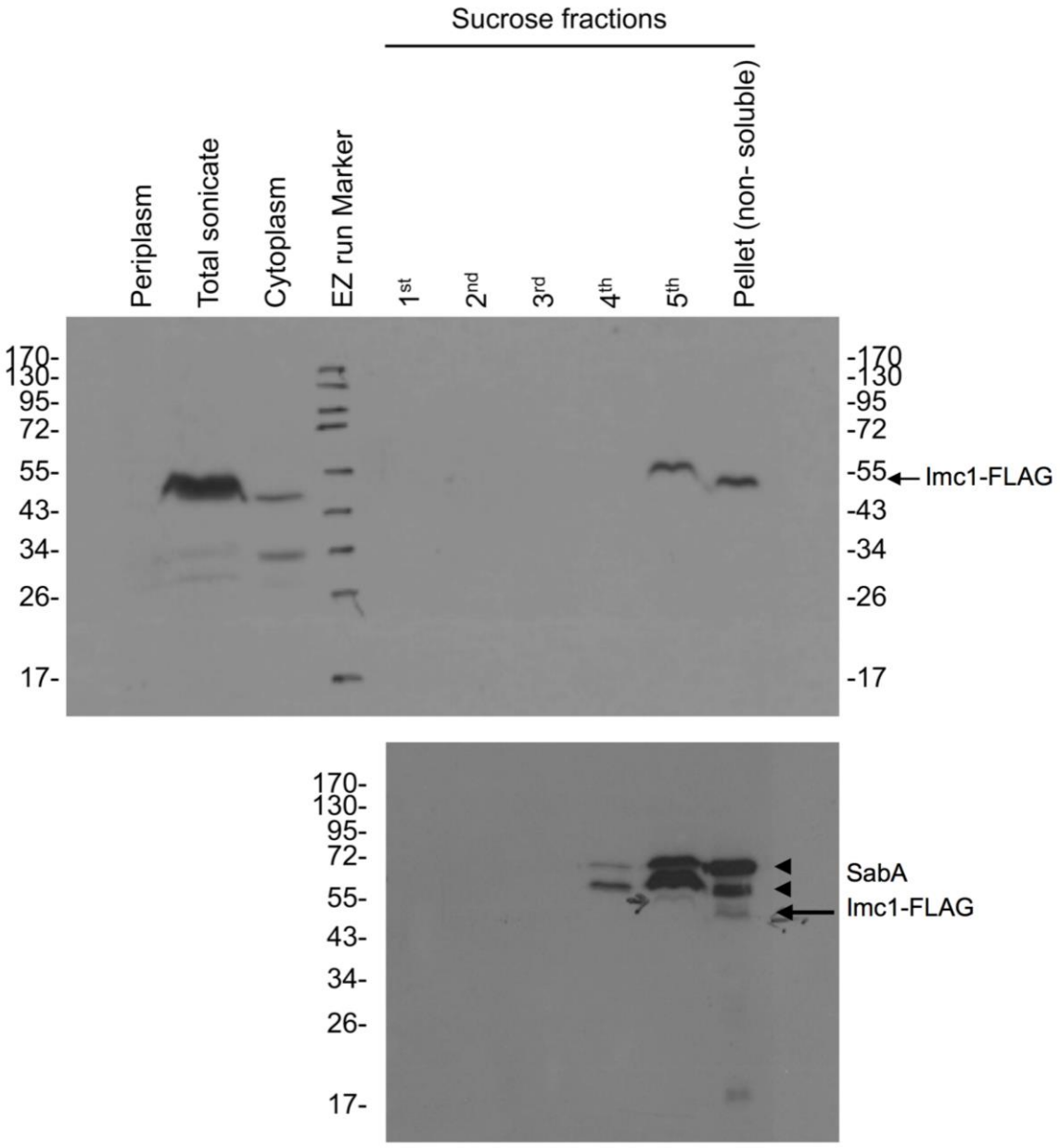

Figure 3.8 Imc1 is expressed in the *H. pylori* outer membrane. Fractionation of bacterial lysates into cytoplasmic, periplasmic, and membrane compartments revealed Imc1-FLAG in dense membrane fractions, co-fractionating with the outer membrane protein SabA. Top: anti-FLAG, bottom: same membrane re-probed with anti-SabA without stripping. Arrowheads indicate SabA, arrow indicates Imc1-FLAG.

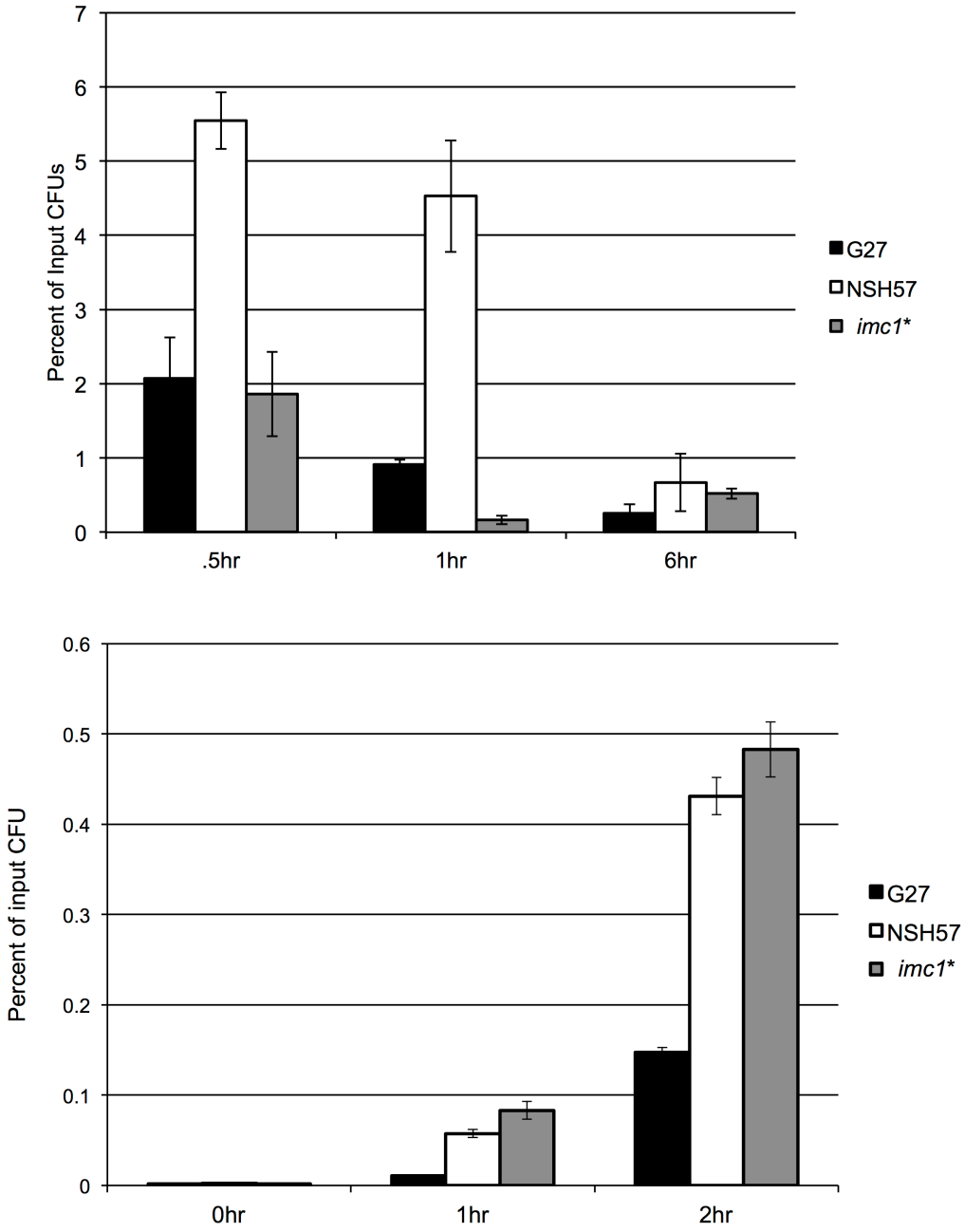

Figure 3.9 Mouse adaptation enhances *H. pylori* adherence to host cells. Mouse adapted strain NSH57 displayed increased adherence relative to parental strain G27. The mutation in *imc1* did not affect adherence to human cells, but appeared to mediate the increased adherence to murine cells. Human AGS (A) or murine conditionally immortalized (B) gastric epithelial cells were co-cultured with the indicated *H. pylori* strains at a multiplicity of infection of 10:1 (A) or 100:1 (B) for the indicated time periods. Bacterial adherence to cells was assessed by plating of cell suspensions after extensive washing to remove non-adherent bacteria. CFU counts normalized to the inoculum are plotted. All experiments consist of biological triplicates and are representative of 3 (A) or 2 (B) independent experiments.

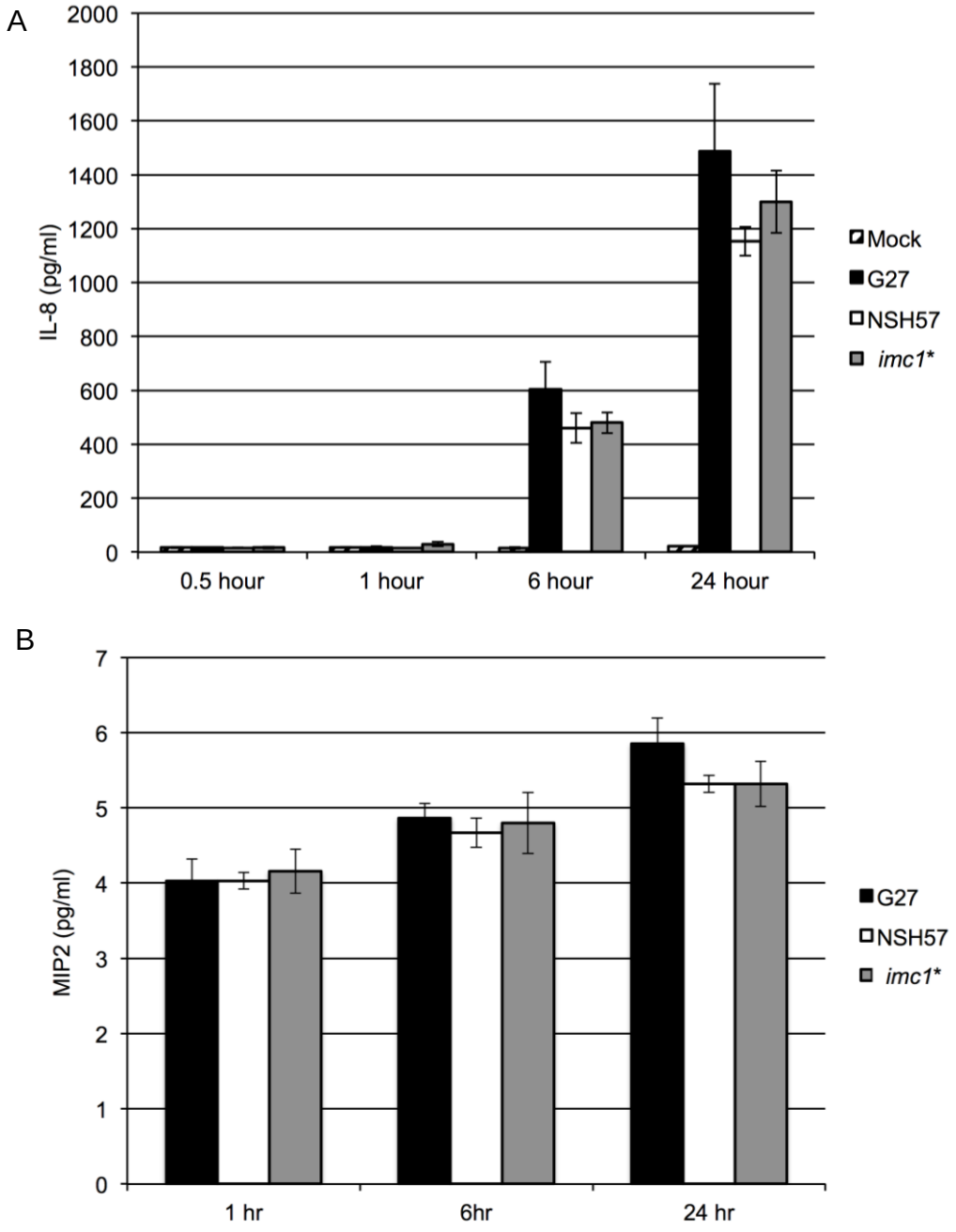

Figure 3.10 Mutation in *imc1* does not affect induction of IL-8 or MIP-2 in cocultured gastric epithelial cells. Release of cytokines into the co-culture medium was measured by ELISA for human IL-8 (A) or murine MIP-2 (B) following the coculture experiments described in Fig. 3.9.

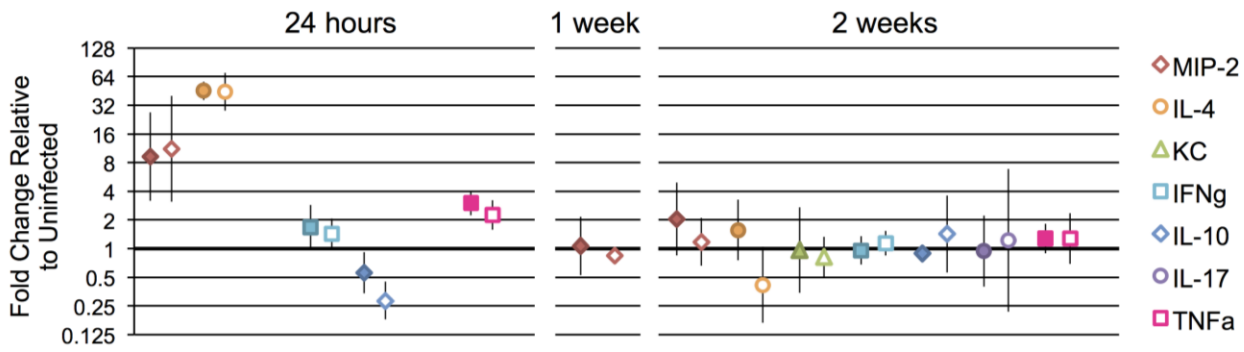

Figure 3.11 Mutation in *imc1* does not affect induction of proinflammatory cytokines during *H. pylori* infection of mice. Mice were infected with 5×10^7 CFU of parental (G27) or *imc1** (adapted allele in parental background) *H. pylori* or mock infected and sacrificed after 24 hours, 1 week, or 2 weeks. Total RNA was extracted from the mouse stomach and cytokine gene expression was measured by one-step qRT-PCR. Cytokine transcript levels were normalized to the levels of the endogenous control beta-2 microglobulin and expressed as fold change relative to the levels in mock-infected mice from the same timepoint. Relative transcript levels for each cytokine are plotted for G27-infected (filled) and *imc1**-infected (open) groups. 24 hour and 1 week data are from a single experiment, N = 5 mice per infection group (24 hours) or 8 mice per group (1 week). 2 week data are from a single experiment with 4 mice per group, except IL-4, which was measured in 2 experiments with a total of 13 mice per group, and MIP-2, which was measured in 3 experiments with a total of 21 mice per group.

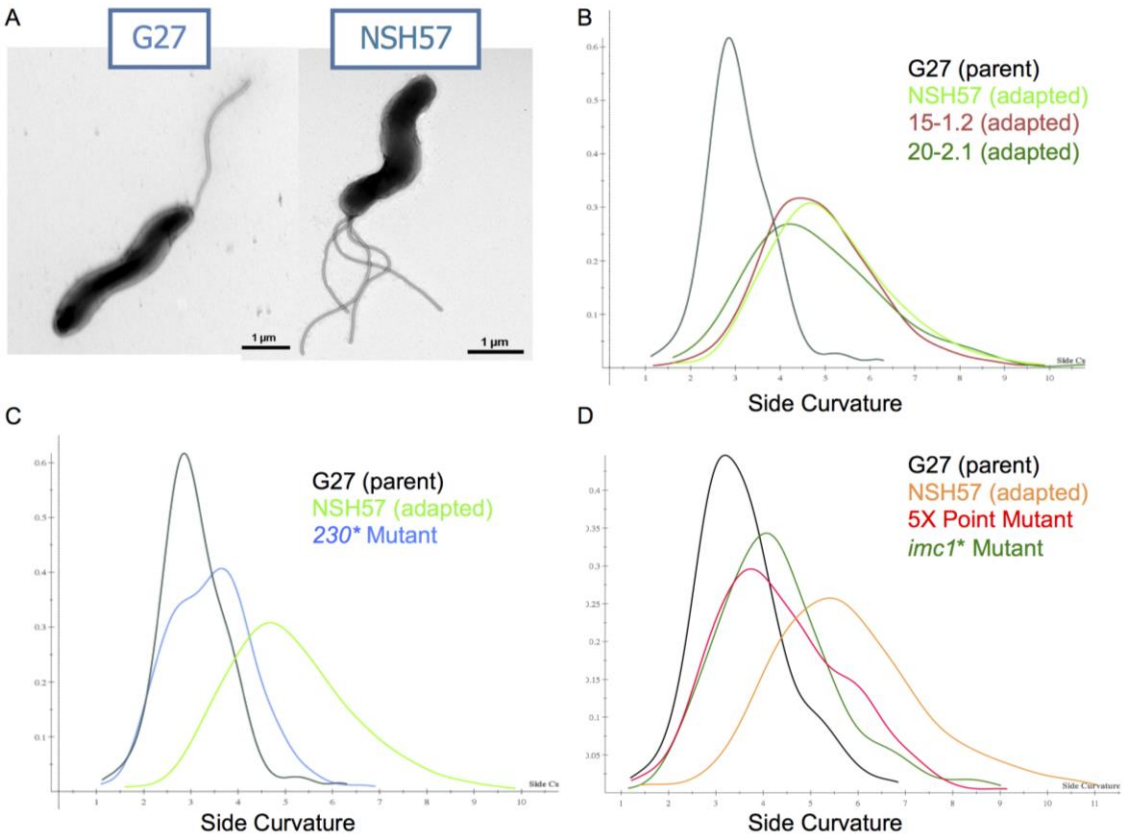

Figure 3.12 The identified mutations do not fully explain the increased helical side curvature of mouse-adapted strains. (A) Electron micrographs of parental strain G27 (left) and mouse-adapted strain NSH57 (right). (B, C, D) Smooth histograms displaying the cell curvature (x-axis) as a density function (y-axis). Curvature is measured using CellTool software on thresholded phase contrast images obtained with a 100X objective, with two biological replicates of at least 100 cells each measured per strain. (B) Mouse-adapted strains NSH57, 15-1.2, and 20-2.1 (independent isolates derived from the same mouse passaging experiment) display increased side curvature compared to parental strain G27. (C) Mutation to the adapted allele of gene *HPG27_230*, encoding the peptidoglycan recycling factor OppD, in the G27 background does not recapitulate the increased curvature of NSH57. (D) Mutation to the NSH57 allele of either *imc1*, encoding a novel cell-envelope-associated protein, or all five genes found to be mutated in NSH57 (*HPG27_230, 557, 1188, fliM,* and *imc1*) in the G27 background does not recapitulate the increased curvature of NSH57. Images in (A) courtesy of Laura Sycuro, personal communication.

Table 3.1 Bacteria with a mutation in *imc1* were recovered from the majority of mice infected with a pool of four mutants.

Mouse	***230* Mutant**	***557* Mutant**	***fliM* Mutant**	***imc1* Mutant**
1				✔
2	✔	✔		
3				✔
4				✔
5	✔			✔
6	✔			✔
7				✔
8	✔			✔
9		✔		
10		✔		
11	✔			✔
12				✔
13	✔	✔		✔
14	✔			
15				✔
	7 of 15	4 of 15	0 of 15	11 of 15

Table 3.2 Proteins identified by affinity chromatography and mass spectrometry as interacting specifically with Imc1-FLAG[a].

	Run 1: 30-40 kDa		Run 1: 40-50 kDa		Run 2		
Gene Number	**Imc1-FLAG**	**G27**	**Imc1-FLAG**	**G27**	**Imc1-FLAG**	**G27**	**Function**
HPG27_792	24	0	22	1	17	0	Inhibitor of mouse colonization Imc1
HPG27_230	-	-	26	0	3	1	Oligopeptide transport ATP-binding protein OppD
HPG27_268	-	-	8	0	-	-	Putative vacuolating cytotoxin (VacA)-like protein
HPG27_277	-	-	18	2	5	0	Periplasmic dipeptide binding protein HbpA/DppA
HPG27_475	9	0	-	-	-	-	ATP-dependent hsl protease ATP-binding subunit
HPG27_590	-	-	35	3	18	9	Hypothetical protein
HPG27_677	-	-	22	0	-	-	Outer membrane protein SabB
HPG27_689	7	0	-	-	-	-	Hypothetical protein
HPG27_924	6	0	-	-	-	-	Peptidyl-prolyl cis-trans isomerase
HPG27_958	-	-	15	0	-	-	Potassium efflux
HPG27_1156	-	-	-	-	5	0	Hypothetical protein
HPG27_1413	-	-	6	0	-	-	Putative integral membrane protein; TlyC-like hemolysin domain

[a] Numbers indicate number of peptides by which the corresponding protein was identified in a given sample

Table 3.3 Strains used in this chapter.

Strain ID	Strain Name	Genotype	Reference	Derived from	Markers	Oligos Used (See T 3.4)	Genomic Positions Affected
G27_O	G27 Original	wild-type G27	(99)	endoscopy patient, Italy	none	N/A	N/A
G27_15-1.2	G27 15-1.2	mouse adapted from G27	this work	G27	none	N/A	N/A
G27_20-2.1	G27 20-2.1	mouse adapted from G27	this work	G27	none	N/A	N/A
NSH57	NSH57	mouse adapted from G27	(59)	G27	none	N/A	N/A
IEC 60	G27-cat	G27 McGee::cat	this work, (38)	MSD132cat (TSH3)	cmR	N/A (genomic DNA used)	204322 to 204768 replaced w/cat
TSH 4	NSH57-cat	NSH57 McGee::cat	this work, (38)	MSD132cat (TSH3)	cmR	N/A (genomic DNA used)	204322 to 204768 replaced w/cat
IEC 2	Δ230	G27 230::catsacB	this work	G27 original	cmR, sucS	oIEC97-100	253555 to 255561 replaced w/catsacB
IEC 25	230*	230*, IEC2::NSH57 230	this work	IEC2	none	oIEC97, 98	253258 to 256017 replaced w/new allele
IEC 34	ΔfliM	G27 fliM::catsacB	this work	G27 original	cmR, sucS	oIEC93-96	426929 to 428229 replaced w/catsacB
IEC 37	Δ557	G27 557::catsacB	this work	G27 original	cmR, sucS	oIEC112-115	598410 to 598483 replaced w/catsacB
IEC 38	Δ792	G27 792::catsacB	this work	G27 original	cmR, sucS	oIEC117-120	856667 to 856845 replaced w/catsacB
IEC 39	fliM*	fliM*, IEC34:: NSH57	this work	IEC34	none	oIEC93, 94	426587 to 426587

		fliM					replaced w/new allele
IEC 40	557*	557*, IEC37::NSH57 557	this work	IEC37	none	oIEC112, 116	598028 to 599268 replaced w/new allele
IEC 41	792*	792*, IEC38::NSH57 792	this work	IEC38	none	oIEC117, 118	856173 to 857262 replaced w/new allele
IEC 43		230* ΔfliM, IEC25 fliM::catsacB	this work	IEC25	cmR, sucS	oIEC93-96	426929 to 428229 replaced w/catsacB
IEC 44		230*fliM*, IEC43::NSH57 fliM	this work	IEC43	cmS, sucR	oIEC93, 94	426587 to 426587 replaced w/new allele
IEC 45		230*fliM* Δ557, IEC44 557::catsacB	this work	IEC44	cmR	oIEC112-115	598410 to 598483 replaced w/catsacB
IEC 46	Δ1188	G27 1188::catsacB	this work	G27 original	cmR, sucS	oIEC159-162	1307417 to 1308089 replaced w/catsacB
IEC 47		230*fliM*557*, IEC45::NSH57 557	this work	IEC45	cmS, sucR	oIEC112, 116	598028 to 599268 replaced w/new allele
IEC 48	1188*	1188*, IEC46::NSH57 1188	this work	IEC46	cmS, sucR	oIEC159, 160	1307010 to 1308295 replaced w/new allele
IEC 53		230*fliM*557* Δ792, IEC47 792::catsacB	this work	IEC47	cmR, sucS	oIEC117-120	856667 to 856845 replaced w/catsacB
IEC 54		230*fliM*557*792*, IEC53::NSH57 792	this work	IEC53	cmS, sucR	oIEC117, 118	856173 to 857262 replaced w/new allele
IEC 55		230*fliM*557*792*	this work	IEC54	cmR,	oIEC159-162	1307417 to

		Δ1188, IEC54 1188::catsacB			sucS		1308089 replaced w/catsacB
IEC 56	5X Mutant	230*fliM*557*792*1188*, IEC55::NSH57 1188	this work	IEC55	cmS, sucR	oIEC159, 160	1307010 to 1308295 replaced w/new allele
IEC 67	792-FLAG	FLAG tag 792 full-length	this work	G27	cmR	oIEC183, 185-187	insert 2X FLAG and cat at 856976
IEC 68	792*-FLAG	FLAG tag 792 truncated	this work	NSH57	cmR	oIEC188-191	insert 2X FLAG and cat at 856793
IEC 75		230*fliM*557*Δ1188, IEC47 1188::catsacB	this work	IEC47	cmR, sucS	oIEC159-162	1307417 to 1308089 replaced w/catsacB
IEC 76	4X Mutant	230*fliM*557*1188*, IEC75::NSH57 1188	this work	IEC75	none	oIEC159, 160	1307010 to 1308295 replaced w/new allele

Table 3.4 Oligonucleotides used in this chapter.

Oligo ID	Oligo Name	Oligo Sequence[a]	Targeted Gene	Reference (if app.)	Purpose of Oligo
oIEC93	fliM_1	AGCCTAGCGATGCGTATTTAGCAGA	fliM		Mutant construction primers
oIEC94	fliM_2	TGGCGCTCAAAAGCCTCCGT	fliM		Mutant construction primers
oIEC95	fliM_3	atccacttttcaatctatatcTCCACTCCTAAGGATTTACGCACCT	fliM		Mutant construction primers
oIEC96	fliM_4	caaaagaaaatgccgatatccTGATGCTAGGAGGTGAGGGGGC	fliM		Mutant construction primers
oIEC97	230_1	TCATGCTCACTCCAGGCGGGG	G27_230		Mutant construction primers
oIEC98	230_2	TGGAGCGCGATCCCCATGC	G27_230		Mutant construction primers
oIEC99	230_3	atccacttttcaatctatatcGGCATGCTTAAAAACGCGCCCA	G27_230		Mutant construction primers
oIEC100	230_4	caaaagaaaatgccgatatccAAGCATCGCGCACCCCTTCG	G27_230		Mutant construction primers
oIEC110	catsacF_3	GATATAGATTGAAAAGTGGAT	catsacB cassette		catsacB cassette primers with overhangs
oIEC111	catcasR_4	GGATATCGGCATTTTCTTTTG	catsacB cassette		catsacB cassette primers with overhangs
oIEC112	new557_1	TGGGGTAAGAGCCCACCAGAGC	G27_557		Mutant construction primers
oIEC113	new557_2	AGCCACAAGGATACTCACCGCT	G27_557		Mutant construction primers
oIEC114	new557_3	atccacttttcaatctatatcTTGGGCAAGTGCAAGCCTGCA	G27_557		Mutant construction primers
oIEC115	new557_4	caaaagaaaatgccgatatccCGCAAGGGGTGGATAGCTCGC	G27_557		Mutant construction primers
oIEC116	new557_2'	GCACAGGGAGTTTGGCTCGCA	G27_557		Mutant construction

				primers
oIEC117	new792_1	ACTGATGGCTTTATGCTGGCTGTC	G27_792	Mutant construction primers
oIEC118	new792_2	AGCGTCTTTAGCCATGCCCCC	G27_792	Mutant construction primers
oIEC119	new792_3	atccacttttcaatctatatcCCGCCCCTATACCAAAACGATCCC	G27_792	Mutant construction primers
oIEC120	new792_4	caaaagaaaatgccgatatccGCTTGGGCGTTTAGAGCCGTCA	G27_792	Mutant construction primers
oIEC159	1188_1	CGCACATTGTCCCATTGCCGC	G27_1188	Mutant construction primers
oIEC160	1188_2	TCCCCCTTAAAACCACCCAAGCA	G27_1188	Mutant construction primers
oIEC161	1188_3	atccacttttcaatctatatcAGCCAGGACATGGTGGGCCT	G27_1188	Mutant construction primers
oIEC162	1188_4	caaaagaaaatgccgatatccGCGGTAAAAATGGGGGTTTAGTTGC	G27_1188	Mutant construction primers
oIEC163	230_B	AATTGGTTTCTACTATCTCGCCTTTTTCTT	G27_230	PCR genotyping: tetra-primer ARMS; T allele
oIEC164	230_C	CGCTTGGCCGATAGGGTTTATGTACT	G27_230	PCR genotyping: tetra-primer ARMS; A allele
oIEC165	230_A	AATCAAGAGCTTTGAATATTCGTGCTTG	G27_230	PCR genotyping: tetra-primer ARMS; outer F
oIEC166	230_D	ATCCAAAACCAGATTTTAGACTTGCTCA	G27_230	PCR genotyping: tetra-primer ARMS; outer R
oIEC167	557_B	ATTTCTTTACGCAAGCCTTTGGGTAT	G27_557	PCR genotyping: tetra-primer ARMS: T allele
oIEC168	557_C	AGCCTTCCTTAAAAAGAAAGTTTGCTG	G27_557	PCR genotyping: tetra-primer ARMS; C allele
oIEC169	557_A	CGAAGGGTTGCATAAAATGTCTTTAGAC	G27_557	PCR genotyping: tetra-primer ARMS; outer F
oIEC170	557_D	ACTTCTAACAACAATATTGACGCCTGGT	G27_557	PCR genotyping: tetra-primer ARMS; outer R

oIEC171	792_B	TCTTCTGTAGGCTGATACACCCTA	G27_792	PCR genotyping: tetra-primer ARMS; A allele
oIEC172	792_C	GCGAAACGATTTTTTTACAACAC	G27_792	PCR genotyping: tetra-primer ARMS; G allele
oIEC173	792_A	GTAGAAATCTTGTTGGGAGTATTCAAT	G27_792	PCR genotyping: tetra-primer ARMS; outer F
oIEC174	792_D	TAGTCAAAAAAATAAGCGTGGTAGTAGT	G27_792	PCR genotyping: tetra-primer ARMS; outer R
oIEC175	1188_B	GTATTATAGCATAAAGCGTTTTTTTTTTTGTC	G27_1188	PCR genotyping: tetra-primer ARMS; C allele
oIEC176	1188_C	AAAATGACAAAATTTTCCCCAAAGGA	G27_1188	PCR genotyping: tetra-primer ARMS; T allele
oIEC177	1188_A	GGATTTTTTTCATGTTTTCTCCTTTTTG	G27_1188	PCR genotyping: tetra-primer ARMS; outer F
oIEC178	1188_D	GGGAATGTTGTTTTAAAATACCCCCTTA	G27_1188	PCR genotyping: tetra-primer ARMS; outer R
oIEC179	fliM_B	CTAATCGTGTGAGTAAGGAGCAATGGT	fliM	PCR genotyping: tetra-primer ARMS; T allele
oIEC180	fliM_C	CATTTTGTCATGGATACTCCTAAAAGATCG	fliM	PCR genotyping: tetra-primer ARMS; C allele
oIEC181	fliM_A	GCTTTTAGAAGTCGTTGATGAAAATGTG	fliM	PCR genotyping: tetra-primer ARMS; outer F
oIEC182	fliM_D	GCTTCATGGAAAAGACATTAAAACTCGT	fliM	PCR genotyping: tetra-primer ARMS; outer R
oIEC183	F1G27792whole	TGACAGCGATTTGAGTTTAAG	G27_792	FLAG tag HPG27_792 full length
oIEC184	CatFLAGUp	GATTATAAAGACGATGACGATAAGGATTATAAAG		FLAG tag HPG27_792
oIEC185	F4G27792dswho	CCCAGTTTGTCGCACTGATAAGGCGTTTATTACGACAATTTA	G27_792	FLAG tag HPG27_792 full length
oIEC186	R3G27792whole	CTTTATAATCCTTATCGTCATCGTCTTTATAATCACTTTCCTTAAGACATTCTAA	G27_792	FLAG tag HPG27_792 full length

oIEC187	R2G27792dswhole	ATTCTTGAACAACAACTGGCG	G27_792		FLAG tag HPG27_792 full length
oIEC188	F1G27792trunc	CTTAAATAAGGGTCTGGCATT	G27_792		FLAG tag HPG27_792 truncated
oIEC189	F4G27792trunc	CCCAGTTTGTCGCACTGATAAAGATGAAAACGATTCTTTTAA	G27_792		FLAG tag HPG27_792 truncated
oIEC190	R3G27792trunc	CTTTATAATCCTTATCGTCATCGTCTTTATAATCGTCAATTGGGCGAAAGAATCT	G27_792		FLAG tag HPG27_792 truncated
oIEC191	R2G27792trunc	CAGCCATTGATCCCCTTCTTT	G27_792		FLAG tag HPG27_792 truncated
oIEC213	mIFNg-F	GATCCTTTGGACCCTCTGACTT	IFN-γ	(126)	qRT-PCR mouse cytokines
oIEC214	mIFNg-R	TGACTGTGCCGTGGCAGTAA	IFN-γ	(126)	qRT-PCR mouse cytokines
oIEC215	mIFNg-ProbeFAM/TAMRA	CAATGAACGCTACACACTGCATCTTGGC	IFN-γ	(126)	qRT-PCR mouse cytokines
oIEC216	mIL4-F	GAGCTGCAGAGACTCTTTCG	IL-4	(126)	qRT-PCR mouse cytokines
oIEC217	mIL4-R	ACTCATTCATGGTGCAGCTTA	IL-4	(126)	qRT-PCR mouse cytokines
oIEC218	mIL4-ProbeFAM/TAMRA	GCTTTTCGATGCCTGGATTCATCG	IL-4	(126)	qRT-PCR mouse cytokines
oIEC219	mIL17-F	GAATGTGAAGGTCAACCTCAAAGTC	IL-17	(126)	qRT-PCR mouse cytokines
oIEC220	mIL17-R	TTCCCAGATCACAGAGGGATATCT	IL-17	(126)	qRT-PCR mouse cytokines
oIEC221	mIL17-ProbeFAM/TAMRA	TCAACCGTTCCACGTCACCCTGG	IL-17	(126)	qRT-PCR mouse cytokines
oIEC222	mTNFa-F	ACAAGGCTGCCCCGACTAC	TNF-α	(128)	qRT-PCR mouse

					cytokines
oIEC223	mTNFa-R	CGCAGAGAGGAGGTTGACTT	TNF-α	(128)	qRT-PCR mouse cytokines
oIEC224	mTNFa-ProbeFAM/TAMRA	CCTCACCCACACCGTCAGCCG	TNF-α	(128)	qRT-PCR mouse cytokines
oIEC225	mMIP2-F	AGCTTGAGTGTGACGCCCC	MIP-2 (CXCL2)	(125)	qRT-PCR mouse cytokines
oIEC226	mMIP2-R	TTGACCGCCCTTGAGAGTG	MIP-2 (CXCL2)	(125)	qRT-PCR mouse cytokines
oIEC227	mMIP2-ProbeFAM/TAMRA	CCCCACTGCGCCCAGACAGAAGT	MIP-2 (CXCL2)	(125)	qRT-PCR mouse cytokines
oIEC228	mKC-F	GTGTCTAGTTGGTAGCCCATAAT	KC (CXCL1)	(125)	qRT-PCR mouse cytokines
oIEC229	mKC-R	TGTAACAGTCCTTTGAACGTCTC	KC (CXCL1)	(125)	qRT-PCR mouse cytokines
oIEC230	mKC-ProbeFAM/TAMRA	CCGAGCGAGACGAGACCAGGA	KC (CXCL1)	(125)	qRT-PCR mouse cytokines
oIEC231	mIL10-F	GTTGCCAAGCCTTATCGGAA	IL-10	(127)	qRT-PCR mouse cytokines
oIEC232	mIL10-R	CCGCATCCTGAGGGTCTTC	IL-10	(127)	qRT-PCR mouse cytokines
oIEC233	mIL10ProbeFAM/TAMRA	CAGTTTTACCTGGTAGAAGTGATGCCCCAGG	IL-10	(127)	qRT-PCR mouse cytokines
oIEC234	mB2M-F	CCTGCAGAGTTAAGCATGCCAG	beta2-microglobulin	(129)	qRT-PCR mouse cytokines
oIEC235	mB2M-R	TGCTTGATCACATGTCTCGATCC	beta2-microglobulin	(129)	qRT-PCR mouse cytokines
oIEC236	mB2M-ProbeFAM/TAMRA	TGGCCGAGCCCAAGACCGTCTAC	beta2-microglobulin	(129)	qRT-PCR mouse cytokines

oIEC237	mEla2-F	GTAGTGCTGGGAGCCCATGAC	neutrophil elastase	(130)	qRT-PCR mouse cytokines
oIEC238	mEla2-R	ACATGGAGTTCTGTCACCCA	neutrophil elastase	(130)	qRT-PCR mouse cytokines
oIEC239	mEla2-P/FAM/ TAMRA	CCAACGTGCAGGTGGCCCAG	neutrophil elastase	(130)	qRT-PCR mouse cytokines
oIEC240	mB2M-ProbeHEX/ TAMRA	TGGCCGAGCCCAAGACCGTCTAC	beta2-microglobulin	(129)	qRT-PCR mouse cytokines

[a]: lowercase letters indicate common overhangs with homology for stitching to the *catsacB* cassette.

CHAPTER 4

Characterization of later stages of mouse adaptation

INTRODUCTION

Chapter 3 characterized the mutations involved in adaptation of an *H. pylori* clinical isolate to become a mouse-colonizing strain. Further adaptation occurred with additional passaging in mice, resulting in increased loads and persistence for longer periods in the mouse stomach (38). A number of mutations specific to these later stages of adaptation have been identified, and characterization of their functional roles can provide additional insights into the mechanisms by which *H. pylori* adapts to new hosts.

The mouse-colonizing strain described in Chapter 3, NSH57, survives in the mouse stomach for 1 to 4 weeks before being cleared, and typically results in bacterial loads of approximately 100,000 CFU per gram of stomach tissue after a one-week infection (38), see Figs. 3.4 and 3.5. This strain was serially passaged for an additional six rounds in mice, with infection durations of 4-8 weeks per round (see Fig. 1.2 A). Strains derived from this process exhibited progressively increased loads and longer persistence in the mouse stomach (see Fig. 1.2 B). The final strain, MSD132, has been shown to persist in the mouse stomach for at least 7 months (38).

A subset of *H. pylori* strains harbor a type IV secretion system (T4SS) encoded by the *cag* pathogenicity island (Cag PAI), which functions to deliver bacterial products into host cells (71). Both a protein effector, CagA, and bacterial peptidoglycan enter host cells, resulting in wide-ranging consequences, including stimulation of inflammatory signaling and effects on cell growth and polarity (68, 71). In humans, infection with an *H. pylori* strain carrying a functional

Cag T4SS is associated with more severe disease outcomes, including higher rates of gastric cancer, possibly due to increased inflammation and/or direct effects on gastric cells (1).

Previous studies have demonstrated that *H. pylori* strains that lack a functional Cag T4SS often colonize mice to higher loads and/or for longer time periods than unrelated Cag^+ strains (58, 113). Additionally, loss of Cag T4SS function of initially Cag^+ *H. pylori* strains has been observed during the course of infection in mice. This has occurred via gene deletion (112, 114) as well as through point mutations that result in decreased stimulation of inflammatory responses by adapted bacteria despite the continued presence of the *cag* genes (113).

Although the mouse-colonizing strain NSH57 has been previously demonstrated to have an active Cag T4SS (59), this function was lost in the strains selected for persistent infection of mice (38). Whole genome sequencing of mouse-persistence strains revealed a single base insertion in a gene of the T4SS-encoding *cag* PAI, *cagY* (see Table 2.4). CagY is a surface-exposed component of the T4SS machinery (26) and has been demonstrated to be essential for effector delivery (131).

This chapter describes the characterization of the effects of the identified mutation in *cagY*, as well as other mutations occurring during the later stages of *H. pylori* murine adaptation that may contribute to the ability of these strains to persistently infect the mouse stomach.

RESULTS

The Cag T4SS is non-functional in strains selected for persistence in mice

Cag T4SS function was assessed by measurement of two downstream effects on gastric epithelial cells cocultured with *H. pylori*. Mouse-persistence strain MSD85 lost both the ability to induce IL-8 production (Fig. 4.1 A) and to translocate the bacterial secreted effector CagA for phosphorylation by host Src family kinases (Fig. 4.1 B). Both of these phenotypes are dependent

on a functional Cag T4SS for entry of either peptidoglycan (for Nod1-mediated IL-8 production) or CagA into the gastric epithelial cell cytosol. Mouse-colonizing strain NSH57 and mouse-persistence strain MSD86, isolated from the same mouse as strain MSD85 during serial passaging, both retain the capacity to induce IL-8 and translocate CagA (Fig. 4.1). Despite differences in phosphorylation of CagA, which occurs in the host cell after translocation, all of the tested *H. pylori* strains produce CagA protein (Fig. 4.1 B, top), suggesting a defect in the translocation machinery in strain MSD85.

A frameshift mutation in *cagY* causes loss of T4SS function

A single base insertion in the gene *cagY,* which encodes a surface-exposed component of the Cag secretion system complex (26), was identified in strain MSD85 but was absent in strain MSD86 (see Table 2.4). This mutation was also present in the strains subjected to further rounds of serial passaging in mice, MSD101, 124, 132A, and 132B, all of which were derived from strain MSD85. The mutation, insertion of a T at genomic position 525,134, creates a stop codon corresponding to amino acid 462, of 1903 total amino acids. This truncation results in loss of the C-terminal VirB10-like domain (132), believed to be essential for function (Fig. 4.2 A).

Because CagY is essential for Cag T4SS function (131) and because the presence of this mutation correlated with loss of T4SS activity in mouse-persistence strains, its functional role was directly tested using isogenic mutant strains. The MSD85 (truncated) allele of *cagY* was introduced into NSH57 by allelic exchange (see Fig. 3.1); this mutant is referred to as *cagY** (Table 4.1).

Both the *cagY** mutant and a strain with a *catsacB* insertion-deletion cassette in *cagY* (Δ*cagY*) demonstrated reduced induction of IL-8 (Fig. 4.2 B) and no translocation of CagA for phosphorylation (Fig. 4.2 C), similar to strain MSD85 and in contrast to strain NSH57, despite

these mutants being made in the NSH57 strain background. This indicates that the identified single base insertion in *cagY* is sufficient to cause loss of Cag T4SS function in *H. pylori*.

Additional mutations also contribute to loss of Cag T4SS function in mouse-persistence strains

To assess whether the identified single base mutation in *cagY* is the sole cause of loss of Cag T4SS function, the MSD85 allele was reverted to the NSH57 genotype (*cagY#*). NSH57 does not harbor the single base insertion in *cagY* and retains Cag T4SS function. If this mutation is the only defect in the *cag* PAI, reversion to the NSH57 genotype would be expected to restore Cag T4SS function, despite the MSD85 strain background of the *cagY#* mutant. However, as shown in Fig. 4.3, the *cagY#* revertant retains the MSD85 phenotype of inability to induce IL-8 production by cocultured gastric epithelial cells.

No other coding mutations in *cag* PAI genes, either within *cagY* or in other genes, were identified in the whole genome sequence data (see Table 2.4). However, *cagY* has two large regions of repetitive sequence (Fig. 4.2 A), which hamper mapping of both short Illumina reads and longer 454 reads, as described in Chapter 2. To more thoroughly search for any additional mutations in *cagY*, Sanger capillary sequencing was performed with a series of primers to span this large (> 5000 bp) gene (cagYA-O, Table 4.2). No additional base changes between strains NSH57 (Cag T4SS functional) and MSD85 (Cag T4SS non-functional) were identified, although even with Sanger read lengths approaching 1000 bp, the repetitive regions of *cagY* were still not fully covered, with a gap remaining in Repeat 2. PCR using primers flanking the two repeat regions of *cagY* identified a slight contraction of Repeat 2 in NSH57 and all derived strains (Fig. 4.4). Because this change is present in NSH57, it cannot explain the phenotypic difference between NSH57 and MSD85. No change in the length of Repeat 1 was observed.

A mutation in the outer membrane protein *hopH* was also identified in mouse-persistence strains (see Table 2.4) and could cause altered interactions with host cells, possibly resulting in effects on Cag T4SS function. Reversion of this mutation to the NSH57 genotype in both the MSD85 background (*hopH#*) and in the *cagY* reversion background (*cagY#*/*hopH#*) did not restore Cag T4SS function, as measured by IL-8 production by cocultured gastric epithelial cells (Fig. 4.3, right). The additional mutation or mutations present in MSD85 that contribute to loss of Cag T4SS function remain to be identified.

Allelic exchange was used to generate isogenic mutants with the MSD85 alleles of genes *hopC, hopH, dnaE, HPG27_676,* and *HPG27_1024* in the NSH57 strain background (Table 4.1) for use in future phenotypic characterization studies. An additional mutant with the entire *cagY* gene of MSD85 replaced with the entire *cagY* gene of NSH57 (MSD85 *wcagY#*) has also been generated for elucidation of the roles of additional mutations in *cagY* versus elsewhere in the *cag* PAI.

DISCUSSION

H. pylori isolates selected for their ability to persist in mice for weeks to months lost Cag T4SS function, as measured by inability to induce IL-8 production by cocultured human gastric epithelial cells or to translocate the bacterial protein CagA into host cells. This phenotype was found to correlate with the acquisition of a single base insertion, predicted to cause loss of an essential domain of the protein, in the gene encoding T4SS component CagY. Introduction of this mutation alone into strain NSH57, which has a functional T4SS, caused loss of T4SS function.

There remains one (or more) additional unidentified mutation causing loss of Cag T4SS function in strain MSD85, as reversion of the identified single base insertion to the NSH57

genotype failed to restore Cag T4SS function in the MSD85 background. This mutation could be within *cagY*, as part of the Repeat 2 region could be not be sequenced despite extensive efforts, or could be in another gene of the *cag* pathogenicity island that was missed in Illumina whole genome sequencing due to insufficient coverage. Difficulty sequencing and assembling *cagY* has been previously reported (133).

Recombination between numerous and non-randomly distributed direct repeats in the *cagY* gene can lead to in-frame deletions that alter the size, and potentially the antigenicity, of this surface-exposed component of the Cag T4SS. The presence of direct repeats at certain sites may allow some degree of targeting of diversification to particular loci that are best suited for producing adaptive change (134). The resulting CagY protein size variation has been observed both in vitro and in vivo (26), and is consistent with the contraction of *cagY* Repeat 2 observed here (Fig. 4.4). This contraction is not thought to cause loss of CagY function, as it is present in strain NSH57, which retains Cag T4SS activity.

A functional Cag T4SS is immune-stimulatory, and immune pressure may contribute to its loss, both during murine adaptation (38, 112–114) and as has been seen during the course of human chronic infection (24, 32). The Cag T4SS facilitates the delivery of fragments of *H. pylori*'s peptidoglycan cell wall to the gastric epithelial cell cytosol. The cytosolic pattern recognition receptor Nod1 (Nucleotide-binding oligomerization domain-containing protein 1) recognizes the tri-peptide signature of Gram-negative peptidoglycan (human Nod1) or both tri- and tetra-peptide fragments (murine Nod1, (135)), leading to NF-κB activation and upregulation of IL-8 (94). IL-8 and its murine functional homologs MIP-2 (CXCL2) and KC (CXCL1) act as chemoattractants for neutrophils (115, 116), beginning the process of active gastritis. Infiltrating immune cells produce additional cytokines, such as interferon gamma (IFNγ), leading to a

predominantly T_H1-polarized immune response to *H. pylori* colonization (117). Inflammatory cytokines, including IFNγ, have been shown to be important for clearance of the related bacterium *H. felis* from mice (95) and reducing the induction of inflammatory mediators may help *H. pylori* avoid clearance from the mouse stomach and establish a chronic infection.

This work provides an additional example of the phenomenon of loss of Cag T4SS function during adaptation to mice, in this case via point mutation. However, the loss of Cag T4SS function described here did not occur until after substantial improvements in murine colonization potential had already been achieved, highlighting the important contributions of other, previously uncharacterized genes.

MATERIALS AND METHODS

Strains

Strains used are listed in Table 4.1. Isogenic mutant strains were constructed by allelic exchange (102). A *cat sacB* insertion cassette, conferring chloramphenicol resistance and sucrose sensitivity and flanked by regions of homology to the site of interest, was constructed using PCR with splicing by overlapping extension (SOEing) (118) using the primers indicated in Tables 4.1 and 4.2 for each gene of interest. This insertion cassette was introduced into parental strain NSH57 by natural transformation and replaced the portion of the targeted gene indicated for each strain (Table 4.1). The resulting gene disruption mutants were transformed with DNA amplified from adapted strain MSD85 using the primers indicated in Table 4.2 or with genomic DNA isolated from strain MSD85, and clones in which the *cat sacB* cassette had been replaced with the adapted allele of the gene of interest were selected based on sucrose resistance and chloramphenicol sensitivity. PCR and Sanger sequencing were used to confirm the allelic exchange. For reversion of MSD85 alleles to the NSH57 genotype, the *catsacB* insertion cassette

was transformed into strain MSD85 (or mutants made in MSD85), and then replaced with PCR product amplified from NSH57 or with NSH57 genomic DNA. Primers used are indicated in Table 4.2, and the allelic exchange was confirmed by PCR and Sanger sequencing.

Coculture with AGS gastric epithelial cells

Coculture of *H. pylori* with human AGS gastric epithelial cells at a multiplicity of infection of 10:1 and measurement of IL-8 release into supernatants by ELISA were performed as described in Chapter 3.

Detection of CagA translocation into host cells

Total and phosphorylated CagA were detected by immunoblotting as previously described (71). Briefly, after coculture of AGS cells with *H. pylori*, monolayers were washed twice with PBS and resuspended in Laemmli buffer (50 mM Tris pH 6.8, 2% SDS, 10% glycerol, 0.075% bromophenol blue, 5% β-mercaptoethanol). Samples were run on 6.5% SDS-PAGE gels and transferred to PVDF membranes using an iBlot system (Invitrogen). Goat anti-CagA primary antiserum bN-13 (Santa Cruz Biotechnology) was used at a dilution of 1:10,000 to detect CagA, followed by donkey anti-goat secondary antibody conjugated to horseradish peroxidase (Promega) at a dilution of 1:10,000. Tyrosine phosphorylation was detected using 4G10 mouse monoclonal antibody (Millipore) at a dilution of 1:1000 followed by goat anti-mouse secondary antibody conjugated to horseradish peroxidase (Jackson ImmunoResearch Laboratories) at a dilution of 1:10,000.

Sequencing of full *cagY* gene

PCR primers cagY A through cagY O (Table 4.2) were designed to target the *cagY* gene every 600 to 700 basepairs, with primers in both forward and reverse directions flanking regions of repetitive sequence. The entire *cagY* gene was amplified using primers cagY A and cagY K Rev

with Phusion high fidelity polymerase (NEB) to produce an approximately 5800 bp PCR product. Capillary sequencing using each primer (A-O) was performed by the FHCRC Genomics Shared Resource using ABI's BigDye terminator cycle sequencing reagent on an ABI 3730xl DNA analyzer (Applied Biosystems).

Detection of changes in *cagY* repeat length

Primers flanking Repeat 1 (cagY D and cagY H Rev) and Repeat 2 (cagY J and cagY K Rev) were used to amplify the entire repeat regions from strains G27, NSH57, MSD85, MSD86, MSD101, MSD124, MSD132A, and MSD132B. Size differences in PCR products were identified after agarose gel electrophoresis.

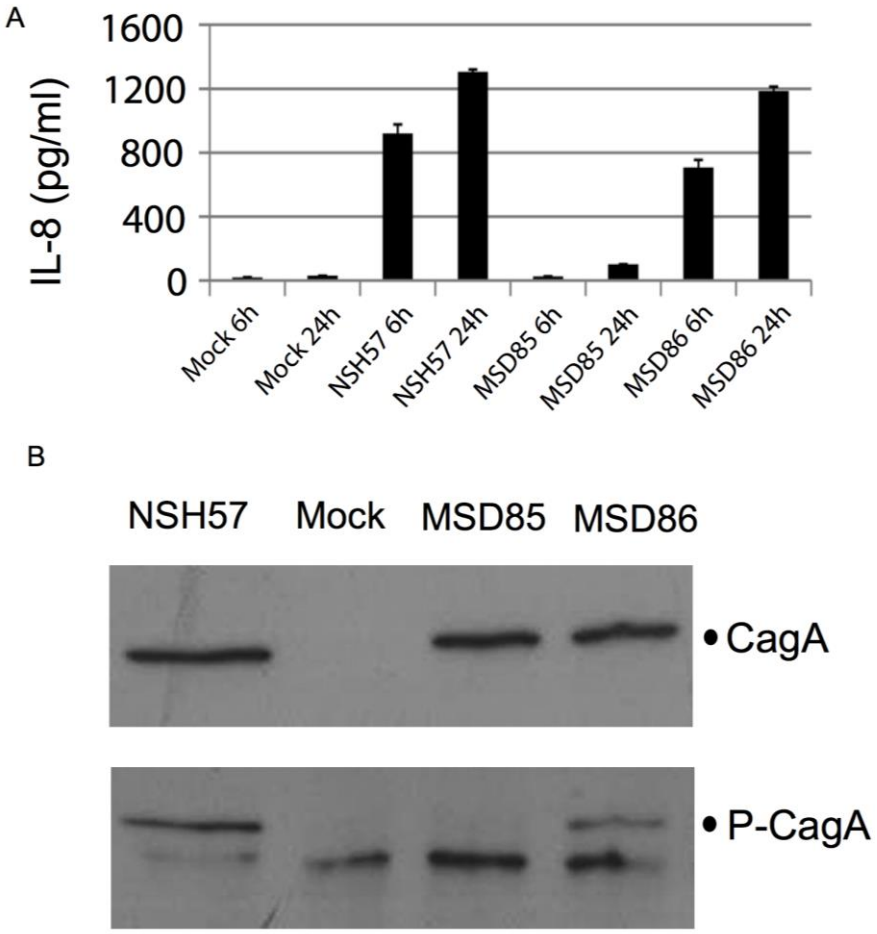

Figure 4.1 The Cag T4SS is non-functional in strains selected for persistence in mice. Human AGS gastric epithelial cells were co-cultured with the indicated *H. pylori* strains at an MOI of 10:1. (A) IL-8 release was measured by ELISA of the co-culture supernatant at the indicated timepoints. Strain MSD85 has lost the ability to induce IL-8 release, while strain MSD86 (isolated from the same round of passaging in mice) retains this ability. (B) Phosphorylation of the bacterial effector protein CagA after translocation into AGS cells, measured by Western blot after 24 hours of co-culture. MSD85 has lost the ability to translocate CagA, while MSD86 retains this ability. Adapted from (38).

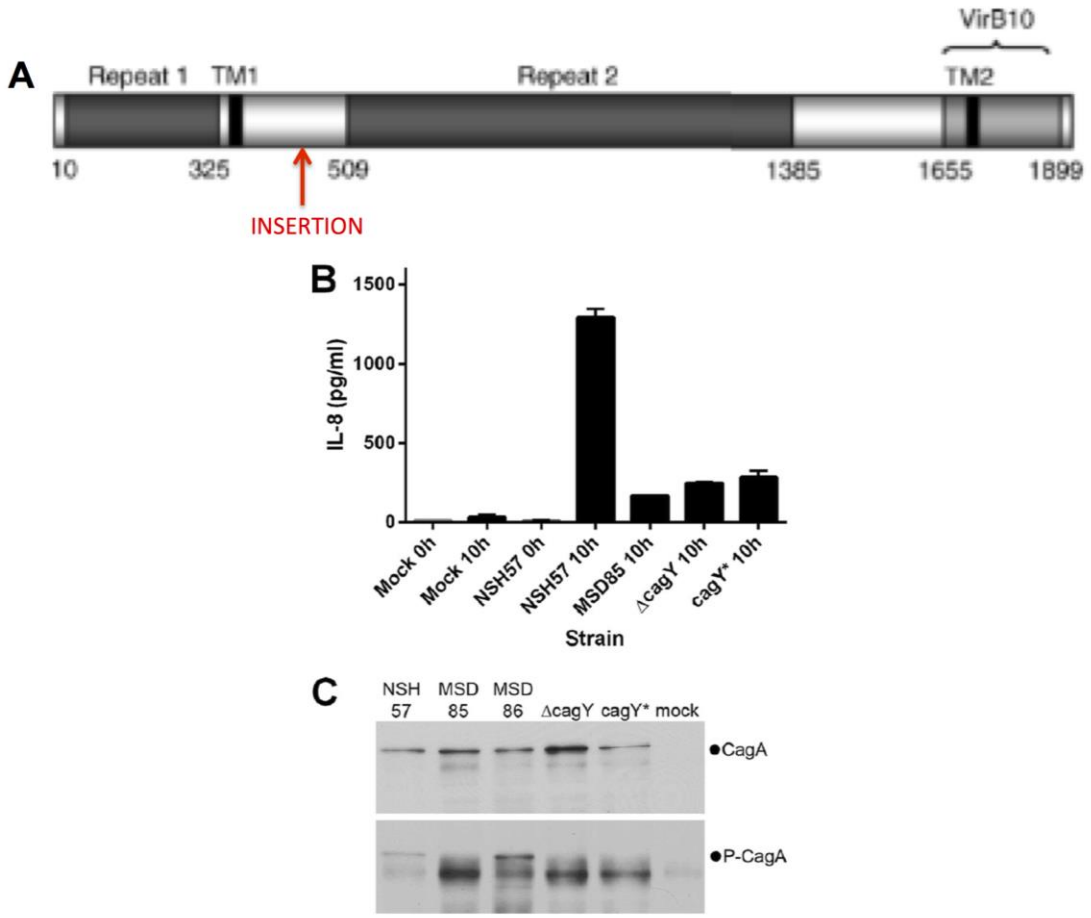

Figure 4.2 A frameshift mutation in *cagY* causes loss of Cag T4SS function. (A) Schematic of the domain structure of the CagY protein. The position of the single base insertion, leading to a frameshift, identified in strain MSD85 and descendants is indicated. (B, C) AGS gastric epithelial cells were co-cultured with the indicated *H. pylori* strains at an MOI of 10:1 for the indicated times. Isogenic mutants with either a *catsacB* insertion-deletion mutation in *cagY* or the MSD85 allele of *cagY* in the NSH57 genetic background (*cagY**) fail to (B) induce release of IL-8 or (C) translocate the bacterial effector protein CagA for phosphorylation by AGS cells. A adapted from (132); B and C reproduced with permission from (38).

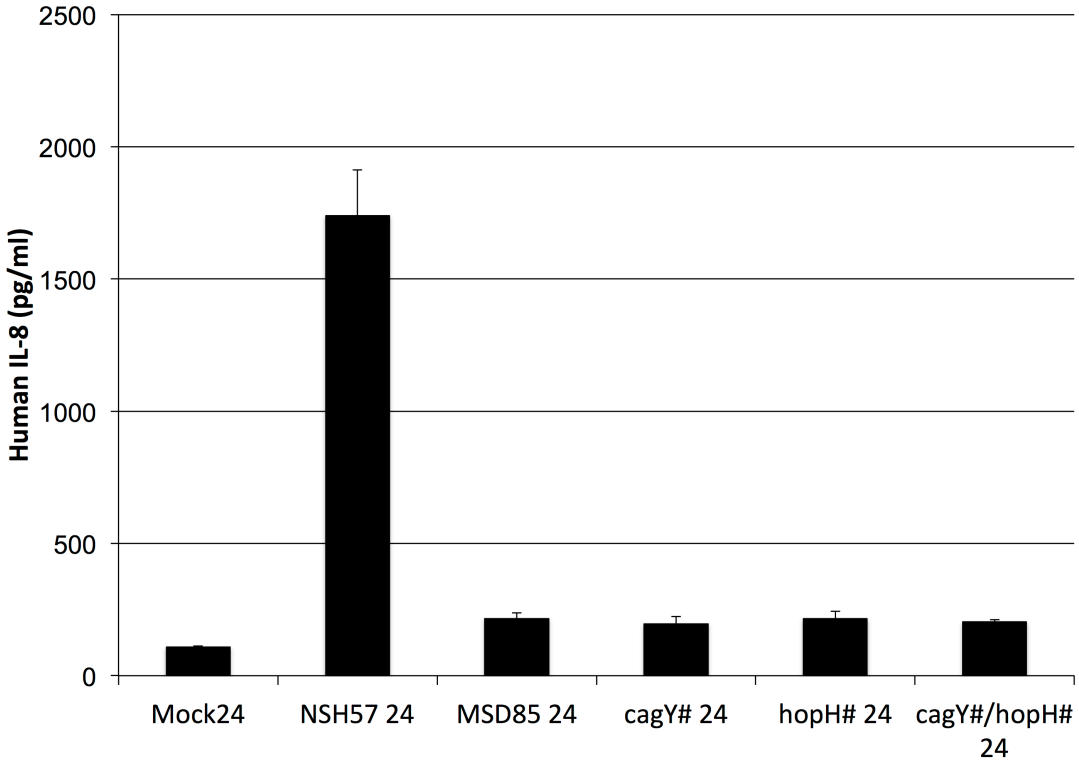

Figure 4.3 Additional, unidentified mutation(s) also contribute to loss of Cag T4SS function in strain selected for persistence in mice. Restoration of the MSD85 *cagY* allele to the NSH57 genotype (cagY#) does not restore the phenotype of IL-8 induction from co-cultured AGS cells. Restoration of the MSD85 *hopH* allele to the NSH57 genotype (hopH#), alone or in combination with restoration of the *cagY* allele (cagY#), also does not restore the IL-8 phenotype. Human AGS gastric epithelial cells were co-cultured with the indicated *H. pylori* strains at an MOI of 10:1 for 24 hours, and IL-8 release was measured by ELISA of the co-culture supernatant.

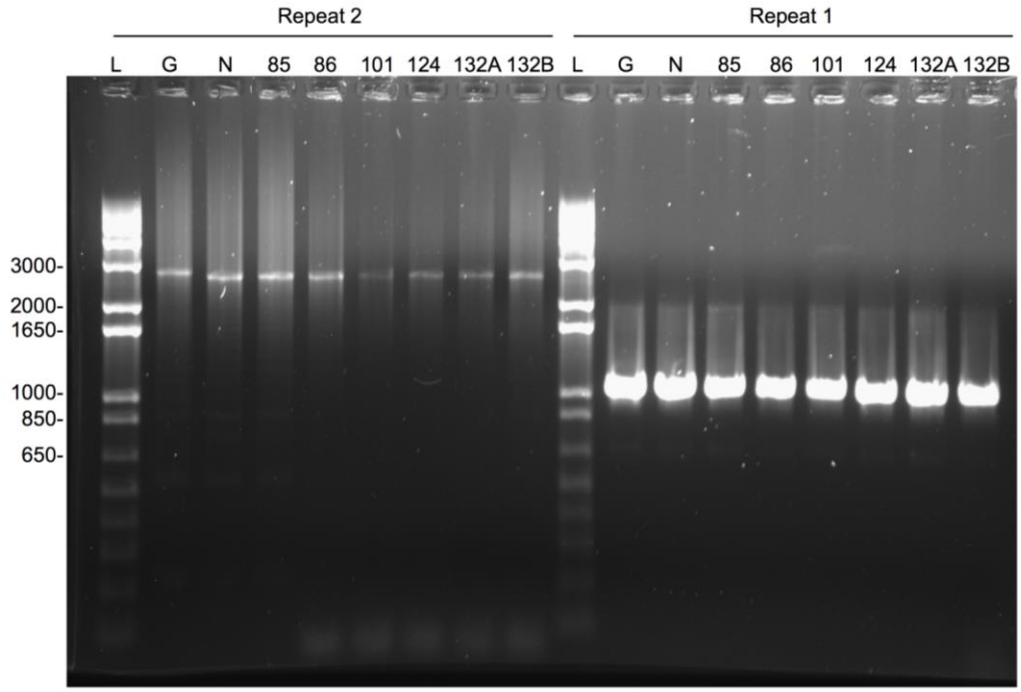

Figure 4.4 Contraction of Repeat 2 of *cagY* in mouse-adapted *H. pylori* strains. The Repeat 1 and Repeat 2 regions of *cagY* (see Fig. 4.2 A) were amplified using primers cagY D and cagY H Rev (Repeat 1) and cagY J and cagY K Rev (Repeat 2) from strains G27 (G), NSH57 (N), MSD85 (85), MSD86 (86), MSD101 (101), MSD124 (124), MSD132A (132A), and MSD132B (132B). Size differences in PCR products were identified after agarose gel electrophoresis, and a slight decrease in product size for Repeat 2 was observed for all strains except parental strain G27. No change was observed in the length of Repeat 1.

Table 4.1 Strains used in this chapter.

Strain ID	Strain Name	Genotype	Reference	Derived from	Markers	Oligos Used (See T 4.2)	Genomic Positions Affected
NSH57	NSH57	mouse adapted from G27	(59)	G27	none	N/A	N/A
MSD86	MSD86	mouse adapted from NSH57	(38)	NSH57	none	N/A	N/A
MSD85	MSD85	mouse adapted from NSH57	(38)	NSH57	none	N/A	N/A
MSD101	MSD101	mouse adapted from MSD85	(38)	MSD85	none	N/A	N/A
MSD124	MSD124	mouse adapted from MSD101	(38)	MSD101	none	N/A	N/A
MSD132A	MSD132A	mouse adapted from MSD124	(38)	MSD124	none	N/A	N/A
MSD132B	MSD132B	mouse adapted from MSD124	(38)	MSD124	none	N/A	N/A
IEC 7	ΔhopC	NSH57 hopC::catsacB	this work	NSH57	cmR, sucS	oIEC128-131	931829 to 932922 replaced w/catsacB
IEC 11	ΔdnaE	NSH57 dnaE::catsacB	this work	NSH57	cmR, sucS	oIEC123-126	1496715 to 1496734 replaced w/catsacB
IEC 13	Δ1024	NSH57 1024::catsacB	this work	NSH57	cmR, sucS	oIEC124-127	1127929 to 1128667 replaced w/catsacB
IEC 16	ΔhopH	NSH57 hopH::catsacB	this work	NSH57	cmR, sucS	oIEC140-143	652622 to 653309 replaced w/catsacB
IEC 22	Δ676	NSH57 676::catsacB	this work	NSH57	cmR, sucS	oIEC132-135	740732 to 741326 replaced w/catsacB
IEC 26	hopC*	hopC*, IEC7::MSD85 hopC	this work	IEC7	none	oIEC128, 129	931492 to 933423 replaced w/new allele
IEC 27	dnaE*	dnaE*, IEC11::MSD85 dnaE	this work	IEC11	none	oIEC127, 124	1495020 to 1497033 replaced w/new allele
IEC 28	1024*	1024*, IEC13::MSD86	this work	IEC13	none	oIEC124, 125	1127560 to 1129238

		1024					replaced w/new allele
IEC 29	hopH*	hopH*, IEC16::MSD85 hopH	this work	IEC16	none	oIEC140, 141	652296 to 653567 replaced w/new allele
IEC 30	676*	676*, IEC22::MSD85 676	this work	IEC22	none	oIEC132, 133	740368 to 741698 replaced w/new allele
IEC 42	ΔcagY	ΔcagY, NSH57 cagY::catsacB	Dorer 2013 IAI	NSH57	cmR, sucS	oIEC148-151	525113 to 525154 replaced w/catsacB
IEC 51	cagY*	cagY*, IEC42::MSD85 gDNA	Dorer 2013 IAI	IEC42	cmS, sucR	oIEC148, 149	524989 to 525379 replaced w/new allele
IEC 57	MSD85 ΔcagY	MSD85 ΔcagY::catsacB	this work	MSD85	cmR, sucS	oIEC148-151	525113 to 525154 replaced w/catsacB
IEC 61	MSD85 cagY*	MSD85cagY*, IEC57::NSH57 cagY	this work	IEC57	none	oIEC148, 149	524989 to 525379 replaced w/new allele
IEC 63	MSD85 hopH#	MSD85 hopH# (repaired to NSH57 allele)	this work	MSD85 ΔhopH, A	cmS, sucR	oIEC140-143	652296 to 653567 replaced w/new allele
IEC 64	cagY* hopH#	IEC61 hopH# (repaired to NSH57 allele)	this work	IEC61 ΔhopH, A	cmS, sucR	oIEC140-143	652296 to 653567 replaced w/new allele
IEC 69	ΔwcagY	MSD85 ΔwholecagY::catsacB	this work	MSD85	cmR, sucS	oIEC207-210	520327 to 526938 replaced w/catsacB
IEC 77	MSD85 wcagY#	IEC69::NSH57 gDNA (revert MSD85 whole cagY to NSH57)	this work	IEC69	cmS, sucR	N/A (gDNA)	520327 to 526938 replaced w/NSH57 allele

Table 4.2 Oligonucleotides used in this chapter.

Oligo ID	Oligo Name	Oligo Sequence[a]	Targeted Gene	Purpose of Oligo
oIEC110	catsacF_3	GATATAGATTGAAAAGTGGAT	catsacB cassette	catsacB cassette primers with overhangs
oIEC111	catcasR_4	GGATATCGGCATTTTCTTTTG	catsacB cassette	catsacB cassette primers with overhangs
oIEC123	dnaE_1	GCCTTTTCTTGCGCTCATGCCC	dnaE	Mutant construction primers
oIEC124	dnaE_2	TGACCACCCCAAAAGGCTGC	dnaE	Mutant construction primers
oIEC125	dnaE_3	atccacttttcaatctatatcTCGGCCCTTGCGCTTGTTCG	dnaE	Mutant construction primers
oIEC126	dnaE_4	caaaagaaaatgccgatatccTGGGGGCTGATCCCAAAGTGG	dnaE	Mutant construction primers
oIEC127	dnaE_1'	CCACACGATCAGCATATACCCTGGG	dnaE	Mutant construction primers
oIEC128	hopC_1	GGGCTATTACTGGCTCCCTAGCT	hopC	Mutant construction primers
oIEC129	hopC_2	GCCCCGCTATCACCGCTTGG	hopC	Mutant construction primers
oIEC130	hopC_3	atccacttttcaatctatatcTGGCGGCTACGGCAGAGTTG	hopC	Mutant construction primers
oIEC131	hopC_4	caaaagaaaatgccgatatccTTGCAGGCACAGGGGGCAAT	hopC	Mutant construction primers
oIEC132	676_1	TGAGAAGGAGTTTGGCTTTTTGCCT	G27_676	Mutant construction primers
oIEC133	676_2	GCCGATTTGATAGCCCGCGC	G27_676	Mutant construction primers
oIEC134	676_3	atccacttttcaatctatatcCGTATCGCCGCTGCAAACGC	G27_676	Mutant construction primers

oIEC135	676_4	caaaagaaaatgccgatatccGCAGAATCCTAAACGCTTTGAGCGT	G27_676	Mutant construction primers
oIEC140	hopH_1	GGCACTCCATACCCATAAGGCCC	hopH	Mutant construction primers
oIEC141	hopH_2	TGCAAACCAAGTGCTACCCCCT	hopH	Mutant construction primers
oIEC142	hopH_3	atccacttttcaatctatatcCGCTTGTGGGGCGAATGTGC	hopH	Mutant construction primers
oIEC143	hopH_4	caaaagaaaatgccgatatccTGGTTCTGTTCCTGGCAGCGG	hopH	Mutant construction primers
oIEC144	1024_1	CCGCTTCCTTTGACCTGGTTGA	G27_1024	Mutant construction primers
oIEC145	1024_2	ACAGCCACCCCACAAAGCGTC	G27_1024	Mutant construction primers
oIEC146	1024_3	atccacttttcaatctatatcTGGTCAATGCGATCGCGCAA	G27_1024	Mutant construction primers
oIEC147	1024_4	caaaagaaaatgccgatatccACGGCTTCTTCCCACGATTTAGAGT	G27_1024	Mutant construction primers
oIEC148	newcagY_1	TTCATTCCTAGCTCTTGAAACGCAGTC	cagY	Mutant construction primers
oIEC149	newcagY_2	ACTGAACCAACAAAAAGTTCAAGTGGC	cagY	Mutant construction primers
oIEC150	newcagY_3	atccacttttcaatctatatcACGCTAAAACCGATGAAGAACGAAACG	cagY	Mutant construction primers
oIEC151	newcagY_4	caaaagaaaatgccgatatccAGCTTCTTTAGACAAGCCTTTCAAGCA	cagY	Mutant construction primers
oIEC192	cagY A	CCGCCACTACATCCTTTAGAAG	cagY	sequence cagY full gene
oIEC193	cagY B	CGAGTCATAATAGGCGTGCCACC	cagY	sequence cagY full gene
oIEC194	cagY C	GAAATCCATTGGTAACAATGGATAACC	cagY	sequence cagY full gene
oIEC195	cagY D	GAGCTTCTCTTCATCGCTCAAACC	cagY	sequence cagY full gene
oIEC196	cagY E	GGCAGTCCTTATAAGCTTTAAGGCTCTC	cagY	sequence cagY full gene
oIEC197	cagY F	AGCTTCGTTTTGGCTCTTGATACGCAAT	cagY	sequence cagY full gene
oIEC198	cagY G	TCAGGGGTGAGTAATTTCTCGCATTCTT	cagY	sequence cagY full gene

oIEC199	cagY Hrev	CTACTAGCTGATATGAGCGTCAAG	cagY	sequence cagY full gene
oIEC200	cagY I	CATTCCTAGCTCTTGAAACGCAG	cagY	sequence cagY full gene
oIEC201	cagY J	CTAGAGTCTTTTTCATTTGCTCTTG	cagY	sequence cagY full gene
oIEC202	cagY Krev	GGTGAATTGGAGCGTGTGATTAAAAAG	cagY	sequence cagY full gene
oIEC203	cagY L	GTGGCATCTACTTTAGAAGTCAGAGTGAT	cagY	sequence cagY full gene
oIEC204	cagY M	CATTCCTAGCTCTTGATACGCAGTCCAAATAAG	cagY	sequence cagY full gene
oIEC205	cagY Nrev	GAGGCTAAAGAGAGCGTTAAAGCTTATAAAGACT	cagY	sequence cagY full gene
oIEC206	cagY O	CTTTAGGGAGATTTTTCAAACACTTTTTTCGTTCTTC	cagY	sequence cagY full gene
oIEC207	wholecagY_1	TGCAACAAACTCCCCAAGCGT	cagY	replace whole cagY
oIEC208	wholecagY_2	AGCAAAGGCGCTAGAGACCCA	cagY	replace whole cagY
oIEC209	wholecagY_3	atccacttttcaatctatatcTGTAGTGGCGGAAAAACCAAGCA	cagY	replace whole cagY
oIEC210	wholecagY_4	caaaagaaaatgccgatatccACAAGAGGGAGCTTTTTAATCACACGC	cagY	replace whole cagY

[a]: lowercase letters indicate common overhangs with homology for stitching to the *catsacB* cassette.

CHAPTER 5

Adaptation to other selective pressures

INTRODUCTION

In addition to the selection of strains with increased ability to colonize and persist in mice, the *H. pylori* clinical isolate G27 has also served as the parental strain for selection of both tissue culture-adapted and laboratory-passaged strains.

Strain G27 MA was selected for increased adhesion to the polarized epithelial cell line MDCK, as well as the ability to translocate the bacterial effector protein CagA to the cells via the Cag type IV secretion system (T4SS) (69). MDCK, or Madin-Darby canine kidney, cells are a polarized epithelial cell line that is useful for investigation of the effects of *H. pylori* on cell polarity and barrier functions (69). The parental strain G27 did not adhere strongly to the cell surface and did not deliver CagA to MDCK cells, necessitating the development of an adapted *H. pylori* strain for investigation of its interactions with these cells. G27 was grown in continuous culture with MDCK cells for 4 months and enriched for bacteria with increased adherence. Clones with increased adherence were tested for CagA translocation by immunoblotting for phosphotyrosine, as CagA is phosphorylated by eukaryotic Src family kinases following delivery into host cells (72). The clone chosen for further study, G27 MA, exhibited an approximately 2-log increase in adherence to MDCK cells, and phosphorylation of CagA was detected, indicating delivery of this virulence factor to the cell cytosol.

The MDCK-adapted *H. pylori* isolate adheres to the apical cell surface primarily near cell-cell junctions, where it results in disruption of junctions and of barrier function (69). *H. pylori* grows and forms microcolonies on the intercellular junctions of polarized MDCK monolayers (136), and its cell-surface growth is enhanced by its CagA-dependent ability to

acquire the essential nutrient iron from the polarized epithelium through disruption of host transferrin recycling (70).

H. pylori can be grown on both solid and liquid media under laboratory conditions, and repeated passaging in liquid culture for approximately 1000 generations resulted in increased fitness of adapted isolates in competition with the ancestral strain (17). The increase in fitness was reduced when *H. pylori* that were not naturally competent for DNA uptake were passaged, highlighting the importance of genetic exchange for bringing together beneficial mutations into the same cell for optimal fitness. The passaging strategy employed involved growth in 5 mL liquid cultures, with a 1:50 dilution after every 24 hours of growth. A sample of the population was archived in frozen storage every 5 days (5 passages), and the cultures were restarted from the frozen stocks after each 25 days, resulting in a bottleneck to approximately 1×10^6 cells. (17) The naturally competent ancestral strain and three isolates independently derived from this ancestor after approximately 1000 generations of passaging were studied here.

To investigate the genetic causes of the phenotypic changes observed in *H. pylori* in both of these experimental systems, strains G27 MA (tissue culture-adapted), G27 DB1 (ancestral strain for laboratory passaging), and three isolates from the endpoint of the 1000 generation laboratory passaging experiment (S1175, S2175, and S5175) were sequenced.

RESULTS

Clustered mutations in three genes in a strain selected for growth on cultured mammalian cells

Tightly clustered SNPs in three genes were identified in a strain selected for growth on mammalian cells in tissue culture, G27 MA (Table 5.1). In the gene *HPG27_910*, encoding an uncharacterized protein, 14 single base changes were identified from Illumina sequence data, and

resequencing confirmed all 14 changes while identifying an additional 18 single base changes with respect to the G27 reference sequence. In the gene *hofH*, encoding an outer membrane protein, 10 single base changes were identified from Illumina sequence data, and resequencing confirmed all 10 changes while identifying an additional 2 changes with respect to the G27 reference sequence. In the gene *HPG27_1112*, encoding carbon starvation protein CstA, 20 single base changes were identified from Illumina sequence data, and resequencing confirmed all 20 changes while identifying an additional 2 changes with respect to the G27 reference sequence. The genomic regions affected by these clustered SNPs are basepairs 989,061-989,515 (*HPG27_910*), 1,218,753-1,219,168 (*hofH*), and 1,220,108-1,220,984 (*HPG27_1112*). All of the identified changes in these three genes were restricted in their presence to strain G27 MA, as these genes were found to match the reference sequence in all 16 of the other strains examined.

Recombination between two *H. pylori* strains in the history of a strain adapted to growth in tissue culture

These changes appear to have occurred via at least three separate recombination events between the sequenced strain and the distantly related strain J99. Regions of 455, 416, and 877 basepairs in genes *HPG27_910*, *hofH*, and *HPG27_1112*, respectively, had numerous differences from the G27 reference sequence in sequenced strain G27 MA. Nucleotide BLAST of these regions identified the homologous genes in *H. pylori* strain J99 (*jhp0897*, *jhp1094*, and *jhp1095*, respectively) as exact matches for each altered region (Fig. 5.1). In addition to matching J99 with 100% identity, the altered regions each had at least 10 mismatches from all sequences in G27, including intergenic sequences, confirming that these changes resulted from inter-strain, rather than intra-strain, recombination.

Although *hofH*/*jhp1094* and *HPG27_1112*/*jhp1095* are adjacent in the G27/J99 genomes, the J99 sequence found in these genes in strain G27 MA is the result of two separate recombination events. As shown in Fig. 5.1, the regions matching J99 sequence are separated by a region that is an exact match to G27 sequence, and not to J99 sequence.

Cause and consequences of interstrain recombination during adaptation to growth in tissue culture

There was no deliberate exposure to strain J99, either DNA or live bacteria, during the selection and passaging of strain G27 MA, but it is possible that contact between the strains may have occurred due to contamination with strains in use in the same laboratory. Deletion of either *hofH* or *HPG27_1112* in strain G27 MA had no effect on bacterial adhesion to polarized MDCK monolayers (Manuel Amieva, personal communication). However, neither the ability of J99 to adhere to MDCK cells nor the effects of the recombination between J99 and G27 in these three genes have been directly examined.

Additional mutations in tissue culture-adapted strain

Additional mutations identified with high confidence (at least 3 reads with high quality and mapping scores) in whole genome sequencing of strain G27 MA include SNPs in flagellar basal body L-ring protein *flgH*, outer membrane protein *sabB*, hypothetical protein *HPG27_654*, the beta subunit of DNA-directed RNA polymerase *HPG27_1142*, oligopeptide ABC transporter *HPG27_1197*, restriction-modification system methylase *HPG27_1314*, putative cation-transporting ATPase *HPG27_1427*, and a transferase involved in peptidoglycan synthesis, *HPG27_1518*. Single base insertions or deletions leading to frameshifts were detected in the minor flagellin subunit *flaB,* the periplasmic dipeptide-binding protein *dppA/hbpA*, the

methyltransferase *HPG27_436*, gamma-glutamyl transpeptidase *HPG27_1063,* and the outer membrane protein *hopL*. All mutations are listed in Table 2.4.

Identification of mutations in strains passaged in vitro

The only change in the in vitro-passaged strains that has been validated by PCR and Sanger sequencing is a SNP in *flgH*, which encodes the flagellar basal body L-ring protein, which was also found in G27 MA and in mouse-adapted strain G27 KO but was absent from independently-derived mouse-adapted strains. The SNP in *flgH* appears to be mutually exclusive with the SNP in another flagellar gene, *fliM*, found in the main lineage of mouse-adapted strains and described in Chapter 3. Other changes identified with high confidence from whole genome sequence data (minimum of 3 reads with high quality scores) in these strains include SNPs in sigma factor *rpoD*, polyphosphate kinase *HPG27_418*, potassium channel protein *HPG27_450*, Cag T4SS component *cagS*, permease *HPG27_571*, hypothetical proteins *HPG27_689* and *_961*, nucleotide biosynthesis factor *HPG27_698*, outer membrane protein *horG*, heptose synthase *HPG27_812*, translation elongation factor *HPG27_1139*, and pyrimidine biosynthetic factor *pyrE*. Single base insertions or deletions leading to frameshifts were found in gamma-glutamyl transpeptidase *HPG27_1063* and hypothetical protein *HPG27_1236*. All mutations are listed in Table 2.4.

Normalized read count comparisons, as described in Chapter 2, indicated that the gene *sabB* appears to have been deleted in strain G27 DB1, the ancestral strain used for in vitro passaging, as well as in all three of its derivatives S1175, S2175, and S5175, passaged for approximately 1000 generations in vitro. Additionally, the nitrite extrusion gene *narK* appeared to be partially deleted in strain S5175, with loss of approximately 500 bp from the 5' end.

Neither of these deletions has been validated by PCR (see Table 2.5). A frameshift mutation in *narK* was also identified in strain S1175.

DISCUSSION

Strain G27 MA, adapted for growth and interaction with a polarized epithelial cell line, exhibited evidence of three separate recombination events with an unrelated *H. pylori* strain, J99. The genes affected were approximately 450 bp of *HPG27_910*, encoding an uncharacterized protein, ~400 bp of *hofH*, encoding an OMP, and ~900 bp of *HPG27_1112*, encoding a carbon starvation protein. Each of these regions in G27 MA was found to be an exact match to the homologous region in strain J99, a commonly used wild-type *H. pylori* clinical isolate.

J99 was isolated in 1994 from a patient in the USA with a duodenal ulcer (13). While both J99 and G27, the ancestral strain for G27 MA, are members of the species *H. pylori*, they are relatively divergent. J99 is assigned to the hpAfrica1 geographic group and G27, isolated in Italy, to the hpEurope group (114). These *H. pylori* groups have been defined on the basis of the sequence of seven housekeeping genes and generally correlate with the origins and migrations of the human populations from which they were isolated (7).

H. pylori is naturally competent for uptake of DNA from the environment and homologous recombination into the genome (21), and recombination is a frequent source of diversity (23). Unintended exposure during the selection of strain G27 MA to either intact bacteria or DNA derived from strain J99 is likely to be the source of the recombinant DNA. The size of the imported J99 regions observed here (400 to 900 bp) is slightly lower than has been previously observed. Mean import sizes of approximately 3000 to 4000 bp were measured for import of J99 DNA carrying a selectable marker into both strain 26695, which, like G27, is a member of the hpEurope group (137), and strain NSH57, derived from G27 (138). The presence

of two regions of imported J99 DNA in close proximity but separated by G27 sequence (*hofH* and *HPG27_1112*) is consistent with previously observed interspersed recipient DNA, which has been hypothesized to be a result of multiple recombination events at a single locus (137).

It is possible that the three recombination-derived changes identified in genes *HPG27_910*, *hofH*, and *HPG27_1112* may contribute to the ability of strain G27 MA to adhere to cells in culture and may have provided an advantage during the selection of this strain. In particular, *hofH* encodes an OMP that could have thus-far uncharacterized adhesin activity. Alternatively, these recombination events may have occurred randomly upon contact with strain J99 and have no effect on this strain's fitness in tissue culture. The ability of J99 to adhere to and grow on cultured cells has not been examined. Experiments placing the recombination-derived J99-like regions of these three genes, singly and in combination, into the parental strain G27 could be used to elucidate their effect on adherence and growth in the cell culture system.

Other changes in G27 MA include an overrepresentation of SNPs and indels in genes encoding OMPs and other cell-surface associated factors. This overrepresentation was even stronger for this tissue-cultured adapted strain than had been observed for the mouse-adapted strains. Although the functional consequences of the identified mutations in G27 MA have not been tested, it is likely that changes in OMPs and other cell surface factors were involved in the increase in this strain's ability to adhere to MDCK cells, the primary phenotype for which this strain was selected.

Three strains derived from approximately 1000 generations of daily passaging in liquid culture, as well as the G27 laboratory stock from which they were derived, were also sequenced. The in vitro passaging regimen by which these strains were generated selected for cells that grow

quickly and can survive at both the high densities achieved just prior to passaging, as well as during the bottlenecks experienced during archival storage.

The presence of independent loss of function changes (one frameshift and one large deletion) in *narK* in two different strains subjected to the same in vitro passaging conditions suggests that this gene may influence growth rate in vitro. Although both mutations are predicted to cause loss of NarK function, it is unclear how loss of nitrite export would enhance bacterial replication. A number of SNPs predicted to cause amino acid substitutions were also identified in the in vitro-passaged strains. Many of these are in genes that could plausibly affect nutrient uptake or growth rate, including a potassium channel, a permease, a translation elongation factor, and factors involved in the biosynthesis of nucleotides, pyrimidines, and heptose. The effects of these mutations on growth phenotypes remain to be determined.

MATERIALS AND METHODS

Strains

Strains used in this study are listed in Table 5.2. All strains are derivatives of the sequenced human clinical isolate G27 (77, 99, 100). Strain G27 MA, selected for adhesion and CagA delivery to Madin-Darby canine kidney (MDCK) cells, has been previously described (69). Strains G27 DB1, S1175, S2175, and S5175 are the ancestral G27 clone and three isolates obtained after approximately 1000 generations of in vitro passage, as described (17).

Resequencing

Regions surrounding the genes *HPG27_910*, *hofH*, and *HPG27_1112* were amplified by PCR and sequenced by capillary sequencing using the primers described in Table 5.3.

Sequence analysis

The regions of *HPG27_910, hofH*, and *HPG27_1112* with concentrated mutations were compared to the G27 reference sequence and to the NCBI non-redundant nucleotide database using BLAST for identification of sequence homologs.

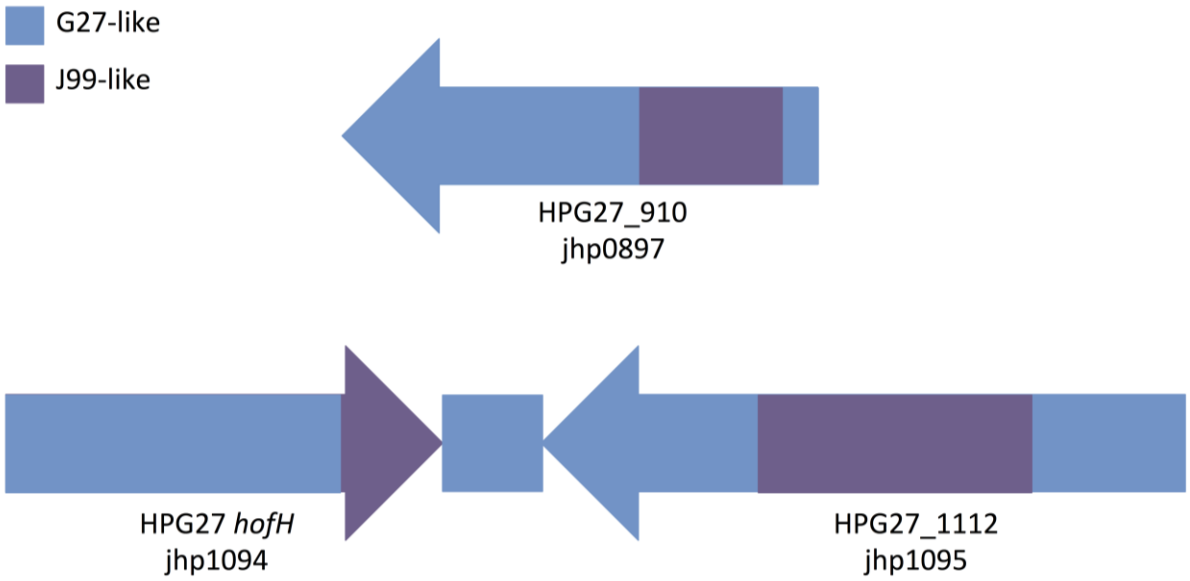

Figure 5.1 Schematic of three separate recombination events importing sequence with an exact match to strain J99 into strain G27 MA. Gene numbers for G27 (HPG27 prefix) and J99 (jhp prefix) are indicated. Regions matching G27 are indicated in blue, and regions matching J99 are indicated in purple.

Table 5.1 All bases changed in three genes of strain G27 MA by recombination with strain J99.

G27 Gene #	J99 Gene #	Position	Ref. Base	New Base	Reads[a]	Amino acid change
910[b]	0897	989061	T	C	12	D187G
910	0897	989081	G	A	11	S180P (w/989083)
910	0897	989083	A	G	13	S180P (w/989081)
910	0897	989111	T	C	16	L170L (syn)
910	0897	989144	A	G	4	L159L (syn)
910	0897	989153	A	G		P156P (syn)
910	0897	989168	C	G	4	L151L (syn) (w/989170)
910	0897	989170	A	G	4	L151L (syn) (w/989168)
910	0897	989201	G	A	9	D140D (syn)
910	0897	989219	C	T	10	L133L (syn)
910	0897	989234	C	T	8	G129G (syn)
910	0897	989261	C	T		E120E (syn)
910	0897	989264	T	C		K119R (w/989265)
910	0897	989265	T	C		K119R (w/989264)
910	0897	989269	T	C		K118E
910	0897	989300	A	G	8	T107T (syn)
910	0897	989314	G	C	4	H103D
910	0897	989336	A	G		G95G (syn)
910	0897	989345	C	G		Q92N (w/989347)
910	0897	989347	G	T		Q92N (w/989345)
910	0897	989371	T	C	3	T84A
910	0897	989387	T	C		V78V (syn)
910	0897	989390	T	C		A77A (syn)
910	0897	989393	G	A		V76V (syn)
910	0897	989396	C	G		M75I
910	0897	989400	T	G		N74T
910	0897	989405	A	T		H72Q
910	0897	989408	A	G		F71F (syn)
910	0897	989463	G	A		A53M (w/989464)
910	0897	989464	C	T		A53M (w/989463)
910	0897	989465	T	C		L52L (syn)
910	0897	989515	C	T	13	E36K
1111[c]	1094	1218753	A	C	25	I358I (syn)
1111	1094	1218777	G	A	21	A366A (syn)
1111	1094	1218834	C	T	14	G385G (syn)
1111	1094	1218909	G	A	11	A410A (syn)
1111	1094	1218929	C	G	10	A417G
1111	1094	1218948	C	T	13	Y423Y (syn)
1111	1094	1218957	C	T	12	G426G (syn)

1111	1094	1219031	*	+A/+ A		T452N, S453F, Q454P, P455A, F456L, I457Y, L458S, T459D, G460R, A461S, R462A, Y463L, T464Y, R465T, P466A, stop467L plus addition of 468V, 469A, 470S, 471F, 472stop
1111	1094	1219032	G	A		K451K (syn)
1111	1094	1219040	A	C	7	Q454P
1111	1094	1219064	G	T	10	R462L
1111	1094	1219076	C	T	7	P466L
1112[d]	1095	1220108	A	G	16	I452I (syn)
1112	1095	1220171	A	G	5	G431G (syn)
1112	1095	1220174	G	A		I430I (syn)
1112	1095	1220183	G	A		T427T (syn)
1112	1095	1220194	T	G	5	S424R
1112	1095	1220202	C	T	6	G421E
1112	1095	1220279	C	G	12	S395S (syn)
1112	1095	1220288	G	A	7	A392A (syn)
1112	1095	1220309	C	T	14	L385L (syn)
1112	1095	1220321	A	G	15	C381C (syn)
1112	1095	1220345	A	G	15	V373V (syn)
1112	1095	1220441	A	G	13	F341F (syn)
1112	1095	1220444	G	T	14	G340G (syn)
1112	1095	1220552	A	G	12	P304P (syn)
1112	1095	1220624	A	G	11	D280D (syn)
1112	1095	1220699	G	A	18	S255S (syn)
1112	1095	1220756	G	A	19	G236G (syn)
1112	1095	1220822	T	C	17	R214R (syn)
1112	1095	1220873	T	A	20	A197A (syn)
1112	1095	1220888	T	G	18	G192G (syn)
1112	1095	1220912	T	C	20	K184K (syn)
1112	1095	1220984	G	C	10	G160G (syn)

[a]The number of Illumina sequence reads supporting the indicated change is given. All listed changes were detected using targeted capillary sequencing of strain G27 MA, and were confirmed to be absent in the other sequenced strains.
[b]*HPG27_910*: Putative uncharacterized protein; genomic position 988097 to 989620; negative strand.
[c]*HPG27_1111* (*hofH*): Outer membrane protein; genomic position 1217680 to 1219080; positive strand.
[d]*HPG27_1112*: Carbon starvation protein; genomic position 1219400 to 1221463; negative strand.

Table 5.2 Strains used in this chapter.

Strain ID	Strain Name	Genotype	Reference	Derived from
G27 MA	G27 MA	adapted to MDCK cells	(69)	G27
G27 DB1	G27 DB1	wild-type G27, working stock	(17)	G27
G27 S1175	G27 S1175	adapted to in vitro passaging	(17)	G27
G27 S2175	G27 S2175	adapted to in vitro passaging	(17)	G27
G27 S5175	G27 S5175	adapted to in vitro passaging	(17)	G27

Table 5.3 Oligonucleotides used in this chapter.

Oligo ID	Oligo Name	Oligo Sequence	Targeted Gene	Purpose of Oligo
oIEC56	910For	CCGCAAAACTCGGACCCGCT	G27_910	Resequencing primers for SNPs/indels
oIEC58	hofHFor	ATGGGGCGTGGAGCTTGGGA	hofH	Resequencing primers for SNPs/indels
oIEC59	1112For	GTTTCAGCCAGCGGCGGGTA	G27_1112	Resequencing primers for SNPs/indels
oIEC68	910Rev	GCGAGATAGGAGCAAAGACGGGC	G27_910	Resequencing primers for SNPs/indels
oIEC70	hofHRev	TGGATGCCGGCACACGAACC	hofH	Resequencing primers for SNPs/indels
oIEC71	1112Ref	CGATAGCAGAGCCACGCCCG	G27_1112	Resequencing primers for SNPs/indels
oIEC80	910Seq	TGTGCGGCTATCAGCCCTACTCA	G27_910	Resequencing primers for SNPs/indels
oIEC82	hofHSeq	GCGGCACAGGCTTTCGCTCT	hofH	Resequencing primers for SNPs/indels
oIEC83	1112Seq1	AACTGCCGGCAAGGCGCTTA	G27_1112	Resequencing primers for SNPs/indels
oIEC84	1112Seq2	TCCCTGCACACACCAACGCC	G27_1112	Resequencing primers for SNPs/indels

CHAPTER 6

Conclusions and Future Directions

H. pylori transmission to a new host, whether person-to-person or in the context of a laboratory animal model, is likely to require alterations in bacterial-host interaction in order to achieve adaptation to the new environment and allow persistent infection. The goals of this dissertation research were:

1) To identify the spectrum of genetic changes that occur during *H. pylori* adaptation to novel environments, including murine infection, interaction with cultured mammalian cells, and laboratory passage; and

2) To characterize the effects of identified mutations on relevant phenotypes, particularly infection of mice.

Chapter 2 addressed the first goal with the identification of approximately 300 single base changes and 25 larger scale deletions and genomic rearrangements across the 17 strains sequenced. Correction of dozens of errors in the reference sequence of clinical isolate G27, resequenced here, will improve the usefulness of this resource for the *H. pylori* research community. This dissertation has demonstrated that adaptation of an *H. pylori* clinical isolate for both murine infection and interaction with cultured cells is accompanied by a bias toward mutations in genes associated with the cell envelope, highlighting the importance of modulating host-bacterium interactions for successful transitions to new environments.

A mutation causing loss of expression of a previously uncharacterized protein, HPG27_792 or Imc1, was found in Chapter 3 to be the largest single contributor to *H. pylori* murine adaption. Imc1 is associated with the bacterial outer membrane and interacts with several periplasmic proteins. Imc1 has a species-specific effect on bacterial adherence to cultured host

cells, but does not appear to affect the induction of cytokine responses against *H. pylori* in vitro or in vivo. Changes in four other genes, *fliM*, *HPG27_230*, *HPG27_557*, and *HPG27_1888*, also contribute to *H. pylori*'s murine colonization potential, although their precise contributions have not been defined.

Chapter 4 turned from the early to the later stages of *H. pylori* adaptation to mice, characterizing the loss of *cag* type IV secretion system (T4SS) activity in strains that persist for months after inoculation of mice. Truncation of the T4SS component CagY was found to be sufficient for abrogation of T4SS function, but evidence indicates that multiple inactivating mutations are present, either within *cagY* or in other genes.

Finally, Chapter 5 briefly described the genetic changes accompanying *H. pylori* adaptation to other selective conditions. Among other changes, recombination with the distinct *H. pylori* strain J99 was observed in the strain selected for interaction with polarized epithelial cells in culture. Parallel inactivation of a nitrite extrusion factor in two of three isolates passaged for 1000 generations in liquid culture suggests adaptive loss of function in *narK*.

As per-base sequencing costs continue to fall, the feasibility of obtaining greater sequence coverage for the strains studied here increases. Higher coverage could facilitate the identification of deletion, duplication, and rearrangement events with more confidence than could be achieved with the 5- to 20-fold coverage obtained here. No shared mutations were found in two mouse-adapted strains generated using the same serial passaging protocol (NSH57 and G27 KO). Additional repeats of the mouse passaging and sequencing of multiple bacterial isolates could identify parallel evolution at the levels of individual genes, or, more likely, functional classes or pathways. Mutations in the same gene or pathway in multiple, independent,

lineages are likely to be adaptive. Such mutations could be prioritized for confirmation by creation and testing of isogenic mutant strains.

Although a loss-of-function mutation in *imc1* was found to be a major driver of murine colonization potential in *H. pylori*, the mechanism by which this occurs has not yet been fully characterized. Although no differences in murine cytokine responses to *H. pylori* with wild type or mutant *imc1* were observed, these responses were measured only at the level of gene expression in the stomach as a whole. Responses of specific, numerically minor, populations of immune cells would likely be undetectable in this system. Microdissection of infected stomach tissue to isolate immune cells could be used to explore this. Alternatively, cultured primary immune cells or cell lines could be exposed to isogenic *H. pylori* mutants and their transcriptional or functional responses measured.

Several putative interaction partners of Imc1 were identified by affinity chromatography and mass spectrometry. The validity of these interactions can be tested by looking for the reverse interaction; that is, can Imc1 be detected among proteins bound to the putative interaction partner? Proteins confirmed as physical interactors with Imc1 can be tested for their functional effects on *H. pylori* colonization of mice, as well as other phenotypes that may be suggested by their annotations.

Several phenotypes of mouse-colonizing strain NSH57 have not been fully explained by the mutations identified in this work. NSH57 exhibits increased adherence to cultured gastric epithelial cells of both human and murine origin, but the isogenic *imc1** mutant carrying the NSH57 allele in the G27 background recapitulates this phenotype for the murine cells only. The adherence phenotypes of isogenic mutants in the other four genes found to differ between G27

and NSH57 remain to be tested, and one or more of these may be responsible for NSH57's increased adherence to cells of human origin.

Similarly, NSH57 displays a distinct shape phenotype from G27, with increased helicity. Isogenic mutants carrying the NSH57 alleles of *imc1* alone or all five genes identified as differentiating the two strains displayed a shape phenotype intermediate between that of parental and adapted bacteria. This result strongly suggests the presence of additional, yet unidentified mutations in mouse-adapted strains, which may be detectable using additional sequencing. Alternatively, as additional genes and pathways contributing to *H. pylori* cell shape are identified in the Salama lab, they can be subjected to targeted sequencing to detect any mutations that occurred during murine adaptation.

Characterization of the later stages of adaptation for persistent infection of mice has thus far been largely restricted to the role of *cagY*. However, isogenic strains with the adapted alleles (as found in MSD85) of genes *hopC, hopH, dnaE, HPG27_676*, and *HPG27_1024* in the NSH57 strain background have been created (see Table 4.1) for use in future phenotypic characterization studies. Investigation of the roles of these genes in multi-week infection of mice is currently ongoing in the lab. To narrow down the location of the presumed additional *cag* T4SS-inactivating mutation to within or outside *cagY*, an isogenic mutant with the entire *cagY* gene of MSD85 replaced with the entire *cagY* gene of NSH57 (MSD85 *wcagY#*) has been created. If the mutation(s) is limited to *cagY*, its reversion to the NSH57 genotype should restore T4SS activity.

Finally, creation and testing of isogenic mutants in the genes identified as having been altered during selection of *H. pylori* for laboratory growth with cultured cells or in liquid culture will allow the identification of the mutations responsible for the adapted phenotypes. This will

hopefully facilitate determination of the mechanisms by which *H. pylori* adapts to these diverse selective pressures.

Understanding how *H. pylori* adapts to mice and to other selective pressures will help clarify important differences between murine and human infection and may lead to an improved mouse model of this disease. The selective pressures involved in the change from human to murine infection, while more dramatic than those involved in transmission between humans, may also contribute to our understanding of how *H. pylori* adapts to a new human host upon transmission.

BIBLIOGRAPHY

1. **Kusters JG, van Vliet AHM, Kuipers EJ**. 2006. Pathogenesis of *Helicobacter pylori* infection. Clin. Microbiol. Rev. **19**:449–90.

2. **Blaser MJ, Kirschner D**. 1999. Dynamics of *Helicobacter pylori* colonization in relation to the host response. Proc. Natl. Acad. Sci. U. S. A. **96**:8359–64.

3. **Blaser MJ, Kirschner D**. 2007. The equilibria that allow bacterial persistence in human hosts. Nature **449**:843–9.

4. **Schreiber S, Konradt M, Groll C, Scheid P, Hanauer G, Werling H-O, Josenhans C, Suerbaum S**. 2004. The spatial orientation of *Helicobacter pylori* in the gastric mucus. Proc. Natl. Acad. Sci. U. S. A. **101**:5024–9.

5. **Akada JK, Ogura K, Dailidiene D, Dailide G, Cheverud JM, Berg DE**. 2003. *Helicobacter pylori* tissue tropism: mouse-colonizing strains can target different gastric niches. Microbiology **149**:1901–1909.

6. **Blaser MJ, Atherton JC**. 2004. *Helicobacter pylori* persistence: biology and disease. J. Clin. Invest. **113**:321–333.

7. **Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, Blaser MJ, Graham DY, Vacher S, Perez-Perez GI, Yamaoka Y, Mégraud F, Otto K, Reichard U, Katzowitsch E, Wang X, Achtman M, Suerbaum S**. 2003. Traces of human migrations in *Helicobacter pylori* populations. Science (80-.). **299**:1582–5.

8. **Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P, Falush D, Stamer C, Prugnolle F, van Der Merwe SW, Yamaoka Y, Graham DY, Perez-Trallero E, Wadström T, Suerbaum S, Achtman M**. 2007. An African origin for the intimate association between humans and *Helicobacter pylori*. Nature **445**:915–8.

9. **Blaser MJ, Falkow S**. 2009. What are the consequences of the disappearing human microbiota? Nat. Rev. Microbiol. **7**:887–94.

10. **Polk DB, Peek RM**. 2010. *Helicobacter pylori*: gastric cancer and beyond. Nat. Rev. Cancer **10**:403–14.

11. **Perry S, de Jong BC, Solnick J V, de la Luz Sanchez M, Yang S, Lin PL, Hansen LM, Talat N, Hill PC, Hussain R, Adegbola R a, Flynn J, Canfield DR, Parsonnet J**. 2010. Infection with *Helicobacter pylori* is associated with protection against tuberculosis. PLoS One **5**:e8804.

12. **Tomb J-F, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill SR, Dougherty BA, Nelson KE, Quackenbush J, Zhou**

L, Kirkness EF, Peterson S, Loftus B, Richardson D, Dodson R, Khalak HG, Glodek A, McKenney K, Fitzegerald LM, Lee N, Adams MD, Hickey EK, Berg DE, Gocayne JD, Utterback TR, Peterson JD, Kelley JM, Cotton MD, Weidman JM, Fujii C, Bowman C, Watthey L, Wallin E, Hayes WS, Borodovsky M, Karp PD, Smith HO, Fraser CM, Venter JC. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature **388**:539–47.

13. **Alm RA, Ling L-SL, Moir DT, King BL, Brown ED, Doig PC, Smith DR, Noonan B, Guild BC, DeJonge BL, Carmel G, Tummino PJ, Caruso A, Uria-Nickelsen M, Mills DM, Ives C, Gibson R, Merberg D, Mills SD, Jiang Q, Taylor DE, Vovis GF, Trust TJ**. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. Nature **397**:176–80.

14. **Ahmed N, Loke MF, Kumar N, Vadivelu J**. 2013. *Helicobacter pylori* in 2013: Multiplying Genomes, Emerging Insights. Helicobacter **18 Suppl 1**:1–4.

15. **Cooke CL, Huff JL, Solnick J V**. 2005. The role of genome diversity and immune evasion in persistent infection with *Helicobacter pylori*. FEMS Immunol. Med. Microbiol. **45**:11–23.

16. **Kang JM, Blaser MJ**. 2006. Bacterial populations as perfect gases: genomic integrity and diversification tensions in *Helicobacter pylori*. Nat. Rev. Microbiol. **4**:826–36.

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

18. **Luria SE, Delbrück M**. 1943. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. Genetics **28**:491–511.

19. **Björkholm B, Sjölund M, Falk PG, Berg OG, Engstrand L, Andersson DI**. 2001. Mutation frequency and biological cost of antibiotic resistance in *Helicobacter pylori*. Proc. Natl. Acad. Sci. U. S. A. **98**:14607–12.

20. **Kang JM, Iovine NM, Blaser MJ**. 2006. A paradigm for direct stress-induced mutation in prokaryotes. FASEB J. **20**:2476–85.

21. **Yeh Y-C, Chang K-C, Yang J-C, Fang C-T, Wang J-T**. 2002. Association of Metronidazole Resistance and Natural Competence in *Helicobacter pylori*. Antimicrob. Agents Chemother. **46**:1564–1567.

22. **Hofreuter D, Odenbreit S, Haas R**. 2001. Natural transformation competence in *Helicobacter pylori* is mediated by the basic components of a type IV secretion system. Mol. Microbiol. **41**:379–391.

23. **Dorer MS, Sessler TH, Salama NR**. 2011. Recombination and DNA Repair in *Helicobacter pylori*. Annu. Rev. Microbiol. **65**:329–348.

24. **Kersulyte D, Chalkauskas H, Berg DE**. 1999. Emergence of recombinant strains of *Helicobacter pylori* during human infection. Mol. Microbiol. **31**:31–43.

25. **Pride DT, Meinersmann RJ, Blaser MJ**. 2001. Allelic variation within *Helicobacter pylori babA* and *babB*. Infect. Immun. **69**:1160–1171.

26. **Aras RA, Fischer W, Perez-Perez GI, Crosatti M, Ando T, Haas R, Blaser MJ**. 2003. Plasticity of repetitive DNA sequences within a bacterial (Type IV) secretion system component. J. Exp. Med. **198**:1349–60.

27. **Blaser MJ**. 1997. Ecology of *Helicobacter pylori* in the Human Stomach. J. Clin. Invest. **100**:759–762.

28. **Kuipers EJ, Israel DA, Kusters JG, Gerrits MM, Weel J, van Der Ende A, van Der Hulst RWM, Wirth H-P, Höök-Nikanne J, Thompson SA, Blaser MJ**. 2000. Quasispecies development of *Helicobacter pylori* observed in paired isolates obtained years apart from the same host. J. Infect. Dis. **181**:273–82.

29. **Israel DA, Salama NR, Krishna US, Rieger UM, Atherton JC, Falkow S, Peek RM**. 2001. *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. Proc. Natl. Acad. Sci. U. S. A. **98**:14625–30.

30. **Giannakis M, Chen SL, Karam SM, Engstrand LG, Gordon JI**. 2008. *Helicobacter pylori* evolution during progression from chronic atrophic gastritis to gastric cancer and its impact on gastric stem cells. Proc. Natl. Acad. Sci. U. S. A. **105**:4358–63.

31. **Falush D, Kraft C, Taylor NS, Correa P, Fox JG, Achtman M, Suerbaum S**. 2001. Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. Proc. Natl. Acad. Sci. U. S. A. **98**:15056–61.

32. **Kraft C, Stack A, Josenhans C, Niehus E, Dietrich G, Correa P, Fox JG, Falush D, Suerbaum S**. 2006. Genomic changes during chronic *Helicobacter pylori* infection. J. Bacteriol. **188**:249–254.

33. **Morelli G, Didelot X, Kusecek B, Schwarz S, Bahlawane C, Falush D, Suerbaum S, Achtman M**. 2010. Microevolution of *Helicobacter pylori* during prolonged infection of single hosts and within families. PLoS Genet. **6**:e1001036.

34. **Kennemann L, Didelot X, Aebischer T, Kuhn S, Drescher B, Droege M, Reinhardt R, Correa P, Meyer TF, Josenhans C, Falush D, Suerbaum S**. 2011. *Helicobacter pylori* genome evolution during human infection. Proc. Natl. Acad. Sci. U. S. A. **108**:5033–5038.

35. **Salama NR, Gonzalez-Valencia G, Deatherage B, Aviles-Jimenez F, Atherton JC, Graham DY, Torres J**. 2007. Genetic analysis of *Helicobacter pylori* strain populations colonizing the stomach at different times postinfection. J. Bacteriol. **189**:3834–45.

36. **Björkholm BM, Lundin A, Sillen A, Guillemin K, Salama NR, Rubio C, Gordon JI, Falk PG, Engstrand LG**. 2001. Comparison of genetic divergence and fitness between two subclones of *Helicobacter pylori*. Infect. Immun. **69**:7832–7838.

37. **Dorer MS, Fero J, Salama NR**. 2010. DNA Damage Triggers Genetic Exchange in *Helicobacter pylori*. PLoS Pathog. **6**:e1001026.

38. **Dorer MS, Cohen IE, Sessler TH, Fero J, Salama NR**. 2013. Natural Competence Promotes *Helicobacter pylori* Chronic Infection. Infect. Immun. **81**:209–15.

39. **Bardhan PK**. 1997. Epidemiological features of *Helicobacter pylori* infection in developing countries. Clin. Infect. Dis. **25**:973–8.

40. **Amieva MR, El-Omar EM**. 2008. Host-bacterial interactions in *Helicobacter pylori* infection. Gastroenterology **134**:306–23.

41. **Weyermann M, Adler G, Brenner H, Rothenbacher D**. 2006. The mother as source of *Helicobacter pylori* infection. Epidemiology **17**:332–4.

42. **Kivi M, Johansson AL V, Reilly M, Tindberg Y**. 2005. *Helicobacter pylori* status in family members as risk factors for infection in children. Epidemiol. Infect. **133**:645–52.

43. **Schwarz S, Morelli G, Kusecek B, Manica A, Balloux F, Owen RJ, Graham DY, van Der Merwe SW, Achtman M, Suerbaum S**. 2008. Horizontal versus familial transmission of *Helicobacter pylori*. PLoS Pathog. **4**:e1000180.

44. **Didelot X, Nell S, Yang I, Woltemate S, van Der Merwe SW, Suerbaum S**. 2013. Genomic evolution and transmission of *Helicobacter pylori* in two South African families. Proc. Natl. Acad. Sci. U. S. A. **110**:13880–13885.

45. **McCallion WA, Murray LJ, Bailie AG, Dalzell AM, O'Reilly DPJ, Bamford KB**. 1996. *Helicobacter pylori* infection in children: relation with current household living conditions. Gut **39**:18–21.

46. **Raymond J, Thiberg J-M, Chevalier C, Kalach N, Bergeret M, Labigne A, Dauga C**. 2004. Genetic and transmission analysis of *Helicobacter pylori* strains within a family. Emerg. Infect. Dis. **10**:1816–21.

47. **Kivi M, Rodin S, Kupershmidt I, Lundin A, Tindberg Y, Granström M, Engstrand L**. 2007. *Helicobacter pylori* genome variability in a framework of familial transmission. BMC Microbiol. **7**:54.

48. **Raymond J, Thiberge J-M, Kalach N, Bergeret M, Dupont C, Labigne A, Dauga C**. 2008. Using macro-arrays to study routes of infection of *Helicobacter pylori* in three families. PLoS One **3**:e2259.

49. **Linz B, Windsor HM, Gajewski JP, Hake CM, Drautz DI, Schuster SC, Marshall BJ**. 2013. *Helicobacter pylori* Genomic Microevolution during Naturally Occurring Transmission between Adults. PLoS One **8**:e82187.

50. **Dubois A, Berg DE, Incecik ET, Fiala N, Heman-Ackah LM, Perez-Perez GI, Blaser MJ**. 1996. Transient and persistent experimental infection of nonhuman primates with *Helicobacter pylori*: implications for human disease. Infect. Immun. **64**:2885–91.

51. **Peek RM**. 2008. *Helicobacter pylori* infection and disease: from humans to animal models. Dis. Model. Mech. **1**:50–5.

52. **Franco AT, Israel DA, Washington MK, Krishna US, Fox JG, Rogers AB, Neish AS, Collier-Hyams L, Perez-Perez GI, Hatakeyama M, Whitehead R, Gaus K, O'Brien DP, Romero-Gallo J, Peek RM**. 2005. Activation of beta-catenin by carcinogenic *Helicobacter pylori.* Proc. Natl. Acad. Sci. U. S. A. **102**:10646–51.

53. **Cai X, Carlson J, Stoicov C, Li H, Wang TC, Houghton J**. 2005. Eradication Restores Normal Architecture and Inhibits Gastric Cancer Progression in C57BL/6 Mice. Gastroenterology **128**:1937–1952.

54. **Falk PG, Bry L, Holgersson J, Gordon JI**. 1995. Expression of a human a-1,3/4-fucosyltransferase in the pit cell lineage of FVB/N mouse stomach results in production of Leb-containing glycoconjugates: A potential transgenic mouse model for studying *Helicobacter pylori* infection. Proc. Natl. Acad. Sci. U. S. A. **92**:1515–1519.

55. **Hoffman PS, Vats N, Hutchison D, Butler J, Chisholm K, Sisson G, Raudonikiene A, Marshall JS, Veldhuyzen Van Zanten SJO**. 2003. Development of an interleukin-12-deficient mouse model that is permissive for colonization by a motile KE26695 strain of *Helicobacter pylori*. Infect. Immun. **71**:2534–2541.

56. **Marchetti M, Aricò B, Burroni D, Figura N, Rappuoli R, Ghiara P**. 1995. Development of a mouse model of *Helicobacter pylori* infection that mimics human disease. Science (80-.). **267**:1655–8.

57. **Lee A, O'Rourke J, de Ungria MC, Robertson B, Daskalopoulos G, Dixon MF**. 1997. A standardized mouse model of *Helicobacter pylori* infection: introducing the Sydney strain. Gastroenterology **112**:1386–1397.

58. **Thompson LJ, Danon SJ, Wilson JE, O'Rourke JL, Salama NR, Falkow S, Mitchell HM, Lee A**. 2004. Chronic *Helicobacter pylori* infection with Sydney strain 1 and a newly identified mouse-adapted strain (Sydney strain 2000) in C57BL/6 and BALB/c mice. Infect. Immun. **72**:4668–4679.

59. **Baldwin DN, Shepherd B, Kraemer P, Hall MK, Sycuro LK, Pinto-Santini DM, Salama NR**. 2007. Identification of *Helicobacter pylori* genes that contribute to stomach colonization. Infect. Immun. **75**:1005–16.

60. **Fox JG, Wang TC, Rogers AB, Poutahidis T, Ge Z, Taylor NS, Dangler CA, Israel DA, Krishna US, Gaus K, Peek RM**. 2003. Host and microbial constituents influence *Helicobacter pylori*-induced cancer in a murine model of hypergastrinemia. Gastroenterology **124**:1879–90.

61. **Ottemann KM, Lowenthal AC**. 2002. *Helicobacter pylori* uses motility for initial colonization and to attain robust infection. Infect. Immun. **70**:1984–1990.

62. **Terry K, Williams SM, Connolly LE, Ottemann KM**. 2005. Chemotaxis plays multiple roles during *Helicobacter pylori* animal infection. Infect. Immun. **73**:803–811.

63. **Snelling WJ, Moran AP, Ryan KA, Scully P, McGourty K, Cooney JC, Annuk H, O'Toole PW**. 2007. HorB (HP0127) is a gastric epithelial cell adhesin. Helicobacter **12**:200–9.

64. **Sycuro LK, Pincus Z, Gutierrez KD, Biboy J, Stern C a, Vollmer W, Salama NR**. 2010. Peptidoglycan crosslinking relaxation promotes *Helicobacter pylori*'s helical shape and stomach colonization. Cell **141**:822–33.

65. **Sycuro LK, Wyckoff TJ, Biboy J, Born P, Pincus Z, Vollmer W, Salama NR**. 2012. Multiple Peptidoglycan Modification Networks Modulate *Helicobacter pylori*'s Cell Shape, Motility, and Colonization Potential. PLoS Pathog. **8**:e1002603.

66. **Sycuro LK, Rule CS, Petersen TW, Wyckoff TJ, Sessler T, Nagarkar DB, Khalid F, Pincus Z, Biboy J, Vollmer W, Salama NR**. 2013. Flow cytometry-based enrichment for cell shape mutants identifies multiple genes that influence *Helicobacter pylori* morphology. Mol. Microbiol. **90**:869–83.

67. **Montecucco C, Rappuoli R**. 2001. Living dangerously: how *Helicobacter pylori* survives in the human stomach. Nat. Rev. Mol. Cell Biol. **2**:457–66.

68. **Sharma SA, Tummuru MKR, Miller GG, Blaser MJ**. 1995. Interleukin-8 response of gastric epithelial cell lines to *Helicobacter pylori* stimulation in vitro. Infect. Immun. **63**:1681–1687.

69. **Amieva MR, Vogelmann R, Covacci A, Tompkins LS, Nelson WJ, Falkow S**. 2003. Disruption of the epithelial apical-junctional complex by *Helicobacter pylori* CagA. Science (80-.). **300**:1430–1434.

70. **Tan S, Noto JM, Romero-Gallo J, Peek RM, Amieva MR**. 2011. *Helicobacter pylori* Perturbs Iron Trafficking in the Epithelium to Grow on the Cell Surface. PLoS Pathog. **7**:e1002050.

71. **Stein M, Rappuoli R, Covacci A**. 2000. Tyrosine phosphorylation of the *Helicobacter pylori* CagA antigen after cag-driven host cell translocation. Proc. Natl. Acad. Sci. U. S. A. **97**:1263–8.

72. **Hatakeyama M**. 2014. *Helicobacter pylori* CagA and Gastric Cancer: A Paradigm for Hit-and-Run Carcinogenesis. Cell Host Microbe **15**:306–316.

73. **Pop M, Salzberg SL**. 2008. Bioinformatics challenges of new sequencing technology. Trends Genet. **24**:142–9.

74. **Mardis ER**. 2013. Next-generation sequencing platforms. Annu. Rev. Anal. Chem. **6**:287–303.

75. **Shendure JA, Ji H**. 2008. Next-generation DNA sequencing. Nat. Biotechnol. **26**:1135–45.

76. **Cahill MJ, Köser CU, Ross NE, Archer JAC**. 2010. Read length and repeat resolution: exploring prokaryote genomes using next-generation sequencing technologies. PLoS One **5**:e11518.

77. **Baltrus DA, Amieva MR, Covacci A, Lowe TM, Merrell DS, Ottemann KM, Stein M, Salama NR, Guillemin K**. 2009. The complete genome sequence of *Helicobacter pylori* strain G27. J. Bacteriol. **191**:447–8.

78. **Dohm JC, Lottaz C, Borodina T, Himmelbauer H**. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res. **36**:e105.

79. **Craig D, Pearson J, Szelinger S, Sekar A, Redman M, Corneveaux JJ, Pawlowski TL, Laub T, Nunn G, Stephan DA, Homer N, Huentelman MJ**. 2008. Identification of genetic variants using bar-coded multiplexed sequencing. Nat. Methods **5**:887–893.

80. **Linnarsson S**. 2010. Recent advances in DNA sequencing methods - general principles of sample preparation. Exp. Cell Res. **316**:1339–43.

81. **Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, Stackhouse B, MacKenzie AP, Caruccio NC, Zhang X, Shendure J**. 2010. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol. **11**:R119.

82. **Li H, Ruan J, Durbin R**. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. **18**:1851–8.

83. **Li H, Durbin R**. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**:1754–60.

84. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup .** 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**:2078–9.

85. **Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF**. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature **461**:1243–7.

86. **Lieberman TD, Michel J-B, Aingaran M, Potter-Bynoe G, Roux D, Davis MR, Skurnik D, Leiby N, LiPuma JJ, Goldberg JB, McAdam AJ, Priebe GP, Kishony R**. 2011. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. Nat. Genet. **43**:1275–80.

87. **Price EP, Sarovich DS, Mayo M, Tuanyok A, Drees KP, Kaestli M, Beckstrom-Sternberg SM, Babic-Sternberg JS, Kidd TJ, Bell SC, Kiem P, Pearson T, Currie BJ**. 2013. Within-host evolution of *Burkholderia pseudomallei* over a twelve-year chronic carriage infection. MBio **4**:e00388–13.

88. **Barroso-Batista J, Sousa A, Lourenço M, Bergman M-L, Sobral D, Demengeot J, Xavier KB, Gordo I**. 2014. The First Steps of Adaptation of *Escherichia coli* to the Gut Are Dominated by Soft Sweeps. PLoS Genet. **10**:e1004182.

89. **Alm RA, Bina J, Andrews BM, Doig PC, Hancock RE, Trust TJ**. 2000. Comparative genomics of *Helicobacter pylori:* analysis of the outer membrane protein families. Infect. Immun. **68**:4155–68.

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

91. **Castillo AR, Arevalo SS, Woodruff AJ, Ottemann KM**. 2008. Experimental analysis of *Helicobacter pylori* transcriptional terminators suggests this microbe uses both intrinsic and factor-dependent termination. Mol. Microbiol. **67**:155–70.

92. **Ilver D, Arnqvist A, Ogren J, Frick I-M, Kersulyte D, Incecik ET, Berg DE, Covacci A, Engstrand LG, Borén T**. 1998. *Helicobacter pylori* Adhesin Binding Fucosylated Histo-Blood Group Antigens Revealed by Retagging. Science (80-.). **279**:373–377.

93. **Mahdavi J, Sondén B, Hurtig M, Olfat FO, Forsberg L, Roche N, Angstrom J, Larsson T, Teneberg S, Karlsson K-A, Altraja S, Wadström T, Kersulyte D, Berg DE, Dubois A, Petersson C, Magnusson K-E, Norberg T, Lindh F, Lundskog BB, Arnqvist A, Hammarström L, Borén T**. 2002. *Helicobacter pylori* SabA adhesin in persistent infection and chronic inflammation. Science (80-.). **297**:573–8.

94. **Viala J, Chaput C, Boneca IG, Cardona A, Girardin SE, Moran AP, Athman R, Mémet S, Huerre MR, Coyle AJ, DiStefano PS, Sansonetti PJ, Labigne A, Bertin J, Philpott DJ, Ferrero RL**. 2004. Nod1 responds to peptidoglycan delivered by the *Helicobacter pylori cag* pathogenicity island. Nat. Immunol. **5**:1166–74.

95. **Sayi A, Kohler E, Hitzler I, Arnold I, Schwendener R, Rehrauer H, Müller A**. 2009. The CD4+ T cell-mediated IFN-gamma response to *Helicobacter* infection is essential for clearance and determines gastric cancer risk. J. Immunol. **182**:7085–101.

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

97. **Wirth H-P, Yang M, Sanabria-Valentín E, Berg DE, Dubois A, Blaser MJ**. 2006. Host Lewis phenotype-dependent *Helicobacter pylori* Lewis antigen expression in rhesus monkeys. FASEB J. **20**:1534–6.

98. **Styer CM, Hansen LM, Cooke CL, Gundersen AM, Choi SS, Berg DE, Benghezal M, Marshall BJ, Peek RM, Borén T, Solnick J V**. 2010. Expression of the BabA adhesin during experimental infection with *Helicobacter pylori*. Infect. Immun. **78**:1593–600.

99. **Covacci A, Censini S, Bugnoli M, Petracca R, Burroni D, Macchia G, Massone A, Papini E, Xiang Z, Figura N, Rappuoli R**. 1993. Molecular characterization of the 128-kDa immunodominant antigen of *Helicobacter pylori* associated with cytotoxicity and duodenal ulcer. Proc. Natl. Acad. Sci. U. S. A. **90**:5791–5.

100. **Xiang Z, Censini S, Bayeli PF, Telford JL, Figura N, Rappuoli R, Covacci A**. 1995. Analysis of expression of CagA and VacA virulence factors in 43 strains of *Helicobacter pylori* reveals that clinical isolates can be divided into two major types and that CagA is not necessary for expression of the vacuolating cytotoxin. Infect. Immun. **63**:94–8.

101. **Censini S, Lange C, Xiang Z, Crabtree JE, Ghiara P, Borodovsky M, Rappuoli R, Covacci A**. 1996. *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. Proc. Natl. Acad. Sci. U. S. A. **93**:14648–53.

102. **Chalker AF, Minehart HW, Hughes NJ, Koretke KK, Lonetto MA, Brinkman KK, Warren P V, Lupas A, Stanhope MJ, Brown JR, Hoffman PS**. 2001. Systematic identification of selective essential genes in *Helicobacter pylori* by genome prioritization and allelic replacement mutagenesis. J. Bacteriol. **183**:1259–1268.

103. **Sycuro LK**. 2009. Ph.D. Thesis. University of Washington.

104. **Eaton K, Dewhirst F, Radin M, Fox J, Paster B, Krakowka S, Morgan D**. 1993. *Helicobacter acinonyx sp. nov*., isolated from cheetahs with gastritis. Int. J. Syst. Bacteriol. **43**:99–106.

105. **Harper C, Feng Y, Xu S, Taylor N, Kinsel M, Dewhirst F, Paster B, Greenwell M, Levine G, Rogers A, Fox J**. 2002. *Helicobacter cetorum sp . nov* ., a Urease-Positive *Helicobacter* Species Isolated from Dolphins and Whales. J. Clin. Microbiol. **40**:4536–4543.

106. **Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R, Stadler PF, Vogel J**. 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature **464**:250–5.

107. **Pinto-Santini DM**. 2008. Ph.D. Thesis. University of Washington.

108. **Whitehead RH, VanEeden PE, Noble MD, Ataliotis P, Jat PS**. 1993. Establishment of conditionally immortalized epithelial cell lines from both colon and small intestine of adult H-2Kb-tsA58 transgenic mice. Proc. Natl. Acad. Sci. U. S. A. **90**:587–91.

109. **Chaturvedi R, de Sablet T, Peek RM, Wilson KT**. 2012. Spermine oxidase, a polyamine catabolic enzyme that links *Helicobacter pylori* CagA and gastric cancer risk. Gut Microbes **3**:48–56.

110. **Algood HMS, Cover TL**. 2006. *Helicobacter pylori* persistence: an overview of interactions between *H. pylori* and host immune defenses. Clin. Microbiol. Rev. **19**:597–613.

111. **Hottes AK, Freddolino PL, Khare A, Donnell ZN, Liu JC, Tavazoie S**. 2013. Bacterial Adaptation through Loss of Function. PLoS Genet. **9**:e1003617.

112. **Sozzi M, Crosatti M, Kim S-K, Romero-Gallo J, Blaser MJ**. 2001. Heterogeneity of *Helicobacter pylori cag* genotypes in experimentally infected mice. FEMS Microbiol. Lett. **203**:109–14.

113. **Philpott DJ, Belaid D, Troubadour P, Thiberge J-M, Tankovic J, Labigne A, Ferrero RL**. 2002. Reduced activation of inflammatory responses in host cells by mouse-adapted *Helicobacter pylori* isolates. Cell. Microbiol. **4**:285–296.

114. **Mane SP, Dominguez-Bello MG, Blaser MJ, Sobral BW, Hontecillas R, Skoneczka J, Mohapatra SK, Crasta OR, Evans C, Modise T, Shallom S, Shukla M, Varon C, Mégraud F, Maldonado-Contreras A, Williams KP, Bassaganya-Riera J**. 2010. Host-interactive genes in Amerindian *Helicobacter pylori* diverge from their Old World homologs and mediate inflammatory responses. J. Bacteriol. **192**:3078–92.

115. **Deuel TF, Senior RM, Chang D, Griffin GL, Heinrikson RL, Kaiser ET**. 1981. Platelet factor 4 is chemotactic for neutrophils and monocytes. Proc. Natl. Acad. Sci. U. S. A. **78**:4584–7.

116. **Wolpe SD, Sherry B, Juers D, Davatelis G, Yurt RW, Cerami A**. 1989. Identification and characterization of macrophage inflammatory protein 2. Proc. Natl. Acad. Sci. U. S. A. **86**:612–6.

117. **Smythies LE, Waites KB, Lindsey JR, Harris PR, Ghiara P, Smith PD**. 2000. *Helicobacter pylori*-induced mucosal inflammation is Th1 mediated and exacerbated in IL-4, but not IFN-gamma, gene-deficient mice. J. Immunol. **165**:1022–9.

118. **Horton R**. 1997. In vitro recombination and mutagenesis of DNA. SOEing together tailor-made genes. Methods Mol. Biol. **67**:141–149.

119. **Salama NR, Otto G, Tompkins LS, Falkow S**. 2001. Vacuolating cytotoxin of *Helicobacter pylori* plays a role during colonization in a mouse model of infection. Infect. Immun. **69**:730–736.

120. **Ye S, Dhillon S, Ke X, Collins AR, Day INM**. 2001. An efficient procedure for genotyping single nucleotide polymorphisms. Nucleic Acids Res. **29**:E88.

121. **Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, Machi D, Mao C, Nordberg EK, Olson R, Overbeek R, Pusch GD, Shukla M, Schulman J, Stevens RL, Sullivan DE, Vonstein V, Warren A, Will R, Wilson MJC, Yoo HS, Zhang C, Zhang Y, Sobral BW**. 2014. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res. **42**:D581–91.

122. **Stöver BC, Müller KF**. 2010. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics **11**:7.

123. **Petersen TN, Brunak S, von Heijne G, Nielsen H**. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods **8**:785–6.

124. **Sheu B-S, Odenbreit S, Hung K-H, Liu C-P, Sheu S-M, Yang H-B, Wu J-J**. 2006. Interaction between host gastric Sialyl-Lewis X and *H. pylori* SabA enhances *H. pylori* density in patients lacking gastric Lewis B antigen. Am. J. Gastroenterol. **101**:36–44.

125. **Chiu BC, Shang X, Frait KA, Hu JS, Komuniecki E, Miller RA, Chensue SW**. 2002. Differential effects of ageing on cytokine and chemokine responses during type-1 (mycobacterial) and type-2 (schistosomal) pulmonary granulomatous inflammation in mice. Mech. Ageing Dev. **123**:313–26.

126. **Shi Y, Liu X-F, Zhuang Y, Zhang J-Y, Liu T, Yin Z, Wu C, Mao X-H, Jia K-R, Wang F-J, Guo H, Flavell RA, Zhao Z, Liu K-Y, Xiao B, Guo Y, Zhang W-J, Zhou W-Y, Guo G, Zou Q-M**. 2010. *Helicobacter pylori*-induced Th17 responses modulate Th1 cell responses, benefit bacterial growth, and contribute to pathology in mice. J. Immunol. **184**:5121–9.

127. **Sommer F, Wilken H, Faller G, Lohoff M**. 2004. Systemic Th1 immunization of mice against *Helicobacter pylori* infection with CpG oligodeoxynucleotides as adjuvants does not protect from infection but enhances gastritis. Infect. Immun. **72**:1029–1035.

128. **Cook WJ, Kramer MF, Walker RM, Burwell TJ, Holman HA, Coen DM, Knipe DM**. 2004. Persistent expression of chemokine and chemokine receptor RNAs at primary and latent sites of herpes simplex virus 1 infection. Virol. J. **1**.

129. **Harrington LE, Hatton RD, Mangan PR, Turner H, Murphy TL, Murphy KM, Weaver CT**. 2005. Interleukin 17-producing CD4+ effector T cells develop via a lineage distinct from the T helper type 1 and 2 lineages. Nat. Immunol. **6**:1123–32.

130. **Grenda DS, Johnson SE, Mayer JR, McLemore ML, Benson KF, Horwitz M, Link DC**. 2002. Mice expressing a neutrophil elastase mutation derived from patients with severe congenital neutropenia have normal granulopoiesis. Blood **100**:3221–8.

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

132. **Delahay RM, Balkwill GD, Bunting KA, Edwards W, Atherton JC, Searle MS**. 2008. The highly repetitive region of the *Helicobacter pylori* CagY protein comprises tandem arrays of an alpha-helical repeat module. J. Mol. Biol. **377**:956–71.

133. **Perkins TT, Tay CY, Thirriot F, Marshall B**. 2013. Choosing a Benchtop Sequencing Machine to Characterise *Helicobacter pylori* Genomes. PLoS One **8**:e67539.

134. **Aras RA, Kang JM, Tschumi AI, Harasaki Y, Blaser MJ**. 2003. Extensive repetitive DNA facilitates prokaryotic genome plasticity. Proc. Natl. Acad. Sci. U. S. A. **100**:13579–84.

135. **Magalhaes JG, Philpott DJ, Nahori M-A, Jéhanno M, Fritz J, Le Bourhis L, Viala J, Hugot J-P, Giovannini M, Bertin J, Lepoivre M, Mengin-Lecreulx D, Sansonetti PJ, Girardin SE**. 2005. Murine Nod1 but not its human orthologue mediates innate immune detection of tracheal cytotoxin. EMBO Rep. **6**:1201–7.

136. **Tan S, Tompkins LS, Amieva MR**. 2009. *Helicobacter pylori* usurps cell polarity to turn the cell surface into a replicative niche. PLoS Pathog. **5**:e1000407.

137. **Lin EA, Zhang X-S, Levine SM, Gill SR, Falush D, Blaser MJ**. 2009. Natural transformation of *Helicobacter pylori* involves the integration of short DNA fragments interrupted by gaps of variable size. PLoS Pathog. **5**:e1000337.

138. **Humbert O, Dorer MS, Salama NR**. 2011. Characterization of *Helicobacter pylori* factors that control transformation frequency and integration length during inter-strain DNA recombination. Mol. Microbiol. **79**:387–401.

VITA

Ilana Cohen was born and raised in the Minneapolis, Minnesota area. She earned a Bachelor of Arts degree in Biological Sciences, with a concentration in Microbiology, from Cornell University in 2006. She received magna cum laude honors for her undergraduate thesis research on the response of dendritic cells to infection with the parasite *Toxoplasma gondii*, as well as academic distinction in all subjects. Ilana moved to Seattle when her husband began his graduate work. She worked for two years as a research technician at the University of Washington, studying the outer membrane of the bacterial pathogens *Francisella tularensis* and *Yersinia pestis*. She began her graduate work in the University of Washington's Molecular and Cellular Biology program in 2008. Ilana's dissertation research was performed in the laboratory of Nina Salama at the Fred Hutchinson Cancer Research Center. She investigated the genetic basis of *Helicobacter pylori* adaptation to novel conditions, and received her doctoral degree in 2014.