# Contribution of the ArsRS two-component system and HomB to *Helicobacter pylori* biofilm formation

by

Stephanie Leigh Servetas

Dissertation submitted to the Faculty of the
Emerging Infectious Diseases Graduate Program
Uniformed Services University of the Health Sciences
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy 2016

**UNIFORMED SERVICES UNIVERSITY, SCHOOL OF MEDICINE GRADUATE PROGRAMS**
Graduate Education Office (A 1045), 4301 Jones Bridge Road, Bethesda, MD 20814

APPROVAL OF THE DOCTORAL DISSERTATION IN THE EMERGING INFECTIOUS DISEASES GRADUATE PROGRAM

Title of Dissertation: "Contribution of the ArsRS Two-Component System and HomB to *Helicobacter pylori* Biofilm Formation"

Name of Candidate: Stephanie L. Servetas
Doctor of Philosophy Degree
December 8, 2016

DISSERTATION AND ABSTRACT APPROVED:

DATE:

[redacted] 12-8-16
Dr. Ann Jerse
DEPARTMENT OF MICROBIOLOGY & IMMUNOLOGY
Committee Chairperson

[redacted] 12-15-16
Dr. D. Scott Merrell
DEPARTMENT OF MICROBIOLOGY & IMMUNOLOGY
Dissertation Advisor

[redacted] 12-8-16
Dr. Patricia Guerry
DEPARTMENT OF MICROBIOLOGY & IMMUNOLOGY
Committee Member

[redacted] 12-8-16
Dr. Jennifer Rusiecki
DEPARTMENT OF PREVENTIVE MEDICINE & BIOSTATISTICS
Committee Member

UNIFORMED SERVICES UNIVERSITY, SCHOOL OF MEDICINE GRADUATE PROGRAMS
Graduate Education Office (A 1045), 4301 Jones Bridge Road, Bethesda, MD 20814

# FINAL EXAMINATION/PRIVATE DEFENSE FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE EMERGING INFECTIOUS DISEASES GRADUATE PROGRAM

Name of Student: Stephanie L. Servetas

Date of Examination: December 8, 2016
Time: 1:00 PM
Place: B4144

DECISION OF EXAMINATION COMMITTEE MEMBERS:

| | PASS | FAIL |
|---|---|---|
| [redacted] | X | |
| Dr. Ann Jerse<br>DEPARTMENT OF MICROBIOLOGY & IMMUNOLOGY<br>Committee Chairperson | | |
| [redacted] | X | |
| Dr. D. Scott Merrell<br>DEPARTMENT OF MICROBIOLOGY & IMMUNOLOGY<br>Dissertation Advisor | | |
| [redacted] | X | |
| Dr. Patricia Guerry<br>DEPARTMENT OF MICROBIOLOGY & IMMUNOLOGY<br>Committee Member | | |
| [redacted] | X | |
| Dr. Jennifer Rusiecki<br>DEPARTMENT OF PREVENTIVE MEDICINE & BIOSTATISTICS<br>Committee Member | | |

# ACKNOWLEDGMENTS

To my advisor, Scotty, thank-you for being an amazing mentor. Thank-you for acknowledging my weaknesses, sharing your wisdom, providing guidance, and helping me grow as a scientist.

To my thesis committee- thank-you for all of your insightful discussions and words of encouragement throughout my graduate career.

To my labmates and classmates, graduate school was not always fun and it was definitely not easy. Thank-you for always being there to celebrate the small things and to commiserate when science didn't go our way.

To the MIC and EID faculty and staff, thank-you for all of your support and making the fourth floor a fun place.

To my little brother, Eric, thank-you for being a constant source of support and comic relief.

To my friends and family, whether you live in the same house, around the beltway, or across the country, thank-you for always being there to listen to me vent and to make me laugh.

To the Carpenters and the Fergusons, thank-you for making me feel like family and for giving me a home away from home.

# DEDICATION

For my Grandmothers,

Thank you for inspiring me, encouraging me, and always believing in me.

For my Parents,

Thank you for your endless support and love.

*"It is not what you do for your children, but what you have taught them to do for themselves that will make them successful human beings."*

*-Ann Landers*

## COPYRIGHT STATEMENT

The author hereby certifies that the use of any copyrighted material in the dissertation manuscript entitled:

"Contribution of the ArsRS Two-Component System and HomB to *Helicobacter pylori* Biofilm Formation"

is appropriately acknowledged and, beyond brief excerpts, is with the permission of the copyright owner.

[Signature] [redacted]

Stephanie L. Servetas

February 2, 2017

# ABSTRACT

Title of Dissertation:

Contribution of the ArsRS Two-Component System and HomB to *Helicobacter pylori* Biofilm Formation

Stephanie L. Servetas, Doctor of Philosophy, 2016

Thesis directed by:

D. Scott Merrell, Ph.D.
Professor, Department of Microbiology and Immunology

Discovery of *H. pylori* within the human stomach not only defied the dogma that the human stomach was a sterile environment, but also revolutionized the treatment of gastric and duodenal ulcers. Moreover, *H. pylori* became the first, and only, bacterium to be classified as a class I carcinogen. While, treatment of *H. pylori* infection leads to beneficial outcomes with respect to incidence of peptic ulcer disease and gastric cancers, treatment failure following *H. pylori* eradication therapy can be as high as 20%; rates of resistance continue to rise. For a pathogen in which a two-week course of triple or quadruple therapy is already the current recommendation, the need for a better treatment regimen is evident.

Biofilm formation in *H. pylori* has been demonstrated during natural human infection of *H. pylori* and has been shown to impart increased antibiotic resistance *in vitro*. A greater understanding of the bacterial factors that are involved in biofilm

formation and antibiotic resistance could lead to the development of more targeted therapeutics. The work presented in this dissertation identifies the ArsRS acid responsive two-component system as a regulator of biofilm formation by *H. pylori*. Our findings suggest that biofilm formation by *H. pylori* is repressed by ArsR~P; mutation of ArsRS results in enhanced biofilm formation at a neutral pH.

This is the first two-component regulatory protein that has been shown to modulate *H. pylori* biofilm formation, and these data suggest that pH may act as an environmental signal for *H. pylori* biofilm development. In addition, we identified the outer membrane protein (OMP), *homB* as an ArsRS regulated gene and showed that this OMP contributes to ArsRS-regulated biofilm formation. *En masse*, our data suggest that ArsRS regulates biofilm development by *H. pylori*, and that the regulation of OMP gene expression is a key contributor to this process.

# TABLE OF CONTENTS

LIST OF TABLES ..... xii

LIST OF FIGURES ..... xiii

CHAPTER ONE: Introduction ..... 14

General Overview of *Helicobacter pylori* ..... 14
Historical Prospective ..... 14
Epidemiology ..... 15
Infection and Disease ..... 16
Genetics of *H. pylori* ..... 17
The *H. pylori* Genome ..... 17
Association Between Disease Outcome and Virulence Factor Polymorphism ..... 18
Outer Membrane Proteins ..... 19
*H. pylori* OMPs ..... 20
Hop (The Major OMP Family) ..... 20
Hom (A Family of Four) ..... 23
*H. pylori* Transcriptional Regulation ..... 30
Metalloregulation in H. pylori ..... 31
Acid Adaptation ..... 33
Biofilm Formation by *H. pylori* ..... 36
A Brief History of Biofilms ..... 36
What is a Biofilm? ..... 37
Medically Important Biofilms ..... 38
*H. pylori* Biofilms ..... 38
Protection During Infection ..... 38
Extra-gastric Biofilms ..... 40
Molecular Mechanisms of Biofilm Development ..... 42
Goals and Specific Aims ..... 45

CHAPTER TWO: Characterization of Key *H. pylori* Regulators Identifies a Role for ArsRS in Biofilm Formation ..... 48

Abstract ..... 48
Importance ..... 49
Introduction ..... 49
Materials and Methods ..... 53
Bacterial Strains and Growth ..... 53
Construction of G27 *ΔfurΔnikR* ..... 55
Construction of G27 *ΔfurΔarsS* ..... 55
Construction of G27*ΔfurΔnikRΔarsS* ..... 56
Construction of G27 ArsR-D52N ..... 56
Construction of the Complemented G27*ΔarsS-arsS*$^{C}$ Strain ..... 57
Growth Curves ..... 58

Biofilm Quantification ..... 59
SEM ..... 60
Motility Assay ..... 60
Bacterial Aggregation Assay ..... 61
Plate Adherence Assay ..... 61
Statistical Analysis ..... 62
Results ..... 63
Growth Characteristics of Fur, NikR, and ArsRS Mutant Strains ..... 63
Loss of ArsRS in G27 Exacerbates Biofilm Formation ..... 67
Loss of ArsRS does not Influence *H. pylori* Motility ..... 78
Enhanced Surface Adherence and Auto-Aggregation in Strains with a Mutated ArsRS System ..... 79
Discussion ..... 82
Acknowledgements ..... 86

CHAPTER THREE: The ArsRS Two Component System Regulates the Expression of the *H. pylori* Outer Membrane Protein, *homB* ..... 89

Abstract ..... 89
Introduction ..... 90
Materials and Methods ..... 94
Bacterial Strains and Growth Conditions ..... 94
Sequencing of *homB* Promoter Regions ..... 95
Construction of G27 *ΔhomB* ..... 98
Construction of *ΔhomB arsRS* Double Mutant Strains ..... 101
Construction of *homB* Over Expression Strains ..... 102
*H. pylori* Growth Conditions for RNA Isolation ..... 103
RNA Isolation, cDNA Synthesis and Quantitative Real-Time PCR ..... 104
Biofilm Analysis ..... 105
Cellular Adherence ..... 106
Aggregation Assays ..... 106
Statistical Analysis ..... 107
Results ..... 109
The *homB* Promoter Region is Highly Conserved Regardless of Locus ..... 109
*homB* Expression is Modulated by Growth Phase and Environmental Cues ..... 110
*homB* Expression is Modulated in Response to pH in an ArsRS Dependent Manner ..... 113
*homB* Contributes to Hyper Biofilm Formation in ArsRS Mutant Strains ..... 115
Deletion of *homB* Decreases Bacterial Adherence and Auto-Aggregation ..... 116
*homB* Over Expression Enhances Biofilm Formation ..... 126
Biofilm SEMs Reveal Dramatic *homB*-Dependent Differences in Biofilm Formation ..... 128
Discussion ..... 133
Acknowledgements ..... 141

CHAPTER FOUR: Discussion ..... 142

Summary of Findings and Relevance ..... 142

Biofilm Formation by *H. pylori* ..... 143
Components of the *H. pylori* biofilm ..... 145
Unraveling *H. pylori* Transcriptional Networks ..... 148
Outer Membrane Heterogeneity ..... 150
Conclusion ..... 153

REFERENCES ..... 155

Appendix A: *Helicobacter pylori bab* Paralog Distribution and Association with *cagA*, *vacA*, and *homA/B* Genotypes in American and South Korean Clinical Isolates ..... 174

Abstract ..... 174
Introduction ..... 175
Materials and Methods ..... 179
Bacterial Strains and Culture Conditions ..... 179
*bab* Genes Genotyping ..... 186
Phylogenetic Analysis ..... 189
Statistical Analysis ..... 193
Nucleotide Sequence Accession Numbers ..... 193
Results ..... 193
Sample Population ..... 194
*bab* Genes Genotyping ..... 194
Associations Between *bab* and *hom* Genes ..... 210
Associations Between *bab*, *cagA*, and *vacA* ..... 211
Associations Between *bab* Genotype and Disease ..... 212
Phylogenetic Analysis of AH and KH ..... 214
Discussion ..... 216
Acknowledgment ..... 222
References ..... 223

Appendix B: *Helicobacter pylori* Outer Membrane Protein, HomC, Shows Geographic Dependent Polymorphism that is Influenced by Bab Family ..... 228

Abstract ..... 228
Introduction ..... 229
Materials and Methods ..... 232
Bacterial Isolates and Culture Conditions ..... 232
*homC* Gene Genotyping ..... 233
Statistical Analysis ..... 234
Results ..... 234
Comparative analysis of HomA, HomB, and HomC ..... 234
*homC* Polymorphism ..... 236
Distribution of *homC* Polymorphisms in AH and KH ..... 239
Association Between *bab* Genotype and *homC* Polymorphism in AH ..... 239
Association Between *bab* Chimera and *homC* Polymorphism ..... 243
Discussion ..... 245
Acknowledgements ..... 247
References ..... 248

# LIST OF TABLES

Table 1. Primers, plasmids, and strains used in this study ..... 54
Table 2 (S1). Comparison of the degree of biofilm formation at each time point ..... 73
Table 3. Plasmids and strains used in this study ..... 96
Table 4. Primers used in this study ..... 99
Table 5. Statistical analysis of biofilm quantification ..... 118
Table 6 (S1-AH). *cagA*, *vacA* and *homA/B* genotyping of 80 American *H. pylori* clinical isolates ..... 180
Table 7 (S1-KH) *cagA*, *vacA* and *homA/B* genotyping of 80 Korean *H. pylori* clinical isolates ..... 183
Table 8. Primer sequences for analysis of *H. pylori bab* genes ..... 187
Table 9 (S3). Phylogenetic coding scheme ..... 190
Table 10 (S2-AH). *bab* genotyping of 80 American *H. pylori* clinical isolates ..... 195
Table 11 (S2-KH). *bab* genotyping of 80 South Korean *H. pylori* clinical isolates ..... 197
Table 12. 'IN'-frame or 'OUT'-of-frame status deduced by the number of CT dinucleotide repeats in the *babB* coding region. ..... 204
Table 13. Significant two-way comparisons of *bab* genotype and other factors in both populations[a] ..... 209
Table 14. Three-way comparisons of *bab* genotype and other factors in both populations. ..... 213

# LIST OF FIGURES

Figure 1. *homA*/*homB* genotype schematic ..... 26
Figure 2. HomA and HomB predicted topology ..... 27
Figure 3. *homA*/*homB* genotyping of clinical isolates from two diverse populations. ..... 28
Figure 4. Characterization of wild-type G27 and isogenic regulatory mutant strain growth phenotypes. ..... 64
Figure 5. Biofilm formation at air-liquid interface. ..... 65
Figure 6 (S1). Phase contrast images of the *ΔarsS* mutant strain. ..... 66
Figure 7 (S2). Representative images of the various biofilm phenotypes. ..... 69
Figure 8. Biofilm quantification. ..... 72
Figure 9. Scanning electron micrographs of biofilm formation after 48 hours. ..... 77
Figure 10. Swimming motility of wild-type G27, *ΔarsS,* and ArsR-D52N strains ..... 80
Figure 11. Auto-aggregation and adherence of the *ΔarsS* and ArsR-D52N mutant strains. ..... 81
Figure 12. Model of ArsRS regulation of colonization factors ..... 87
Figure 13. *homB* locus A and locus B promoter alignment ..... 100
Figure 14 (S1). Bacterial aggregation observed at 40X (100X inset) ..... 108
Figure 16. *homB* expression in response to environmental stimuli ..... 114
Figure 17. Biofilm quantification ..... 117
Figure 18 (S2). Crystal violet staining of biofilm formation ..... 119
Figure 19. Bacterial adherence ..... 120
Figure 20. Bacterial aggregation ..... 123
Figure 21 (S3). Representative scatter plots and gates ..... 124
Figure 22. Evaluation of *homB* expression driven by pFlaA and pUreA ..... 127
Figure 23. Biofilm quantification of pHomB strains ..... 129
Figure 24. SEMs of biofilm formation after 48 hours demonstrate a role of HomB in biofilm formation. ..... 131
Figure 25. Outline of *bab* genotyping by PCR. ..... 188
Figure 26. Distribution of the *bab* genotypes in AH and KH (n=160). ..... 200
Figure 27 (S1). Frequency of *bab* gene by locus ..... 202
Figure 28 (S2). Status of *babB* reading frame. ..... 206
Figure 29. Schematic comparison of BabA, BabB, and BabC. ..... 207
Figure 30 (S3). Alignment of the amino acid sequences of 13 BabCs and 1 BabBC (AH-J68) ..... 208
Figure 31. Phylogenetic analysis of AH and KH ..... 215
Figure 32. Schematic comparison of HomA, HomB, and HomC. ..... 235
Figure 33. Outline of *homC* genotyping in *H. pylori* strains. ..... 237
Figure 34. Genotyping of *homC* variants by PCR. ..... 238
Figure 35. Distribution of *homC* polymorphisms. ..... 240
Figure 36. Schematic depiction of the distribution of the *homC* polymorphism stratified by *bab* locus in AH. ..... 242
Figure 37. Schematic depiction of the distribution of the *homC* polymorphism stratified by the presence of *bab* chimeras. ..... 244

# CHAPTER ONE: Introduction

## GENERAL OVERVIEW OF *HELICOBACTER PYLORI*

### Historical Prospective

*Helicobacter pylori* is a Gram-negative, microaerophillic bacterium that colonizes the gastric mucosa of humans and non-human primates. Although its name was coined within the last 50 years, *H. pylori*'s relationship with mankind extends over millennia; *H. pylori* was recently identified in the gastric contents from a 5300 year old iceman, and numerous studies suggest that *H. pylori* has been associated with man for around 100,000 years (49; 54; 128; 137). Articles detailing the presence of a spiral shaped microorganism in the stomachs of animals as well as in humans date back to the late 19$^{th}$ century (113). However, while descriptions of these spirilla within the stomach appeared sporadically throughout the 1900s (75; 190) due to the low pH of the stomach and the inability to recover any bacteria, these findings were usually dismissed as contamination. Beginning in 1979, J. Robin Warren and Barry Marshall sought to characterize this stomach-associated bacterium that appeared to be related to the severity of gastric disease (131-133) Despite many failed attempts, by happy accident, Warren and Marshall discovered "pyloric *Campylobacter*" on agar plates of gastric specimens that were forgotten in the incubator over the Easter holiday (six days) (133). Around the same time, researchers in the UK were also attempting to culture these pyloric *Campylobacter*-like organisms; in May and June of 1984 the first reports of culturable gastric bacteria were published (133;

136). Following isolation of the pylori *campylobacter*, Warren and Marshall went on to fulfill Koch's postulates by demonstrating association between the presence of this organism with development of gastric disease. Furthermore, in the *coup de grâce,* they showed that when the bacterium was ingested by an otherwise healthy individual (Barry Marshall), bacterial colonization, gastric inflammation, and clinical gastritis occurred (132). Since this seminal discovery, pyloric *Campylobacter*, now known as *Helicobacter pylori,* has been identified as a major cause of peptic ulcers (69), and in 1994 was declared a class I carcinogen by the WHO (1)(IARC 61:1-241).

**Epidemiology**

Remarkably, since its discovery in the 1980s, *H. pylori* has been found ubiquitously in human stomachs across the globe. The worldwide distribution of *H. pylori* is not uniform, with much higher rates of infection seen in developing countries where prevalence can reach rates in excess of 80% (85). In comparison, infection rates in developed areas, such as the United States and Europe, remain around 30% (85). The advent of multi-locus sequence typing (MLST) allowed researchers to identify six geographically distinct *H. pylori* strain types that varied across populations (71; 143). Interestingly, the geographic diversity of *H. pylori* strains has been shown to follow predicted human migration patterns (30; 71; 117; 144). These distinct lineages bring to mind the founder effect; whereby distinct *H. pylori* strains dispersed within humans as they populated different regions of the globe, culminating in geographically distinct subpopulations. Although a few environmental reservoirs for *H. pylori* have been proposed, no definitive environmental niche has been identified to date (18). Indeed, the

lack of an environmental source may be a contributing factor to the close association between *H. pylori* subpopulations and human migration.

**Infection and Disease**

*H. pylori* is a highly successful pathogen, as demonstrated by its global distribution and colonization of 50% of the world's population. Even when considered from the perspective of an individual infection, *H. pylori* is impressive. Epidemiological studies suggest *H. pylori* is commonly acquired during childhood, and acquisition likely occurs via a fecal-oral or oral-oral route (69). While the mode of transmission for *H. pylori* has not been demonstrated directly, there are two lines of evidence that support the aforementioned hypotheses: families tend to carry related strains of *H. pylori,* and a higher prevalence of *H. pylori* infections correlates with crowded living conditions that are common in developing countries and in lower socioeconomic groups in developed countries (142). Once *H. pylori* establishes an infection, it can persist for the lifetime of an individual without specific antimicrobial intervention. Following the paradigm of a persistent infection, the chronology of *H. pylori* disease progression is recorded in years rather than in days or weeks. Acute *H. pylori* infection is characterized by infiltration of polymorphonuclear cells (PMNs) into the gastric mucosa (50; 63). Progression to chronic infection is marked by the presence of adaptive immune cells and occurs in all individuals successfully colonized by *H. pylori* (50; 63). In a majority of individuals, chronic inflammation will persist but never progress to clinical disease. In fact, if *H. pylori* clinical manifestations arise, this typically occurs after decades of infection, and a majority (80%) of cases remain asymptomatic. Unfortunately, for approximately 20% of

individuals, infection with *H. pylori* will progress to clinical gastritis, gastric or duodenal ulcers, gastric adenocarcinoma, or mucous associated lymphoid tissue (MALT) lymphoma (113). A major unanswered question in the *Helicobacter* field is why some cases remain quiescent and others develop into severe disease such as gastric cancer.

Epidemiology presents a classic model by which researchers can assess the factors that contribute to infectious disease: environment, host, or pathogen (31). *H. pylori*, is the strongest known risk factor for the development of gastric cancer (55). Indeed, almost 90% of non-cardia gastric carcinomas are attributed to *H. pylori*; this translates to 35.4% of all cancer attributable to an infectious agent, more than any of the other ten infectious class I carcinogens (165). The high level of genetic heterogeneity of *H. pylori* is well documented (193; 194) and thus, presents a very attractive target for unraveling the molecular identity of bacterial factors that influence disease outcome.

## GENETICS OF *H. PYLORI*

### The *H. pylori* Genome

As a bacterium that has adapted to life in a highly specified niche, i.e. the stomach of humans and non-human primates, it may not be surprising that *H. pylori* has a relatively small genome (1.6-1.7Mb). Interestingly, despite the reduced genome size, *H. pylori* displays extensive intraspecies genetic diversity. The *H. pylori* genome encodes between 1,400 and 1,600 genes (6; 7; 65); however, a core genome of only 1,111 genes was deduced based on compiled microarray data that compared over 60 *H. pylori* isolates, (65; 93; 172). This disparity suggests that approximately 25% the of *H. pylori*

genome is comprised of strain-specific "auxiliary" genes (65). Among genes conserved in the core genome are those involved in metabolic, biosynthetic, cellular, and regulatory functions (172). In comparison, the auxiliary group of genes is comprised of genes encoding for proteins with unknown or hypothetical functions, components of the outer membrane, including lipopolysaccharides (LPS) and numerous outer membrane proteins (OMPs), members of the Cag pathogenicity island, and proteins involved in DNA metabolism and modification (93; 172).

**Association Between Disease Outcome and Virulence Factor Polymorphism**

Many of the same genes that can be classified as auxiliary genes have been shown to be highly polymorphic. For example, *H. pylori* carries genes for two traditional virulence factors: cytotoxic gene A (*cagA*) and vacuolating cytoxin A (*vacA*). Both of the cytotoxins encoded for by these genes play a role in modulation of host cell signaling pathways, intoxication, and host cell death (106). Additionally, specific polymorphisms within both of the genes have been associated with variable disease outcomes. CagA positive strains have been shown to pose an increased risk of severe disease as compared to CagA negative strains; however, among CagA positive strains, polymorphisms within the C-terminal EPIYA motif are associated with different levels of risk (31). Similarly, polymorphisms within three regions, designated i/m/s, of the *vacA* gene are associated with differential disease outcome (31). In addition to these cytotoxins, several members of another class of proteins, OMPs, have also been associated with differential disease outcome (55). Although not traditionally viewed as virulence factors, several OMPs are

thought to contribute to *H. pylori*- associated disease pathology (59; 127; 200; 214; 215) (156).

**Outer Membrane Proteins**

*H. pylori* encodes for an abundance of outer membrane proteins (OMPs). The OMPs of Gram-negative bacteria interface with the environment and provide various functions that include protection, solute and protein translocation, and signal transduction (111; 179). Gram-negative OMPs are unique, because unlike other membrane proteins that consist of α-helices, OMPs are made up of parallel β-sheets that fold into β-barrels (111; 179). In many cases, β-barrel OMPs function as porins. These can either be general porins, which accommodate a variety of substrates based on size and charge, or substrate-specific porins. Substrate-specific porins can be used for transport of small molecules, such as sugars, or for large complexes, such as those required for metal transport (179). In addition to functioning as pores that allow for the translocation of solutes and proteins across the outer membrane, bacterial defense, and enzymatic activity have also been attributed to some OMPs (111; 179; 212). *In silico* predictions suggest that two to three percent of a typical Gram-negative bacterial genome encodes for β-barrel OMPs (212). For a bacterium such as *Haemophilus influenzae*, which has a relatively small genome, this corresponds to approximately 40 genes (212). Although *H. influenzae* and *H. pylori* have similar sized genomes, *H. pylori* is predicted to encode more than 60 OMPs (5; 7). Given that the total *H. pylori* genome consists of around 1500 ORFs, genes coding for OMPs make up approximately 4% of the coding potential (5; 7).

### *H. pylori* OMPs

The *H. pylori* genome includes a wide-array of OMPs, many of which are proposed to function as adhesins. For an organism that sustains prolonged colonization within a single host, the repertoire of adhesins likely contributes to the adaptability of the strain. Throughout the course of infection, *H. pylori* likely encounters variable receptors and needs to quickly deploy the appropriate bacterial counterpart. Furthermore, it is plausible that specific sets of receptors are needed for transmission, to establish infection, and to maintain colonization in the face of pathological changes at the gastric surface and robust host defenses. To date, we have only begun to understand the phenotypic variation of *H. pylori*'s adhesins. Many questions still remain concerning the role of specific adhesins and their necessity during the course of infection. Indeed a greater understanding of the molecular basis of *H. pylori* adherence throughout the course of an infection is the next task for researchers.

### *Hop (The Major OMP Family)*

In comparison to other Gram-negative pathogens that typically encode for a highly abundant, nonselective porin such as OmpF (179; 212), *H. pylori* appears to harbor numerous less abundant OMPs (64). Based on C-terminal homology, these OMPs are divided into six paralogous families: Hop (21), Hor (12), Hof (8), Hom (4), iron-regulated OMPs (6), and efflux pumps (3). Furthermore, there are approximately 10 additional orphan OMPs (5). Similar to what is seen in other Gram-negative bacteria, β-barrel proteins that function as transporters have been identified in *H. pylori* and include those involved in ferric iron complex uptake: the FecA-like and FrpB-like transporters

(149). Initial characterization of five members of the Hop family (HopA through E) showed that these porins function as nonspecific water-filled channels that display little selectivity. Based on conductivity studies, the channels are predicted to be more similar to the antibiotic resistance porins of *Pseudomonas* spp rather than to porins of the closely related *Campylobacter* spp (64; 149).

In addition to the more traditional porin roles, a major role attributed to *H. pylori* OMPs is adherence. Two of the original five characterized porins, HopB and HopC, also known as AlpA and AlpB, respectively, were among the first adhesins identified (123; 151; 181). Characterization of AlpA and AlpB showed that these lipoproteins interact with host laminin and contribute to adherence and modulation of the host pro-inflammatory signaling cascade (123; 151; 181). The importance of AlpA and AlpB is highlighted by animal infection models which demonstrate the essentiality of both OMPs for efficient *H. pylori* colonization (reviewed in (156)). Furthermore, analysis of 200 *H. pylori* isolates determined that 100% of strains were positive for AlpA and B, as compared to 35%-73% positivity for other tested OMPs (150). Additional members of the Hop family were subsequently determined to act as adherence factors in *H. pylori*: the Bab paralog, BabA (HopS); the Sab paralog, SabA (HopP); OipA (HopH); HopZ; and HopQ (reviewed in (156)).

BabA is one of the best characterized *H. pylori* OMP adhesins. BabA has been shown to bind fucosylated Lewis b histo-blood antigen (Leb and H1), terminal fucos residues on blood group O, A and B antigens, and certain salivary glycoproteins (reviewed in (156)). Interestingly, *H. pylori* isolates can encode up to three Bab paralogs: BabA, BabB, and/or BabC; numerous combinations of *bab* paralogs have been reported

(15; 97; 110; 166). Although these proteins are highly similar, the functions of BabB and BabC in adherence have yet to be demonstrated. Similarly, two *sabA* paralogs, *sabB* and *hopQ*, have also been identified (198). While SabA is known to bind sialylated structures present on extracellular matrix proteins, erythrocytes, and neutrophils, the role of SabB in adherence has yet to be demonstrated. Moreover, two studies recently showed that HopQ binds CEACAM1, 3, 5, and 6 with high affinity; the association between HopQ and host cells effects CagA translocation into the host cell (26; 103; 112; 122). It is interesting that a bacterium with a reduced genome would harbor this level of redundancy in OMPs, raising questions about the potential importance of harboring paralogous OMPs.

While we often focus on the genetic variability between strains, it is important to remember that phenotypic variability can occur between strains as well as within subpopulations of the same strain (70). Interestingly, variations in both SabA and BabA copy number and genomic location are observed between strains (97; 109; 110; 166; 184; 198). Genomic rearrangement of the *bab* genes has been observed in animal models (118; 119; 187). In many cases, these rearrangements led to loss of *babA*, suggesting that *babA* may be dispensable or detrimental during chronic infection (118; 119; 187). Conversely, while equal rates of gene conversion between *sabA* and *sabB* were reported *in vitro*, the presence of *sabA* was maintained in at least one locus in a collection of clinical isolates (198). Also of note, the presence of *hopQ* at the *sabA* and *sabB* loci has been observed in clinical isolates. However, these strains all carried duplicate copies of *hopQ,* with one copy at the *hopQ* locus as well (109; 198). Furthermore, *hopQ* gene conversion with the *sabA* or *sabB* locus was not observed *in vitro* (109). *En masse*, these observations suggest

that selective pressures encountered during infection likely influence gene conversion or selection of specific genotypes.

It is also important to note that many of the OMPs are predicted to be modulated by slip-strand mispairing (SSM): SabB, BabB, BabC, HopZ, and OipA. Roughly 30 genes are predicted to be subject to phase variation, and a majority of these genes are located at the cell surface (173). Given the prolonged colonization timeframe, the ability to quickly generate diversity between strains is likely paramount to the success of this pathogen and may also contribute to the differences in disease outcome. For instance, *in vitro* analysis of clinical isolates found that BabA was expressed in approximately 60% of clinical isolates, 40% expressed SabA, and 33% appeared negative for both BabA and SabA (150). Specifically in regard to SabA, these findings seem contradictory to the 100% maintenance of *sabA* presence observed in some studies (198). It is plausable that phase variation enables *H. pylori* to easily maintain subpopulations that allow for rapid adaptation; in a small percentage of bacteria, a combination of OMPs remain in the phase "ON" state, while in the majority of the population these OMPs remain "OFF." (3; 96; 118; 173; 187).

***Hom (A Family of Four)***

While much attention has been devoted to the investigation of proteins in the Hop family, a few other families have been studied to a lesser extent. In particular, a member of the smallest OMP family, the Helicobacter Outer Membrane protein family (Hom), was identified as a potential *H. pylori* adherence factor (152). The Hom family of proteins is encoded by four genes, *homA*, *homB*, *homC* and *homD* (5). This family first

came to the attention of the *Helicobacter* field when HomB was identified in a suppressive subtractive hybridization screen designed to identify markers of strain virulence (157). In that study, an isolate from a five year old male with a duodenal ulcer (DU) was compared to an isolate from an age/sex matched control who only had gastritis; *homB* was among the genes specific to the DU strain (157). Further analysis of the presence of HomB among children and adults suffering from peptic ulcer disease (PUD), revealed a significant association between PUD and HomB presence in children but not in adults (152; 157). Given that a majority of *H. pylori*-associated disease typically occurs in the later decades of life (22), one possible explanation for this difference could be that these genes correlate with increased virulence in children that is obscured by other environmental or host factors in the adult population.

Although *homB* was found in association with more severe disease, a role for this putative OMP was unknown until a series of *homB* deletion strains were constructed in clinical isolates (152). For the majority of the clinical isolates tested, deletion of *homB* led to a decrease in adherence to a gastric adenocarcinoma cell line (AGS) (152). Furthermore, in a strain that naturally carried two copies of *homB,* deletion of both copies led to a significant decrease in adherence as compared to a single *homB* deletion in the same strain (152). Similarly, deletion of *homB* caused reduced IL-8 production by AGS cells, despite the fact that the strains associated with reduced IL-8 did not always correspond to decreased adherence (152). Finally, pooled serum from *H. pylori* positive individuals was shown to be reactive against both native and recombinant HomB, supporting *in vivo* expression and a potential role for *homB* during *H. pylori* infection (152).

Follow-up studies showed that another Hom family member, *homA,* could sometimes be found at the same locus where *homB* should be found (154). Reminiscent of the Bab and Sab family of paralogs, it was shown that numerous *homB*/*homA* genotypes exist (Figure 1). As such, *hom* locus A and *hom* locus B can be occupied by *homA* or *homB* or can remain empty. In addition to sharing loci, *homA* and *homB* also share approximately 90% nucleotide identity (5). Based on the *in silico* β-barrel prediction software (PRED-TMBB) the regions of disparity between HomA and HomB are predicted to reside in environmentally exposed loops (Figure 2). A difference in environmentally exposed loops could suggest that these two proteins have different functions or were selected for in response to immunologic pressure. This is of particular interest given that the presence of *homB* has been associated with PUD, while *homA* has been associated with non-ulcerative disease (NUD) (152; 155). Given the interest in identification of genetic markers of *H. pylori* virulence, the connection between *homB* and disease was evaluated in numerous populations (100; 108; 155; 199).

*En masse*, these studies showed that associations between *homB* and disease outcome varied depending on the population studied (100; 108; 155; 199). Interestingly, while an association between *homB* and gastric cancer was observed in individuals from Iran, a similar study in patients from neighboring Turkey and Iraq found no association (100; 199). One observation that arose from the epidemiological investigations was a geographic disparity in the distribution of *homA* and *homB* (Figure 3) (154). Perhaps, the variability in *homA* and *homB* genotypes may explain why the role of *homB* as a marker of severe disease varies based on the population being studied. These findings are strikingly similar to studies of other genes, including OMPs such as *babA*, and the

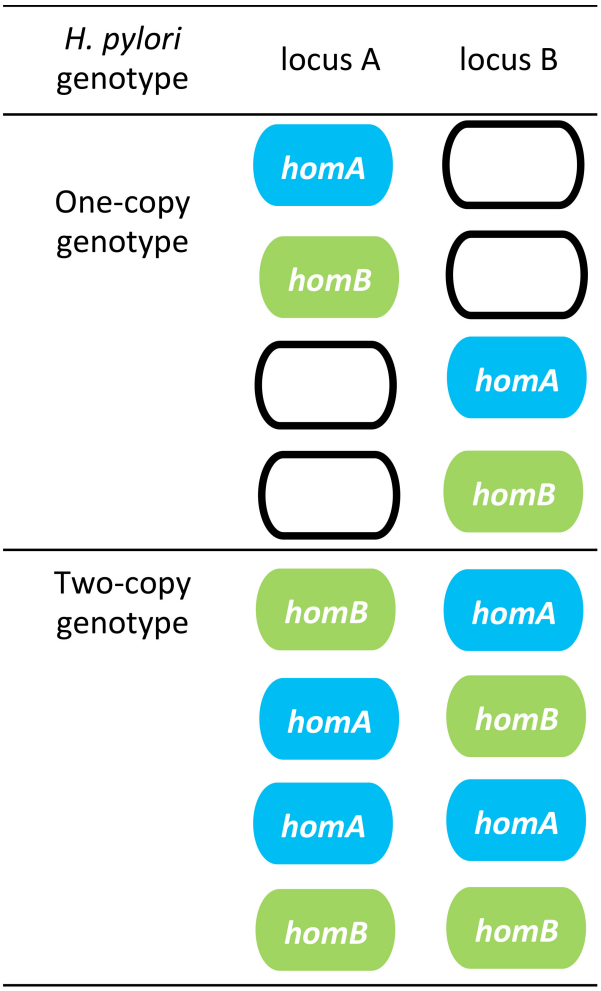
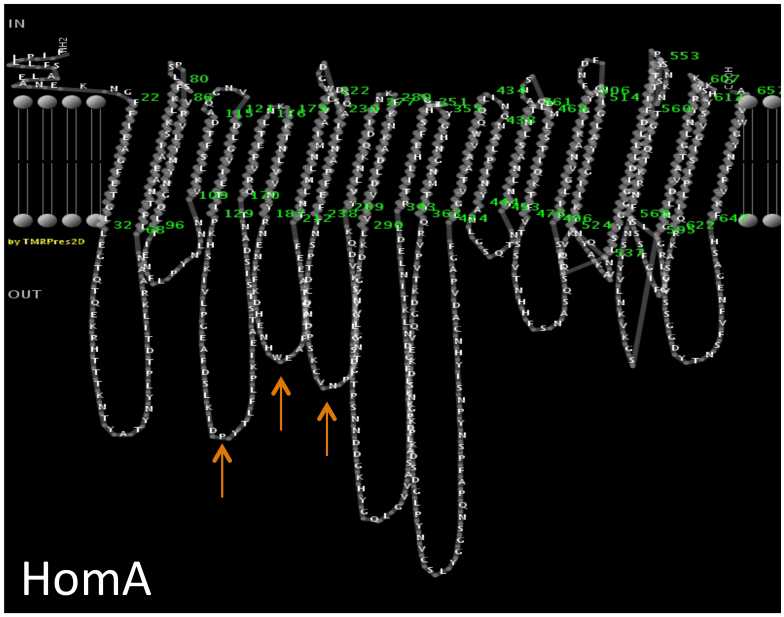



**Figure 1. *homA*/*homB* genotype schematic.**
The various iterations of *homA* and *homB* genotypes are depicted in the schematic. The genotypes indicated above represent the *H. pylori hom* genotypes that have been identified to date; each row corresponds to a single genotype; empty ovals indicate the absence of a gene at the corresponding locus.



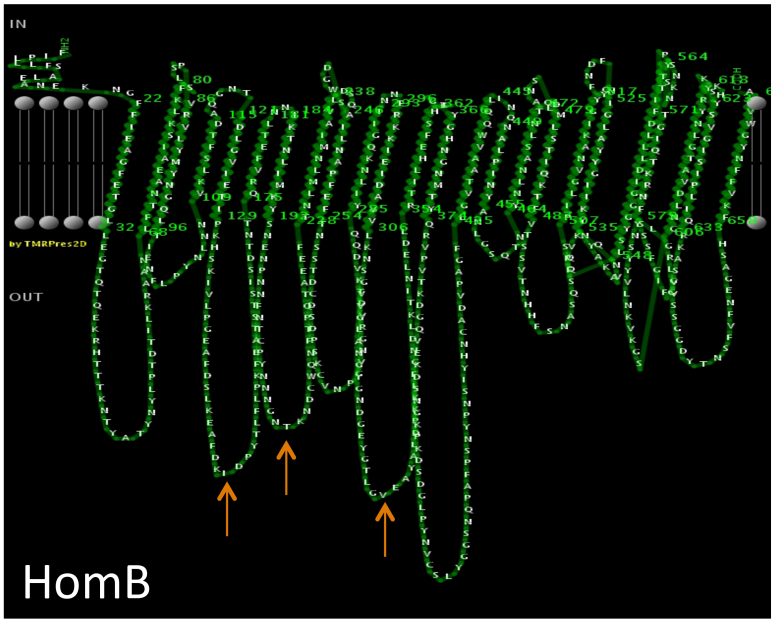




**Figure 2. HomA and HomB predicted topology.**
HomA and HomB secondary structure was predicted using the PRED-TMBB prediction program. The periplasm is located at the top of each image; environmentally exposed loops are below the outer membrane. Orange arrows indicate regions where the amino acid sequence differs between HomA and HomB.

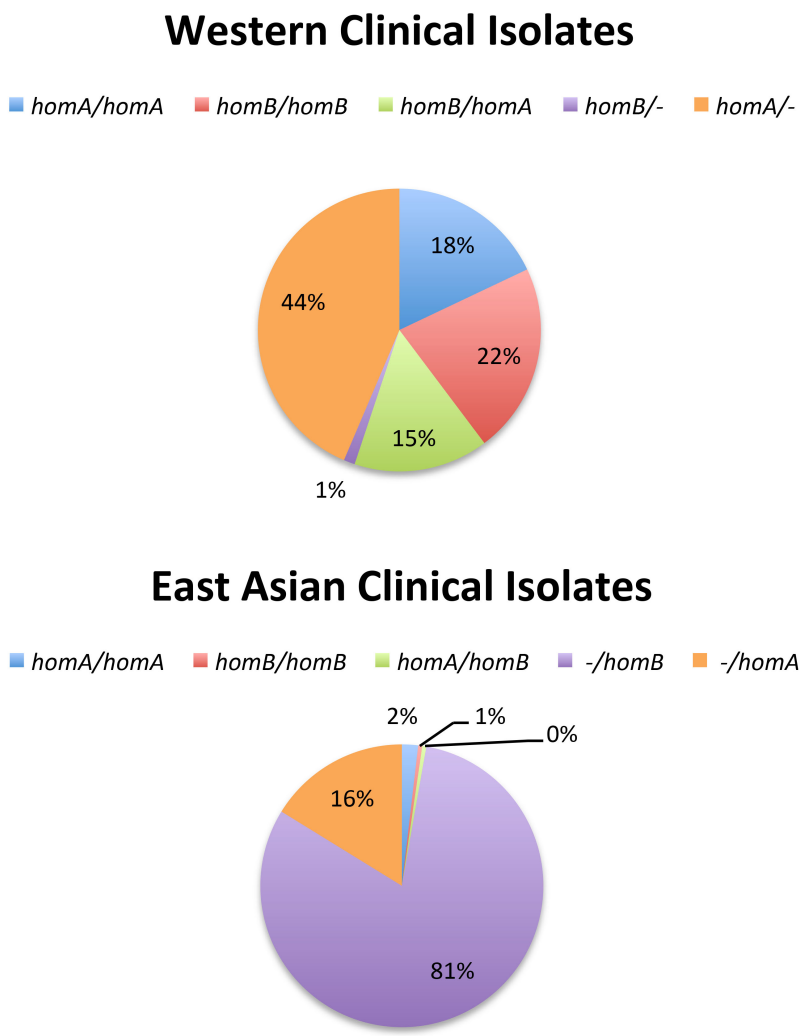



**Figure 3. *homA/homB* genotyping of clinical isolates from two diverse populations.** PCR genotyping of *hom* locusA and *hom* locus B was conducted on recent American (n=78) and South Korean (n= 216) *H. pylori* clinical isolates. The genotypes are listed as locusA/locusB and are presented as a Venn diagram for the individual populations. The relative percentage of each genotype is indicated with each population.

cytotoxins, *cagA* and *vacA*, that revealed difference in genotype that correlate to the geographic origin of the strain (2; 4; 102; 110; 184; 223).

Two major *H. pylori* geographic strain groupings, identified by Cag type, are Western and East Asian. Our group and others have found that East Asian strains typically carry a single-copy of *homA* or *homB* at locus B; a majority of these strains carry a single *homB* (108; 154). In contrast, Western strains are more diverse and carry two copies of the *hom* gene in numerous iterations: *homA/homA, homA/homB, homB/homA, homB/homB,* corresponding to locus A/locus B respectively. Both *hom* loci are occupied in roughly 40% of Western isolates; the other 60% of Western isolates carry a single *homA* or *homB* (108; 110). However, in contrast to the single-copy genotypes in East Asian strains, Western strains carry the single *hom* gene at locus A (Figure 3) ((108; 154) unpublished Merrell). Of note, when multiple *hom* copies are considered, carriage of two copies of *homB* was more frequently observed in PUD patients and was the rarest genotype in NUD individuals (155).

A majority of the research to date has focused on *homA* and *homB*; less is known about the other two members of the Hom family, *homC* and *homD*. While HomA and HomB share 90% identity to one another, comparison of all four Hom proteins from the same strain reveals that HomC and HomD are more divergent from HomA (52%, 48% identity, respectively) and HomB (51%, 47% identity, respectively) ((5) , Merrell and Servetas unpublished). Additionally, the relationship between HomC and HomD is not analogous to that of HomA and HomB; HomC and HomD share only 75% identity and occupy separate distinct loci (Merrell and Servetas unpublished). One interesting feature

of both *homC* and *homD* is that they appear to be universally present in all strains tested to date; this is in stark contrast to many of the OMPs discussed previously (48; 138; 150). Furthermore, while multiple variants of *homC* exist, analysis of *homD* sequence from 200 isolates suggests that it is highly conserved (48; 138).

Another area that has yet to be evaluated for the Hom family is gene regulation. Indeed, little is known about the expression of the Hom family of OMPs. Based on immunologic evidence, HomB is thought to be expressed during infection, and a decrease in *homB* expression was reported following passage through an animal model (74; 152). In comparison to several members of the Hop family, members of the Hom family do not appear to be subject to phase variation ((173), Servetas and Merrell unpublished). It is possible that *homA* and *homB* are subject to gene conversion given the high degree of similarity between these two genes and the observation that they occupy the same loci; however, gene conversion between *homA* and *homB* has yet to be evaluated *in vitro*. Interestingly, HomD appeared in screens for ArsRS regulated factor, and direct interaction between ArsR and the *homD* promoter has been demonstrated (121). Given that the bacterial outer membrane provides the interface between the environment and the bacterium, it is not surprising that pH fluctuation could modulate OMP expression in this gastric pathogen. In addition, ArsR regulation of SabA and LPS modification in response to pH 5 have been demonstrated (86; 96; 135). Taken together, these findings may reflect more global changes in the outer membrane of *H. pylori* in response to changing pH.

## *H. pylori* Transcriptional Regulation

In direct contrast with the extensive gamut of OMPs, one striking genetic feature uncovered during dissection of the *H. pylori* genome was the low abundance of predicted regulatory proteins. For comparison, *H. pylori* is predicted to encode one tenth of the regulatory proteins of *E. coli* K12. This is less than half of the regulatory proteins encoded by *H. influenzae,* which has a similar sized genome. Furthermore, even the closely related *Campylobacter jejuni* is predicted to have a larger repertoire of regulatory proteins (60; 72; 160). This disparity in abundance of regulatory proteins between *H. pylori* and bacteria with similarly sized genomes may reflect the differences in the ecological niches of each pathogen. In contrast to *H. pylori,* which is restricted to the gastric niche of humans and non-human primates, *H. influenzae* and *C. jejuni* each occupy broader host ranges and environmental reservoirs. Despite the restricted niche, *H. pylori* is still exposed to stressors such as nutrient deprivation, oxidative stress, host responses, and extreme acidity. Some of the prototypical Gram-negative regulatory proteins that are missing in the genome of *H. pylori* are the two stress-related sigma factors: *rpoH* and *rpoS*. Additionally, there are far fewer two-component systems found in the genome of *H. pylori*. *H. pylori* encodes five response regulators and three sensor kinases, while *C. jejuni* encodes 12 two-component regulators and seven sensor kinases, (160; 175). Despite the lack of regulatory proteins, *H. pylori* is highly successful at colonization and persistence within the gastric niche. This fact suggests that it may employ novel regulatory mechanisms, networks, or proteins to coordinate the required environmental adjustments.

### ***Metalloregulation in* H. pylori**

Maintenance of metal homeostasis is one of the most important regulatory tasks for all bacteria. Transition metals such as nickel, iron, copper, and zinc are essential components for many biological processes, yet are toxic if high levels accumulate within the cell. (23; 94; 169). The bioavailability of metals is particularly relevant to the discussion of host-pathogen interactions since both eukaryotes and prokaryotes require metal cofactors for essential cellular processes (23). In fact, the ability of the host to sequester metals such as iron is a form of defense known as "nutritional immunity" (207). Given the limited number of regulatory proteins found in *H. pylori* and the prominent role of metal acquisition in bacterial pathogenesis, it is perhaps not surprising that two metalloregulators, Fur and NikR, have taken on more global roles in *H. pylori* transcriptional regulation (56; 58).

Iron is a central player in the battle for metal that is waged between host and pathogen, and the bacterium's ability to monitor and respond to environmental iron is paramount to maintaining proper iron homeostasis. The ferric uptake regulator (Fur) fulfills this regulatory role for *H. pylori*. Fur is a small (15-17kDa) protein that is highly conserved across bacterial species (38). Of note, functional complementation between Fur derived from different species has been demonstrated (141). The breadth of *H. pylori*'s Fur regulon is exemplified by its multiple mechanisms of action. In addition to functioning as a classical holo-Fur repressor, *H. pylori* Fur has also been shown to function in holo-Fur activation, apo-Fur activation, and apo-Fur repression (38). Accordingly, approximately 200 genes are predicted to be encompassed within in the Fur regulon in *H. pylori* (57). These include genes that respond to iron availability (ex: *fecA*, *frpB*), low pH (ex: *amiE*), and oxidative stress (ex: *sodB*, *napA*), as well as genes

involved in motility/chemotaxis (ex: *flaA*, *tlpB*) and virulence (ex: *cagA*, *vacA*) (43; 57; 67; 78). A number of regulatory proteins found in the Fur regulon further elaborate its central role. In addition to being autoregulatory, *H. pylori* Fur has also been shown to regulate the expression of: *rpoN*, *flgR*, *flgS*, *cheA*, *hspR*, *nikR*, and *arsR* (43; 57; 67; 78). In many ways Fur fits the description of a regulatory hub. The pleotropic nature of Fur regulation demonstrates one way in which *H. pylori* has adapted to life with a limited number of regulatory proteins.

Fur is not the only regulator that has taken on a more pluripotent role in *H. pylori*. NikR, a nickel dependent regulatory protein, also has a rather large and diverse regulon. NikR regulates numerous factors involved in the acid response (ex: *ureA*, *ureB*), stress response (ex: *dnaK*, *grpE*), motility/chemotaxis (*flaA*, *cheV*), outer membrane proteins (ex: *hopA*, *horE*), metal transport proteins (*nixA*, *ceuE1*), and regulatory proteins (ex: *fur*, *hspR*) (46; 68; 204). There are several genes that are members of both the Fur and NikR regulons (32; 56; 58; 61; 175). In addition, while Fur and NikR are known to be auto-regulatory, they also can repress the expression of each other (56; 58; 61; 175). Although we have only examined two of the 17 known *H. pylori* regulators at this stage of our discussion, a complex regulatory network is beginning to emerge.

***Acid Adaptation***

Another vital component of *H. pylori*'s existence within the gastric niche is its ability to acclimate and grow despite the acidic conditions. Despite its survival in an environment where the pH drops to values of ~1 during periods of starvation, *H. pylori* is not an acidophile (186). In fact, *H. pylori* prefers to replicate at a pH between 6.5-7(87).

Other neutrophiles, such as *Vibrio cholera*, *Salmonella typhimurium*, and pathogenic *Escherichia coli* species, have mechanisms that resist or tolerate acid-shock long enough for the bacteria to transit though the gastric lumen to the intestine (73; 210). While *H. pylori* also employs mechanisms to counteract extreme acid shock, for this bacterium this acid adaptation is a constant battle that requires coordination of numerous processes. Indeed, these coping mechanisms of *H. pylori* may be better described as acid acclimation (208). Not surprisingly, there are between 100-200 genes that are predicted to be regulated in response to low pH (32; 139; 164; 210), thus accounting for between 6-12% of the predicted coding regions (163; 193). Three main regulators are responsible for coordination of the acid-acclimation response: the two previously mentioned metalloregulators, Fur and NikR, and the acid-response two-component system, ArsRS ((28; 32; 78; 120; 163; 204).

Since bioavailability of metal ions is associated with low pH, and NikR and Fur are pleotropic regulators, it is not surprising that these metalloregulators play a role in acid-adaptation in *H. pylori* (32; 78; 129; 180; 204). In fact, broad scope microarray and transcriptome analyses show overlap between the Fur, NikR and acid adaptive regulons (58). Not surprisingly, Fur and NikR are both required for efficient colonization in animal studies, and loss of Fur and NikR leads to impaired acid-tolerance (204; 205). (32).

The other key players in acid regulation are ArsRS. Named for their role in acid responsive signaling, *arsRS* encodes a prototypical two-component system. *arsS* encodes the histidine sensor kinase that autophosphorylates in response to low pH (pH<6.2) (162). It is believed that the sensory domain of ArsS resides in and responds to the environment in the periplasm (130; 208). As such, acidification of the periplasm results in auto-

phosphorylation of ArsS. ArsS then trans-phosphorylates the aspartic acid at residue 52 of ArsR, which is an OmpR-like response regulator. The phosphorylation of ArsR results in an increased affinity for a specific set of promoters (163; 208). This system is so critical to the survival of *H. pylori* that ArsR is essential for growth *in vitro*, and *arsS* deficient mutants are unable to colonize mice (158). While ArsR cannot be deleted, point mutations that lead to substitutions of the phospho-accepting residue of ArsR are not lethal. This suggests that ArsR has two sets of targets: essential genes targeted by ArsR and non-essential genes targeted by ArsR~P.

Relatively little is known about the essential processes that ArsR regulates; however, several groups have endeavored to unravel the pH-dependent ArsRS regulon. These studies suggest that the ArsRS regulon is comprised of between 100 and 125 genes (121; 164; 208). While an initial distinction between the ArsR and ArsR~P regulons was made, subsequent studies revealed that ArsRS transcriptional regulation can be further delineated based on pH levels (210). For example, studies suggest that ArsR acts as an auto-repressor in response to low pH (163). Despite the fact that *arsS* is the second gene in an operon with *arsR*, *arsS* appears to be upregulated at a pH of 6.2, whereas *arsR* expression does not show an increase until the pH of the media is decreased to pH 4.5 (210). These findings suggest that the regulation of the *arsRS* regulon may not be a simple negative-regulatory feedback loop. In support of this theory, an elegant study found that Fur acts as a repressor of *arsR* through metal-dependent DNA compaction (170). It was also observed that NikR prevents DNA compaction by antagonizing holo-Fur. In total, three Fur-binding boxes (two which are shared between Fe- and apo-Fur), two NikR-binding regions, and an ArsR-binding sequence were identified upstream of

the *arsR* coding region (170). Together, these data not only suggest that expression of ArsRS is under careful modulation that is initiated by a variety of stimuli other than pH, but also showcase the intricate mechanism by which *H. pylori* can fine tune gene expression with a handful of regulatory proteins. Coordinated regulation is not restricted to just ArsRS, but has also been proposed to occur between Fur and NikR at the *fur*, *nikR*, and *exbB* promoter regions (61). In addition, coordinated regulation between ArsR and NikR at the *ureA* promoter was recently demonstrated and has been proposed for the *oorD* locus as well (37; 83). The overlapping regulatory networks and the use of coordinated regulation allows for fine-tuning of gene expression.

## BIOFILM FORMATION BY *H. PYLORI*

### A Brief History of Biofilms

While a concerted effort in biofilm research began in the late 1970s, the observation that bacteria reside within bacterial communities is centuries old. In fact, Antonie van Leewenhoek commented on the presence of "animalcules" in plaques on teeth in 1623 (reviewed (174)). In the 1920's and 1930's, observations regarding the association of bacterial slime with surfaces present in aquatic environments were common (13; 98; 99; 224). Scientist observed that microorganisms in water were not free floating, but were most commonly attached to a surface (98; 99; 224). For example, in the 1920's a report from the United States Navy Department discussed the biofouling of ships by bacteria (13). In addition to general marine environments, the development of biofilms in water treatment systems and the resistance of these biotic mats to

antimicrobials was also under investigation (42; 105). Many names were used in the early studies of the bacterial communities that we now know as biofilms, but it appears that this term was first used in a publication in 1975 by Mack *et al* to describe "Microbial film development in a trickling filter." (125)

**What is a Biofilm?**

In the simplest terms, a biofilm describes a community of microorganisms that is attached to a surface (148). The microbial community can be a single species, multi-species, or even multi-kingdom. The extracellular matrix that enshrouds these communities typically distinguishes a biofilm from other sessile lifestyles. The importance of the microbial biofilm is demonstrated by its presence in the fossil record dating back over three million years (95). Today, biofilms have been identified in a myriad of ecological niches, and the following global steps in biofilm development have been observed: attachment of bacteria to a surface, expansion or aggregation of bacteria on the surface, and increased production of exopolysaccharides (EPS) that forms the biofilm matrix. Supporting the theory that biofilms form by similar mechanisms in different bacterial species, genes that have been demonstrated to be important in biofilm formation tend to fall into a few categories: adherence, quorum sensing, cell wall synthesis, stress response, carbohydrate metabolism, and cell division (104). Additionally, numerous two-component systems, which provide a key interface between the environment and gene expression, have also been shown to be instrumental in biofilm formation by several species (25; 88).

## Medically Important Biofilms

A majority of the early biofilm work focused on environmental biofilms; however, the importance of medically relevant biofilms that was initially described in the 1600s was revived in 1978 (51). This area of biofilm study has received an increasing amount of attention, especially the impact on infectious disease. Biofilm formation has been shown to be a contributing factor for many bacterial diseases, including native valve endocarditis, osteomyelitis, dental carries, ear infections, and chronic lung infections in cystic fibrosis patients. In addition to the formation of biofilms on biotic surfaces, the formation of biofilms on abiotic devices such as ocular implants, catheters, and other medical devices account for a huge burden of infection (reviewed in (95)). From the host perspective, the formation of bacterial biofilm complicates clearance of the infection; biofilm-associated bacteria demonstrate an increased tolerance to the host immune response as well as to antimicrobials (reviewed in (95; 148)). In fact, biofilm formation likely helps to sustain many chronic infections.

## *H. pylori* Biofilms

### *Protection During Infection*

The ability of *H. pylori* to form biofilms may contribute to the longevity of these infections. Indeed, biofilms are often associated with chronic infectious diseases since they provide an additional level of protection against environmental, host, and antimicrobial assaults (reviewed in (95; 148)). Following infection with *H. pylori,* individuals can remain colonized for their lifetime without proper intervention. Even in those individuals who seek treatment, the treatment failure rate is between 20-30% (206;

222). Given these characteristics, it seems plausible that biofilms may play a role in *H. pylori* infection. Despite this, a PubMed search of "*H. pylori* and Biofilms" conducted in the 3rd week of October 2016 revealed fewer than 100 published articles on this subject. Nearly 20 years ago, biofilm formation by *H. pylori* was first offered as a mechanism of increased antimicrobial resistance (89). *H. pylori* biofilms have since been observed in association with the gastric mucosa of infected humans (33; 39; 41; 52) and this phenomenon was documented in separate reports regarding patients suffering from PUD (52) and dyspepsia (41). Furthermore, biofilm formation was recently demonstrated in *H. pylori* infection of C57Bl/6J mice (17).

During infection, biofilms are believed to protect bacteria from antibiotics by limiting diffusion, dampening the antibiotic effect, allowing time for expression of resistance genes, or perhaps by secreting antibiotic degrading enzymes within the matrix (80). The high rate of treatment failures in *H. pylori* infections is well documented (202; 206; 222). With an 80% intention-to-treat eradication rate, the current treatment strategies are at the minimum acceptable level based on the Maastricht guidelines (34). Additionally, *in vitro* antibiotic susceptibility testing does not always accurately predict *in vivo* results (84). While there are several factors that may contribute to the decreased *in vivo* antibiotic susceptibility for *H. pylori*, one possible explanation is that *H. pylori* reside in a protective biofilm in the stomach. A recent study supported this theory and showed that biofilm associated *H. pylori* displayed increased clarithromycin resistance (217). The increased level of resistance observed in biofilm-associated *H. pylori* is likely due in part to an increased expression of efflux pumps compared to planktonic bacteria (217). Amazingly, bacteria within the biofilm showed an increased rate of point

mutations in the 23s rRNA gene, which resulted in the generation of high level clarithromycin resistance (217).

The protection imparted by the *H. pylori* biofilm has been shown to extend to host antimicrobials as well (76). Resistance to the antimicrobial peptide, calprotectin, was recently found to be related to *H. pylori* lipid A modification and the associated biofilm formation (76). Biofilms are also thought to provide protection against environmental changes. *H. pylori* resides within the gastric niche, yet it is not an acidophile. One hypothesis regarding biofilm formation by *H. pylori* may be that it provides a more stable environment, adding an additional level of protection from the fluctuating pH of the stomach (189).

***Extra-gastric Biofilms***

The only definitive reservoir for *H. pylori* is the stomach of humans and non-human primates; however, several studies have focused on the role that biofilms may play in extra-gastric *H. pylori* colonization. The presence of *H. pylori* has been described in dental plaques (reviewed in(10)), various aquatic biofilms (reviewed in(80)), and biofilms on vegetables (16; 147; 213). The fastidious nature of *H. pylori*, the specificity of the gastric niche, and the tendency of *H. pylori* to transition to the coccoid form when stressed, all seem contrary to the survivability of extra-gastric *H. pylori*, but perhaps biofilm formation is the key to forming a favorable habitat. The implication that water serves as a potential source of *H. pylori* is not novel. Indeed, a majority of the early *H. pylori*-biofilm studies focused on the possibility that *H. pylori* forms a biofilm as a mechanism to survive in aquatic reservoirs (80). Investigators analyzed naturally

occurring biofilms in water systems (40; 159) in addition to constructing artificial models in laboratories to investigate this hypothesis (18). Unfortunately, initial characterizations of water-associated *H. pylori* biofilms relied on methods such as PCR (126; 159) and peptide DNA probes (20) to demonstrate *H. pylori* presence. These methods fail to indicate the presence of viable, infective *H. pylori*. To address this issue, follow-up studies implemented *in situ* fluorescence hybridization (FISH), which can be utilized to assess viability based on cellular rRNA content, to detect *H. pylori* (81; 146). Despite all of these methodologies, the gold standard is still culture, and two groups have obtained positive *H. pylori* cultures from wastewater (124; 146). While these groups did not comment on biofilm formation, *in vitro* evidence suggests that interactions with specific bacteria (ex: *Mycobacterium chelonae*) within a multi-species biofilm may help contribute to *H. pylori* viability (82). Taken together, these data suggest that biofilms may contribute to the survival of *H. pylori* in natural and municipal water sources, and thus contribute to *H. pylori* transmission from an environmental reservoir.

In addition to water systems, dental plaques and vegetables have also been investigated as locations for extra-gastric *H. pylori* biofilms. *H. pylori* has been cultured from numerous vegetable sources (16; 213), and the formation of biofilms on certain vegetables was demonstrated in a laboratory setting (147). These data suggest food-borne transmission cannot be entirely discounted. Finally, while the gastric mucosa is the primary residence for *H. pylori*, several groups have proposed the oral cavity may act as a potential reservoir for reseeding of self infections (8). The prevalence of *H. pylori* in dental plaques varies widely between studies, and methods utilized to detect the presence of *H. pylori* in dental plaques range from PCR to culture (8). One intriguing piece of

evidence found for individuals who tested positive for both oral and gastric *H. pylori* was that concomitant peridontal and triple therapy resulted in a better treatment success rate (77.3%) than triple therapy alone (47.6%) (221). While the evidence of extra-gastric *H. pylori* is not definitive, the potential involvement of biofilms as a means to foster a more favorable colonization niche is worth further exploration.

***Molecular Mechanisms of Biofilm Development***

The first characterization of *H. pylori* biofilms was published in 1999 (189). While previous studies had demonstrated the ability of *H. pylori* to develop biofilms at the air-liquid interface, Stark and colleagues focused on identification of components of the EPS matrix. Polysaccharides similar to those in the LPS of *H. pylori* were identified as major component of the EPS (189). They also tested the solubility of this crude biofilm material in HCl and found that it was insoluble. This led to the conclusion that the biofilm may help to protect *H. pylori* from low pH in the gastric environment (189). A more descriptive analysis of *H. pylori* biofilm formation found a classical course of biofilm development: initiation following the adherence of individual bacteria, expansion into bacterial colonies/aggregates, and formation of 3D structures that produce EPS (45). While the exact composition of the *H. pylori* biofilm matrix has yet to be elucidated, several studies have identified contributing components (45; 76; 92; 216; 218; 220). In one study, mannose was identified as the major sugar component of both four and seven day old biofilms, and subsequent analysis showed that proteomannans were a major biofilm component (216). In addition, it was recently demonstrated that modifications to

lipid A contribute to biofilm formation (76). Together these data suggest that components of LPS play a prominent role in biofilm formation by *H. pylori*.

The presence of outer membrane vesicles (OMVs) as a prominent feature of the *H. pylori* biofilm matrix has been shown by two separate groups, using two different *H. pylori* background strains (92; 218). Interestingly, these OMVs were shown to have distinct protein profiles that varied in mature biofilms (day 2 and 3) as compared to early biofilms (day 1) (218; 219). Another study found that the amount of extracellular DNA (eDNA) associated with biofilm OMVs was 4 fold higher than the eDNA associated with planktonic OMVs (92). Additionally, the composition of the OMV-associated eDNA was found to be different than corresponding intracellular DNA. eDNA is known to act as a structural component in biofilms formed by other species (92). Initial studies suggested that eDNA was not a structural component of the biofilm matrix based on DNAse I analysis; however, those results may be misleading as the association between eDNA and OMVs may have restricted the activity of DNAse I (90; 92). In summary, proteomannans, lipid A structure, OMVs, and eDNA likely all contribute to the structure of the *H. pylori* biofilm matrix.

Several genes known to contribute to biofilm formation by other Gram-negative pathogens were evaluated in an attempt to determine the genetic basis of biofilm formation by *H. pylori* (45). While several known biofilm factors had no effect, deletion of *luxS* was found to increase biofilm formation (45). In addition, a decrease in *luxS* expression was observed in biofilm-associated *H. pylori* following antibiotic treatment (27). LuxS is important for quorum sensing and appears to act as both an inducer of biofilm formation and as a dispersal agent (12). In fact, a study of the quorum signal

auto-inducer 2 (AI-2) verified that AI-2 acts as a signal for dispersal of *H. pylori* biofilms (12). Other genes reported to influence *H. pylori* biofilm formation include *cagE* (45) and *napA* (216). While NapA has many roles, one of its predicted functions is as an adhesin; therefore, it is perhaps not surprising that deletion of NapA leads to decreased adherence and a subsequent reduction in biofilm formation (216). The deletion of *cagE*, which encodes an ATPase that energizes the transport of *cagA* and other substrates through the type IV secretion system, is known to abolish translocation of *cagA* (45; 191). It is not yet clear why deletion of this gene would effect biofilm formation, but it may suggest a role for type IV secretion in this process (45).

In a more global approach, two groups compared the proteomes of planktonic bacteria and biofilm-associated bacteria. Those studies were conducted using different time points, conditions, and techniques (1D vs 2D gels,) which makes it difficult to compare them directly; however, there were a few overlapping proteins identified (183; 216). Furthermore, while many identified proteins were not identical, certain common functional groups emerged, including stress response proteins. Biofilms are often believed to be formed in response to environmental stress, though not all proteins upregulated in response to stress will play a role in biofilm formation. For example, GroEL, a protein chaperone, was upregulated in the biofilm associated *H. pylori*, but deletion of this gene did not affect biofilm formation (183).

In addition to identification of genes associated with *H. pylori* biofilm formation, much attention has been dedicated to elucidation of environmental factors that influence biofilm formation (27; 45; 183; 189; 211). Despite the attention to environmental stimuli, few studies have attempted to parse out how *H. pylori* interprets these singles and

switches to a biofilm life style. As mentioned previously, LuxS was shown to be down-regulated in biofilm associated *H. pylori* (27). Therefore, LuxS represents one regulatory system that may contribute to *H. pylori* biofilm formation. Two-component systems have also been shown to regulate biofilm formation in several bacteria and may dictate biofilm development in *H. pylori* ((195), reviewed in (25; 88))However, deletion of the response regulator HP1365 was shown to have no effect on biofilm formation by *H. pylori;* however, a decrease in ArsR was reported in biofilm-associated *H. pylori* as compared to planktonic cells (45; 183). These data suggest that the acid-responsive two component system may play a role in integration of environmental signals to as a means to direct biofilm formation by *H. pylori*.

## GOALS AND SPECIFIC AIMS

*H. pylori* is a truly unique pathogen. It colonizes a niche that was long believed to be sterile and does so with incredible efficiency. Not only does *H. pylori* colonize approximately 50% of the world's population, but within an individual host, *H. pylori* has the ability to persist for the lifetime of the host. The success of *H. pylori* is intricately linked to its ability to quickly adapt to environmental and host changes. The goal of this dissertational work was to gain a better understanding of the genetic mechanisms that allow *H. pylori* to adapt to and thrive within this incredibly inhospitable niche. To begin to tease out these mechanisms, we took a two-pronged approach and assessed (1) the interactions between the key *H. pylori* regulatory proteins Fur, NikR, and ArsS and (2) the expression pattern of *homB*, an OMP that is associated with symptomatic infection.

To address aim one we constructed a series of isogenic mutant strains, in which three key regulatory proteins (Fur, NikR and ArsS) were deleted individually or in combination. Several prior studies have shown the overlapping regulons of these three regulators; however, few have investigated the effect of their combined loss (32; 37; 61). Of note, while no major defects in growth were observed between the double regulatory knockouts, the *ΔfurΔnikRΔarsS* triple knockout did appear to be less tolerant to long term culture (>20 hours) and dipslayed an increase in the number of coccoid cells at later time points. This finding suggests that loss of all three regulators led to an aberrant stress response. Additionally, through the course of this work, biofilm formation at the air-liquid interface was observed for strains lacking ArsS. A similar phenotype was observed whether ArsS was deleted individually or in combination with other regulatory proteins, and was also replicated with an ArsR-D52N mutant strain. Further investigation of this phenotype revealed that deletion of ArsS resulted in an increased rate of biofilm formation, as well as an increase in the amount of biofilm formed. It is likely that the enhanced biofilm phenotype is related to an increase in adherence and aggregation. These findings led us to hypothesize that the ArsRS two-component system acts as an environmental indicator of when *H. pylori* has reached the epithelial layer. Under favorable conditions, ArsR will modulate the expression of factors important for adherence and biofilm formation, which may aid in colonization of the gastric mucosa. Previous studies have demonstrated biofilm formation during natural *H. pylori* infection (33; 39; 41; 52) as well as in an animal infection model (17). Our work supports the role of ArsRS in regulation of biofilm formation and identifies the first regulatory protein involved in biofilm formation by *H. pylori*.

In parallel to the work described above, we began a molecular characterization of HomB, an OMP that was found to be associated with more severe disease. Through this analysis we discovered that *homB* expression is repressed at low pH, and that this suppression occurs in an ArsRS-dependent manner. In light of the enhanced biofilm formation observed in response to ArsRS mutation, we hypothesized that *homB* contributed to the enhanced biofilm phenotype of the mutant strains. Intriguingly, we showed that deletion of *homB* from both the ΔarsS and ArsR-D52N backgrounds led to a reduction in the biofilm formation by these strains.

In the following chapters, we present the detailed characterization of the *H. pylori* regulatory mutant strains, and detail the discovery that ArsRS regulates biofilm formation in *H. pylori* (chapter 2). Furthermore, we show that the OMP encoded by *homB*, is one of the genes that responds to ArsRS and contributes to biofilm formation (chapter 3). Although biofilm formation by *H. pylori* has been demonstrated *in vivo*, this is the first time that regulation of biofilm formation has been described, or that regulation of an OMP has been shown to contribute to biofilm development in this bacterium. Together, our data suggest a model in which pH is interpreted by ArsRS as a cue to coordinate the expression of pro-biofilm factors. Given that treatment failure of *H. pylori* infections is rising, studies such as the ones presented in this dissertation may help to provide a better understanding of the underlying mechanisms of gastric colonization by *H. pylori* and may subsequently aid in the design of more efficacious treatment regimens.

# CHAPTER TWO: Characterization of Key *H. pylori* Regulators Identifies a Role for ArsRS in Biofilm Formation

Published as: **Stephanie L. Servetas, Beth M. Carpenter, Kathryn P. Haley, Jeremy J. Gilbreath, Jennifer A. Gaddy, D. Scott Merrell. 2016.** Characterization of Key *Helicobacter pylori* Regulators Identifies a Role for ArsRS in Biofilm Formation. *J Bacteriol* **198**:2536-2548.

The work presented in this chapter is the sole work of S.L. Servetas with the following exceptions: B.M. Carpenters assisted in growth curve studies; J.J. Gilbreath constructed DSM 986 and DSM 987; K.P. Haley and J.A. Gaddy conducted the SEM studies.

## ABSTRACT

*Helicobacter pylori* must be able to rapidly respond to fluctuating conditions within the stomach. Despite this need for constant adaptation, *H. pylori* encodes few regulatory proteins. Of the identified regulators, Fur (ferric uptake regulator), NikR (nickel response regulator), and ArsRS (two-component acid response system), are each paramount to the success of this pathogen. While numerous studies have individually examined these regulatory proteins, little is known about their combined effect. Therefore, we constructed a series of isogenic mutant strains that contained all possible single, double and triple regulatory mutations in Fur, NikR and ArsS. Growth curve analysis revealed minor variation in growth kinetics across the strains; these were most pronounced in the triple mutant and in strains lacking ArsS. Visual analysis showed that

strains lacking ArsS formed large aggregates and a biofilm-like matrix at the air liquid interface. Biofilm quantification using crystal violet assays and visualization via scanning electron microscopy (SEM) showed that all strains lacking ArsS or containing a non-phosphorylateable form of ArsR (ArsR-D52N) formed significantly more biofilm than the wild-type strain. Molecular characterization of biofilm formation showed that strains containing mutations in the ArsRS pathway displayed increased levels of cell aggregation and adherence, both of which are key to biofilm development. Furthermore, SEM analysis revealed prevalent coccoid cells and extracellular matrix formation in the ArsR-D52N, *ΔnikRΔarsS*, and *ΔfurΔnikRΔarsS* strains, suggesting that these strains may have an exacerbated stress response that further contributes to biofilm formation. Thus, *H. pylori* ArsRS has a previously unrecognized role in biofilm formation.

**Importance**

Despite a paucity of regulatory proteins, adaptation is key to the survival of *H. pylori* within the stomach. While prior studies have focused on individual regulatory proteins such as Fur, NikR and ArsRS, few studies have examined the combined effect of these factors. Analysis of isogenic mutant strains that contained all possible single, double and triple regulatory mutations in Fur, NikR and ArsS revealed a previously unrecognized role for the acid responsive two-component system, ArsRS, in biofilm formation.

## INTRODUCTION

*Helicobacter pylori* is a Gram-negative pathogen that colonizes the gastric mucosa of humans and non-human primates. Approximately 50% of the world's

population is infected with *H. pylori*, and phylogeographic evidence suggests that the bacterium may have been associated with mankind for over 100,000 years (145). Given this long-standing association with its host and the specificity of the gastric niche, it is not surprising that *H. pylori* exhibits a minimalist genome indicative of reductive evolution (6; 203). As such, the *H. pylori* genome encodes approximately half the number of regulatory proteins seen in *Haemophilus influenzae*, which has a similar sized genome (175; 203); moreover, the genome of the closely related species *Campylobacter jejuni,* also appears to encode for a larger repertoire of regulatory proteins (160). Within the human stomach, *H. pylori* faces fluctuating stressors such as acidity, nutrient availability, osmotic pressure, and host defenses. In order to adapt and persist within this hostile niche, *H. pylori* has developed a complex regulatory network that optimizes the activities of the few regulators carried in the genome (56; 58; 175). Three key nodes within this regulatory structure are represented by the following regulatory proteins: Fur, the ferric uptake regulator; NikR, a nickel response regulator, and ArsRS an acid responsive two-component system (58). The prominent roles of these regulators are highlighted by the fact that all three are required for wild-type colonization levels in animals models (32; 78; 141; 158). The importance of the OmpR-like response regulator ArsR is further demonstrated by the fact that it is essential for *in vitro* survival. Intriguingly, the cognate histidine kinase, ArsS, is dispensable *in vitro* and a non-phosphorylateable form of ArsR (ArsR-D52N) also supports viability of *H. pylori* (24; 134; 176). Thus, essentiality of ArsR must depend on regulation of essential genes by the non-phosphorylated form of the protein (130). The nature of this regulation is not clearly understood, and despite the two states of ArsR, most transcriptional studies focus on elucidating the targets of the

pH-dependent, phosphorylated ArsR (ArsR~P) regulon. Based upon this literature, current evidence indicates that ArsR~P can act to repress or activate expression of different sets of genes (121; 163).

In an effort to understand adaptation, numerous studies have sought to define the regulon of genes controlled by Fur, NikR and ArsRS (43; 46; 62; 67; 78; 121; 163; 208). Examination of these transcriptome studies *en masse* demonstrates the vast and overlapping nature of the Fur, NikR, and ArsRS regulons. Furthermore, evidence suggests that these regulators affect expression of each other; *nikR* and *arsR* are part of the Fur regulon (57; 114) whereas NikR appears to regulate *fur* (46; 61). Given the overlapping nature of these regulatory cascades, it's likely not surprising that the role of each of these factors appears to be expanded as compared to what is seen in other bacteria. For example, although Fur and NikR are classically thought of as metallo-regulators, in *H. pylori* they also play an important role in acid acclimation (28; 78). Furthermore Fur has been shown to be important for adaptation to oxidative, nitrosative and osmotic stress (35; 79; 167). Thus, further study of the intricate interaction between each of these regulatory factors may provide increased information about the adaptation process of *H. pylori*.

One general adaptive mechanism used by many bacteria is the formation of biofilm structures that provide increased protection from a number of different stresses. Biofilm formation by *H. pylori* has been demonstrated both *in vitro* (45; 189; 219) and *in vivo* (17; 33; 39; 41; 52). While it is unclear exactly what drives biofilm formation in *H. pylori,* protection from antibiotics and host defenses during infection (33; 76; 217) and/or protection in putative environmental reservoir (19; 81; 116; 161) have been proposed.

Basic characterization has shown that *in vitro* biofilm formation is variable across strains (219) and appears to be induced by conditions such as serum/nutrient starvation (27; 211). Initial investigations of the extracellular polymeric substances (EPS) matrix observed in *in vitro* biofilms revealed the presence of lipids, amino acids, and monosaccharides. Further investigation of the EPS matrix demonstrated the presence of extracellular DNA (eDNA); however, in contrast to what has been shown for other Gram-negative species, eDNA does not appear to play a structural role in *H. pylori* biofilms (91). An increased number of outer membrane vesicles (OMVs) are associated with *H. pylori* found within a biofilm as compared to planktonic bacteria, suggesting a potential link between OMV production and biofilm formation (92; 218). Interestingly, OMVs have been observed in association with other Gram-negative biofilms such as those formed by *Pseudomonas aeruginosa* (177).

Apart from what can be inferred from the descriptive biofilm studies, little is known about the molecular mechanisms that drive biofilm formation or the genes that are required for biofilms in *H. pylori*. Notable exceptions include studies that indicate that deletion of *cagE* and *luxS* result in increased biofilm formation, whereas deletion of additional genes (*crdR, clpA, ppk, hpA*) known to contribute to biofilm development in other bacteria had no effect on biofilm formation by *H. pylori* (45). Proteomic analysis designed to identify proteins involved in biofilm formation identified numerous proteins that were differentially regulated between biofilm and planktonic *H. pylori*; intriguingly, ArsR was among the proteins upregulated (>2 fold) in the biofilm community (183). Overall, relatively little is understood about the molecular mechanisms that lead to biofilm formation in this important pathogen.

In an effort to understand the complex regulatory interaction between Fur, NikR, and ArsRS, herein we describe the creation and basic characterization of a series of isogenic *H. pylori* mutant strains in which each regulatory protein was knocked out singularly or in combination. Basic characterization of these strains led to the novel observation that the ArsRS system plays a previously unappreciated role in biofilm formation in this important pathogen.

## MATERIALS AND METHODS

### Bacterial Strains and Growth

Strains and plasmids used in the study are listed in Table 1. All isogenic mutant strains were created using *H. pylori* G27 as the parental strain (21; 53). Unless otherwise noted, *H. pylori* strains were cultured as previously described (37). Briefly, all cultures were grown at 37°C, in gas evacuation jars, under microaerobic conditions (5% $O_2$, 10% $CO_2$ , and 85% $N_2$ ) generated with an Anoxomat gas evacuation and replacement system (Advanced Instruments, Inc). For plate cultures, strains were grown on horse blood agar (HBA) comprised of 4% Columbia agar (Neogean Corporation), 5% defibrinated horse blood (HemoStat Laboratoris, Dixon, CA), 2 mg/mL β-cyclodextran (Sigma), and an antibiotic/antifungal cocktail [10 μg/ml vancomycin (Amresco), 5 μg/ml cefsulodin (Sigma), 2.5 U/ml polymyxin B (Sigma), 5 μg/ml trimethoprim (Sigma), and 8 μg/ml amphotericin B (Amresco)]. *H. pylori* liquid media consisted of brucella broth (Neogen Corporation) supplemented with 10% fetal bovine serum (FBS) (Gibco) and 10ug/mL vancomycin. Liquid cultures were grown shaking at 100rpm. *H. pylori* strains

**Table 1. Primers, plasmids, and strains used in this study.**

| Primers | Sequence | Source |
|---|---|---|
| ArsR-D52N construction | | |
| Up1F | GGGGATTTTTTGAGCGTTGAG | this study |
| Up2R | gttagtcacccgggtaccgagctcCACTCTGTGTTTTCTCGCTCC[a] | this study |
| Dn1F | ctagagtcgacctgcaggcatgcaagCCATGAAAACAAAGCCTTAATTTCC[a] | this study |
| Dn2R | CTAAATTAGGCAAAGTCAAATtCAACAACAACAAATC[b] | this study |
| Dn3F | GATTTGTTGTTGTTGaATTTGACTTTGCCTAATTTAG[b] | this study |
| Dn4R | CGCAAACGGCCAATGATCAC | this study |
| KanF | GTTAGTCACCCGGGTACCGAGCTC | this study |
| KanR | CTAGAGTCGACCTGCAGGCATGCAAG | this study |
| *Δfur* Confirmation | | |
| Fur-expressF1 | ATGAAAAGATTAGAAACTTTGGAA | This study |
| Fur-expressR1 | ACATTCACTCTCTTGGCATTC | This study |
| *ΔnikR* Confirmation | | |
| HP1338_Del_Conf_F | GGCCAACCATTTAAATCCAG | this study |
| HP1338_Del_Conf_R | GTTTTGATAAGCGGGCAAGA | this study |
| *ΔarsS* Confirmation | | |
| HP0165_del_ver_F | TGAAAGCATTGCGATTGAGA | this study |
| HP0165_del_ver_R | AAAACGGCTTTGATGCCTAA | this study |
| | *ΔarsS* Complementation | |
| ArsS_F | GGGGTTTTGCGTTTCTCTATC | this study |
| ArsS_compR | ***AATTCCGCGG***TTATTGTTTTTCTTTAACCCCAC[c] | this study |
| ArsS_compF | ***AATTTCTAGA***TTAAGATATTTTCAAGGTATCTTTAATAATAAAA TGCTAGAATTAGTCCCTTAAAACATTAGCAAAGTTAATCGTT TTTAAAGTTGCAGAAATCATTGAAGGAGTTAATTGAAGTttgcgttt ctctatctttttttaagg[a, c] | this study |
| ArsRProm_F | ***AATTTCTAGA***TTAAGATATTTTCAAGGTATC[c] | this study |
| **Plasmids** | **Description** | **Source** |
| pDSM924 | pGEM T-easy::Δ*nikR*::*cat* | (37) |
| pDSM920 | pBluescript::*ΔarsS*::kan | (209) |
| pΔHP1027-K7 | pCRII-Topo::*Δfur*::kan | (78) |
| pDSM1525 | PTM117:*arsS*$^{C}$ | this study |
| **Strains** | **Description** | **Source** |
| DSM 1 | WT G27 | (21) |
| DSM 41 | G27 *ΔflgR::cat,* Cm$^{R}$ | (134) |
| DSM 145 | G27Δ*fur::kan*, Kan$^{R}$ | (36) |
| DSM 300 | G27Δ*fur::cat*, Cm$^{R}$ | (36) |
| DSM 975 | G27Δ*nikR::cat*, Cm$^{R}$ | (37) |
| DSM 983 | G27*ΔarsS* markerless | (37) |
| DSM 986 | G27*Δfur::kan/ΔnikR::cat,* Kan$^{R}$ Cm$^{R}$ | this study |
| DSM 987 | G27*Δfur::cat/ΔarsS::kan,* Kan$^{R}$ Cm$^{R}$ | this study |
| DSM 1071 | G27*ΔarsS/nikR::cat*, Cm$^{R}$ | (37) |
| DSM 1441 | G27*ΔarsS*/*Δfur::kan/ΔnikR::cat,* Kan$^{R}$ Cm$^{R}$ | this study |
| DSM 1446 | G27 ArsR-D52N, unphosphorylatable, KanR | this study |
| DSM 1526 | G27*ΔarsS* markerless PTM117:*arsS*$^{C}$ | this study |

[a] lower case letter indicate regions of overlap sequence for SOE PCR

[b] underlined lower case nucleotide indicates the point mutation that encodes for the D52N.

were maintained as stock cultures at -80°C. Stock cultures were frozen in brain heart infusion broth (BD Biosciences) containing 10% FBS and 20% glycerol (EMD chemicals, Inc).

**Construction of G27 *ΔfurΔnikR***

The G27 *ΔfurΔnikR* mutant strain was constructed by naturally transforming DSM145 (*Δfur*::kan) (36) with pDSM924 (37), which carries a *ΔnikR::cat* cassette. Transformants were selected on HBA plates supplemented with 25µg/mL kanamycin (Kan) (Gibco) and 8µg/mL chloramphenicol (Cm) (EMD Chemicals, Inc.) to ensure colonies selected were lacking both *fur* (Kan) and *nikR* (Cm). Integration and subsequent deletion of *nikR* was confirmed by PCR using primers HP1338_Del_Conf_F and HP1338_Del_Conf_R. Sequences of primers are listed in Table 1. Colonies with amplicons of the expected size for *nikR*::cat were sequenced to verify deletion of *nikR*. Fur deletion confirmation primers Fur-expressF1 and Fur-expressR1 are modified versions of 1027-F1 (SalI) and 1027-R6 (NotI), respectively (78), in which the restriction site has been removed. Fur-expressF1 and Fur-expressR1 were used in PCR and sequencing to verify that the *Δfur* mutation was unaltered and the G27*ΔfurΔnikR* strain was archived as DSM986.

**Construction of G27 *ΔfurΔarsS***

The G27 *ΔfurΔarsS* mutant strain was constructed by naturally transforming pDSM920 (209), carrying a *ΔarsS*::*kan* cassette, into DSM300 (*Δfur*::cm) (36). Transformants were selected on HBA plates supplemented with 25µg/mL Kan and

8μg/mL Cm. Kan/Cm resistant colonies were verified by PCR using primer pairs HP0165_del_ver_F/ HP0165_del_ver_R and Fur-expressF1/ Fur-expressR1 for *ΔarsR* and *Δfur*, respectively. Products displaying the expected size change for deletion of *arsS* and *fur* were sequenced using the primers above and the *ΔfurΔarsS* mutant strain was archived as DSM987.

**Construction of G27*ΔfurΔnikRΔarsS***

The G27 mutant strain containing the combined deletions of *fur*, *nikR* and *arsS* was constructed by transforming DSM1071 (*ΔnikRΔarsS*) (37) with pΔHP1027-K7 (78), which contains a *Δfur::kan* cassette. Transformants were selected on 25μg/mL Kan. PCR with the following primer sets: Fur-expressF1/Fur-expressR1, HP1338_Del_Conf_F/HP1338_Del_Conf_R, and HP0165_del_ver_F/HP0165_del_ver_R was used to screen Kan resistant colonies for deletion of *fur*, *nikR* and *arsS*, respectively. The Fur deletion mutant was further confirmed by sequencing with Fur-expressF1/Fur-expressR1 and the resulting *ΔfurΔnikRΔarsS* strain was archived as DSM1441.

**Construction of G27 ArsR-D52N**

Site directed mutagenesis was used to create a non-phosphorylatable ArsR mutant strain (ArsR-D52N) (130; 176). The ArsR-D52N mutant strain was constructed in a step-wise manner using consecutive splicing by overlap extension (SOE) PCR followed by allelic exchange. The initial SOE PCR was performed to introduce an adenine at position 154 of *arsR*, which results in an asparagine at amino acid position 52 in lieu of an

aspartic acid. To accomplish this, *arsR* was amplified in two fragments: fragment one was amplified with Dn1F and Dn2R and included the *arsR* promoter and about 1/4 of the coding region (nucleotides 169811-170269, gpCP001173.1), fragment two was amplified with Dn3F and Dn4R and included a 30bp overlap with fragment one in addition to a majority of *arsR* 3' region (nucleotides 170234-170584). Dn2R and Dn3F both contain the G→A transition, which is introduced into *arsR* during SOE PCR of the two *arsR* fragments. The resulting *arsR* construct, which carries the single base pair modification, was then used as the downstream section for the subsequent SOE PCRs. In addition to the *arsR* construct, two standard PCRs were run to amplify the region upstream of *arsR* (nucleotides 170645-171164, amplified with Up1F and Up2R) and a $Kan^R$ cassette (amplified with KanF and KanR). Next, the upstream fragment was fused to the 5' end of the $Kan^R$ cassette through SOE PCR with Up1F and KanR. In the final step, the upstream-$Kan^R$ PCR product was fused the 5' end of the *arsR* construct by SOE PCR using Up1F and Dn4R. The resulting SOE PCR product was used to transform G27 (DSM1) (21). Transformants were selected on 25μg/mL Kan. Dn1F and Dn4R primers were used to PCR amplify Kan resistant colonies and the resulting amplicons were screened by restriction digest with Apo1; G→A transition introduces an Apo1 site. Dn1F and Dn4R were used to sequence the ArsR region of colonies for which PCR products yielded the correct Apo1 digest pattern. 2X sequence coverage for the region was obtained to confirm the transition mutation and the resulting strain was archived as DSM1446.

**Construction of the Complemented G27*ΔarsS-arsS$^C$* Strain**

The *arsS* deletion was complemented in trans. Since *arsS* is the second gene in an operon with *arsR*, the promoter region of *arsR* was fused upstream of *arsS* using SOE PCR. Briefly, the *arsR* promoter region was synthesized as an oligonucleotide named ArsS_compF; this oligonucleotide was designed to contain an overlapping complementary sequence to *arsS*. The *arsS* coding region was amplified with primers ArsS_F and ArsS_compR. The promoter fragment and the *arsS* fragment were then mixed together and fused in a SOE PCR that employed the ArsR_promF and ArsS_compR primers; these primers were designed to incorporate restriction sites needed for downstream cloning. The resulting fragment was digested with XbaI and SacII and ligated into the similarly digested PTM117 vector (36). The ligation was then transformed into Top10 chemically competent *E. coli,* and transformants were selected on LB containing 25µg/mL Kan. PTM117:*arsS* positive colonies were confirmed by PCR, and the resulting PTM117:*arsS* complementation construct was transformed into the G27*ΔarsS* strain. Transformants were selected on HBA containing 25µg/mL Kan. To prove reproducibility, five separate complementation colonies (designated *ΔarsS-arsS*$^{C}$ ) were selected, were shown to express arsS by quantitative real-time (qRT) –PCR, and were tested in the biofilm assay. Given that all 5 isolates behaved similarly, one isolate was archived as DSM1526.

**Growth Curves**

Bacterial cultures were patched from freezer stocks onto HBA plates. After 48 hours, sterile swabs were used to inoculate a 10mL liquid starter culture in *H. pylori* liquid media. Liquid starter cultures were grown for 20-24 hours and then 60mL liquid

cultures were standardized to a starting OD 600 of 0.05 ODU from the overnight starter culture. Each 60mL culture was then divided in half into two separate 125mL flasks and grown under standard lab conditions; CFU counts and OD readings were taken from alternating culture flasks at 0, 6, 12, 20, 30, and 48 hours post inoculation. Data shown represents three biologically independent experiments.

**Biofilm Quantification**

Bacterial cultures were started following the same procedure used in the growth curves; however, biofilm assays were conducted in 24 well tissue culture treated (hydrophilic, ionic coating) plates (Corning Inc.). Starter cultures were used to inoculate 1mL of liquid media per well to an OD 600 of 0.1 ODU. A separate plate was used for each of the five time points: T6, T12, T24, T48, and T72. At each time point the media was aspirated and each well was washed 2X with PBS. Plates were dried for 5 minutes at 37°C and biofilms were subsequently fixed with methanol (J.T. Baker). 1% Gram's crystal violet solution (Sigma-Aldrich) was then added to each well until full (approximately 2 mLs), including one empty control well to serve as a blank. The plates were incubated on the bench top for 15 minutes, after which crystal violet was removed from the wells and each well was washed 3X with distilled water. Plates were again dried for 5 minutes at 37°C. At this point, 1.5 mL of differentiation solution (Sigma) were added to solubilize the dye contained within the biofilms. Plates were incubated with differentiation solution for 15 minutes, 1mL of solubilized crystal violet solution was then transferred to a cuvette, and the absorbance (OD 590) for each sample was read. Data shown represents three biologically independent experiments.

**SEM**

Stock cultures (0.3-0.7 OD 600) were inoculated 1:10 into fresh Brucella broth in 12 or 24 well plates. Biofilm were captured on cell-culture treated cover slips at the bottom of the wells and analyzed by scanning electron microscopy as previously described (77). Briefly, samples were fixed with 2.0% paraformaldehyde (Electron Microscopy Sciences), 2.5% gluteraldehyde (Electron Microscopy Sciences) in 0.05M sodium cacodylate, pH 7.4 (Electron Microscopy Sciences) buffer for 24 hours. After primary fixation, samples were washed three times with 0.05M sodium cacodylate buffer before sequential dehydration with increasing concentrations of ethanol. After ethanol dehydration, samples were dried at the critical point using a Tousimis Critical Point Dryer machine, mounted onto aluminum SEM sample stubs (Electron Microscopy Sciences), and sputter-coated with 5 nm of gold-palladium. Afterward, samples were painted with a thin strip of colloidal silver (Electron Microscopy Sciences) at the edge to facilitate charge dissipation. Biofilms were imaged with an FEI Quanta 250 field-emission gun scanning electron microscope. Micrographs shown are representative of three biological replicates.

**Motility Assay**

*H. pylori* cultures were started as described in the growth curve assay. After approximately 20 hours of growth, each culture was passed through a 30 gauge needle to disrupt large aggregates that form in some strains, which could influence OD measurement. After passage through the needle, 0.5 ODU of each culture was pelleted

and resuspended in 50µl of *H. pylori* liquid media. From this suspension, 2µL was inoculated into motility agar plates with a 10µL pipette tip. Motility agar consisted of 2.5% brucella broth, 0.3% agar-agar (EMD chemicals), 10% FBS, and 10µg/mL vancomycin. Plates were incubated at 37°C under microaerobic conditions and the area of the halo of motility was measured 3, 5 and 7 days post-inoculation. At each time point plates were removed and imaged on an Imagequant Las4000 (GE). Area measurements were completed in ImageJ 1.48v (http://imagej.nih.gov/ij/) using a scale set by the known dimensions of the plate. Data presented represent three independent biological replicates.

**Bacterial Aggregation Assay**

Bacterial auto-aggregation was assessed as previously described (219) with the following modifications. *H. pylori* cultures were prepared as described in the bacterial motility assay. 1.0 ODU of each culture was pelleted, washed with PBS, re-suspended in 2mL of PBS in 12 X 75mm glass tubes, and an initial OD at 600 nm was taken ($T_0$). The tubes were then placed in a shaking incubator (190 rpm, 37°C) and additional OD readings at 600 nm were taken after 2 hours. Percent auto-aggregation was calculated as: [(OD600 $T_0$- OD600 $T_2$)/ OD 600 $T_0$)] x 100. Data shown represents a minimum of three biological replicates.

**Plate Adherence Assay**

The plate adherence assay was conducted as previously described (185) with several modifications. Overnight, liquid cultures of each strain were inoculated into 6 separate wells of a 96 well tissue culture treated plate (Corning) at an OD 600 of 0.1

ODU per well in 100μL of *H. pylori* liquid media. At various time points post inoculation, the liquid was aspirated from three wells per strain. These same wells were washed three times with PBS after which 200μL of 5% alamarBlue (Invitrogen) diluted in *H. pylori* liquid media were added to each well. The 5% alamarBlue solution was also added to three empty wells to serve as the blank. In the three remaining wells for each strain, 100μL of 10% alamarBlue in *H. pylori* liquid media was added. These wells were used to determine the number of starting bacteria. After addition of alamarBlue, plates were wrapped in foil to prevent light exposure and were incubated at 37°C, under microaerobic conditions, in a static incubator for 2 hours. The plate was then read using an excitation wavelength of 544nm and emission wavelength of 590nm, which resulted in Relative Fluorescence Units (RFU) for each well. Triplicate wells for each experiment were averaged. From the averaged values a percent adherence was calculated as: (RFU of the washed wells/RFU inoculum) x100. Data shown represents a minimum of three biologically independent experiments.

**Statistical Analysis**

Statistical analyses were conducted using GraphPad Prism v6.0g (GraphPad Software, Inc). A Two-way ANOVA with Tukey's correction for multiple comparisons was used to analyze increases in biofilm formation over time for individual strains and also for differences in biofilm formation between strains at each time point. For aggregation and adherence assays, wild-type G27 was compared to either *ΔarsS* or ArsR-D52N using a one-way ANOVA with Dunnett's multiple comparison test.

# RESULTS

## Growth Characteristics of Fur, NikR, and ArsRS Mutant Strains

Growth curve analysis under standard laboratory conditions was conducted for each of the single, double and triple regulatory mutants as a means to identify any substantial changes in strain health. While, some minor differences were evident (Figure 4), no overt growth defects were observed for the single or double regulatory knockout strains. However, DSM1441, the strain which lacks Fur, NikR and ArsR, appeared to grow more slowly (Figure 4) and took longer to recover from the frozen stock (data not shown). Moreover, visually this strain appeared more stressed and transitioned to the coccoid form more quickly compared to the other strains (data not shown). For the remaining strains, while growth was fairly similar throughout exponential phase (through T24), strains lacking ArsS appeared to plateau at a lower CFU and ODU compared to the other strains. The difference in ODU and CFU for strains lacking ArsS was most pronounced between 24 and 30 hours of growth as cultures entered stationary phase. Around the same time point, noticeable clumping was observed in these cultures as well as a biofilm like ring at the air liquid interface (Figure 5). Given this phenotype, it is possible that the slight differences in growth observed for *ΔarsS*, *ΔfurΔarsS*, *ΔnikRΔarsS*, and *ΔfurΔnikRΔarsS* is partially the result of bacterial clumping that skewed both optical density reading and CFU enumeration. To determine the nature of the bacterial clumping, liquid bacterial cultures were viewed using bright field microscopy; aggregates composed primarily of spiral *H. pylori* with some extracellular material were observed (Figure 6).

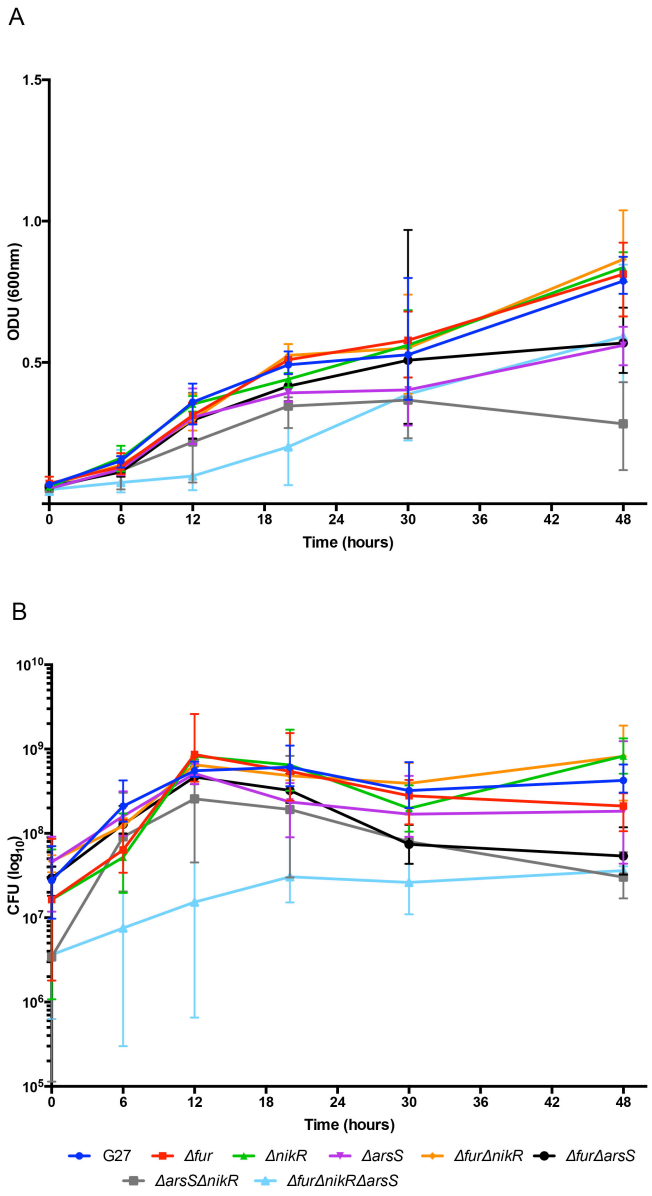
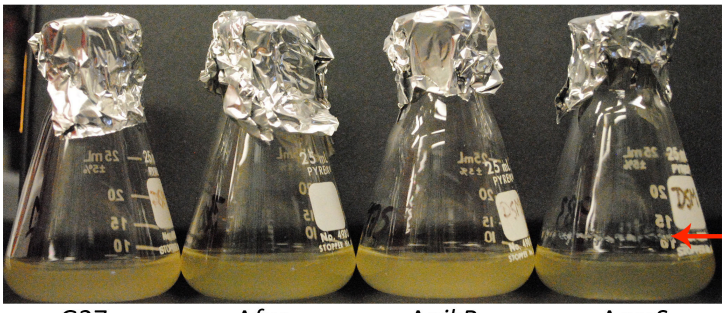



**Figure 4. Characterization of wild-type G27 and isogenic regulatory mutant strain growth phenotypes.**

**A.** OD 600 readings were taken after 6, 12, 20, and 48 hours. Mean ODU for each strain and time point is indicated by the symbol; error bars display standard error of the mean. Graph is representative of three biologically independent experiments. **B.** Enumeration of CFUs was performed after 6, 12, 20, 20, and 48 hours, corresponding to the OD values graphed in panel A. The mean CFU was calculated for each time point and is indicated by the symbol; error bars display standard error of the mean. Graph is representative of three biologically independent experiments.

A



G27 Δ*fur* Δ*nikR* Δ*arsS*

B

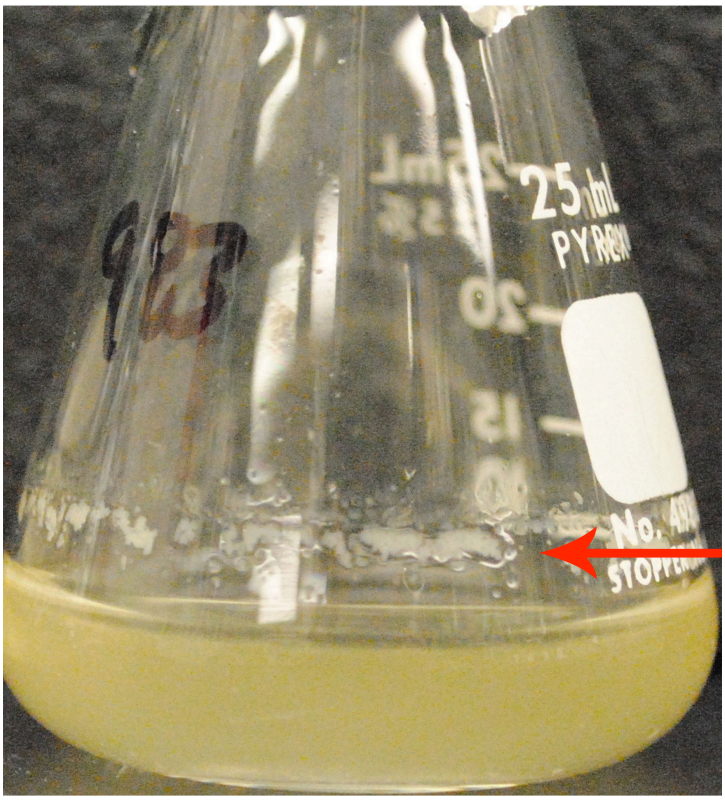



**Figure 5. Biofilm formation at air-liquid interface.**

**A.** Glass flasks similar to those used for the growth curve analyses are shown. From left to right, flasks containing WT G27, *Δfur*, *ΔnikR*, and *ΔarsS* were imaged after 72 hours of growth. In comparison to the other strains, a biofilm-like rim (red arrow) can be seen on the flask containing the *ΔarsS* strain. **B.** A close up images of the *ΔarsS* strain shows formation of the biofilm-like ring (red arrow) at the air-liquid interface.

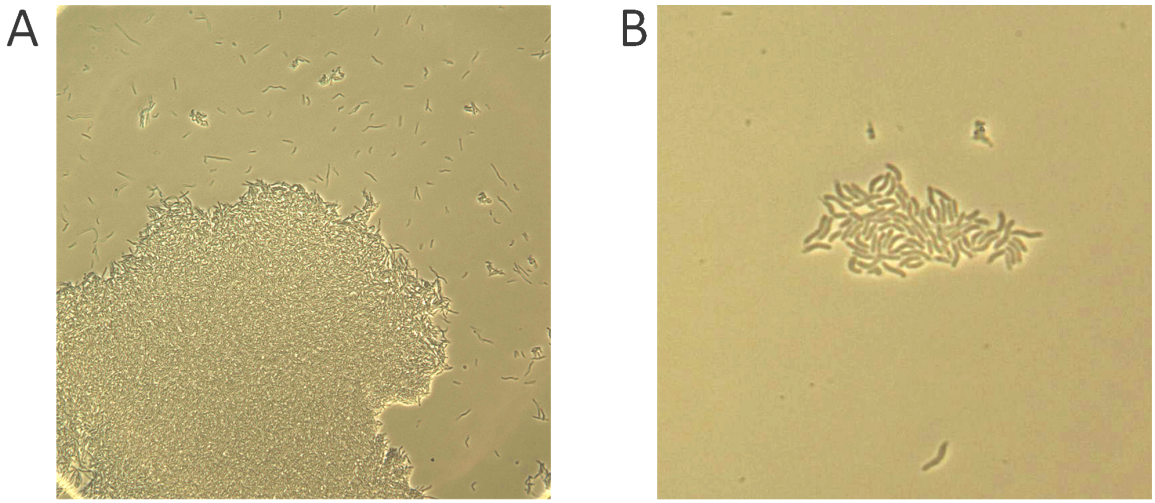



**Figure 6 (S1). Phase contrast images of the *ΔarsS* mutant strain.**
Phase contrast images of the *ΔarsS* mutant strain were taken after 72 hours of growth. **A.** 40X magnification and **B.** 100x oil emersion.

**Loss of ArsRS in G27 Exacerbates Biofilm Formation**

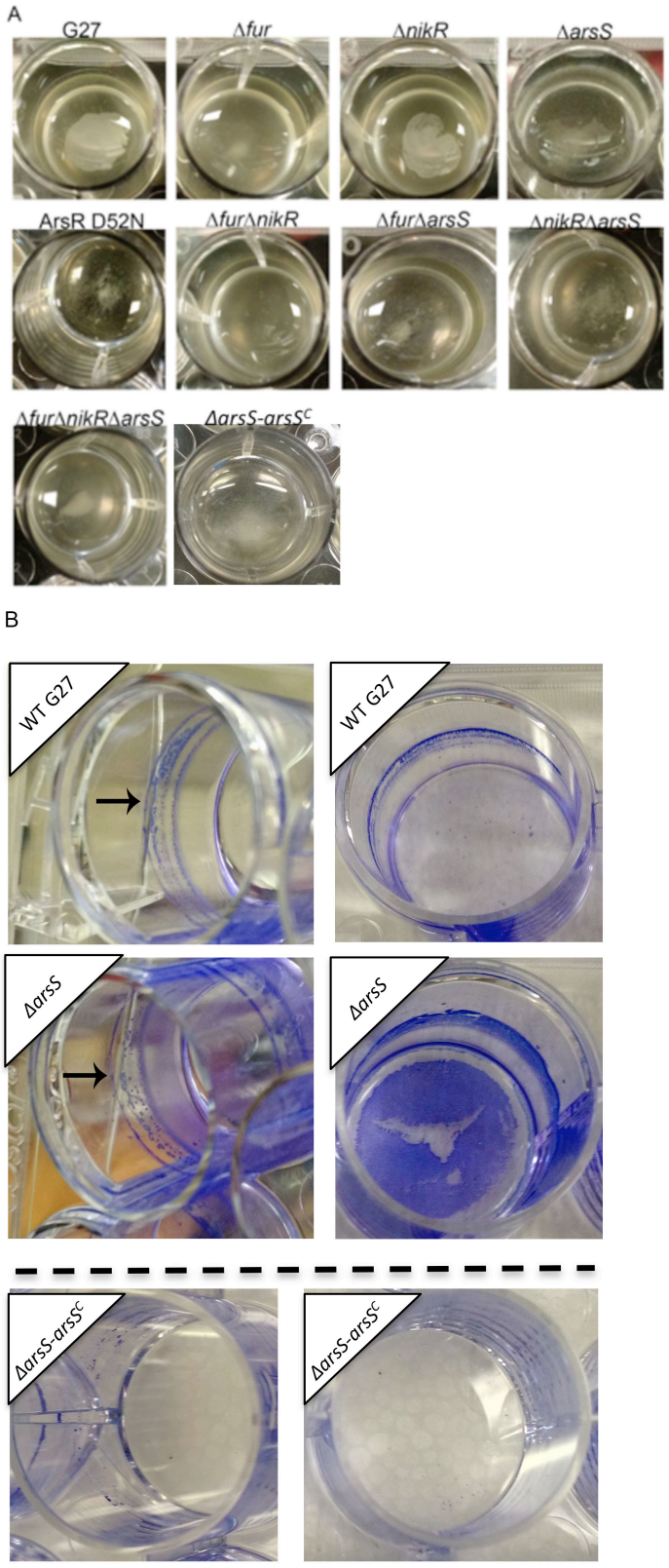
Based on the observations made during the growth curve analyses, we decided to further investigate the biofilm-like phenotypes observed with the isogenic strains. First, we conducted a time course study, in which biofilm formation was visually monitored over 72 hours (Figure 7A). Three distinct biofilm types were observed during this assay and some strains displayed multiple types of biofilm formation. The major phenotypes observed were biofilm formation at the liquid-air interface, biofilm formation on the bottom of the well, and formation of a pellicle. For wild-type G27, biofilm formation was not apparent at early time points but first became visible after 24 hours at the air-liquid interface (data not shown). At 48 and 72 hours of growth, noticeable biofilm had developed at the air-liquid interface; additionally, by 72 hours wild-type G27 also exhibited pellicle formation (Figure 7A). The *ΔnikR* strain showed similar patterns of biofilm formation. Conversely, *Δfur*, and *ΔfurΔnikR* appeared to show decreased biofilm formation at the liquid-air interface and did not form a pellicle. In contrast, *ΔarsS*, ArsR-D52N, *ΔfurΔarsS*, *ΔnikRΔarsS*, and *ΔfurΔnikRΔarsS* formed biofilms at the liquid-air interface as well as showed varying degrees of biofilm formation that extended to the bottom of the well; in some cases, the entire surface of the well was covered with biofilm (Figure 7B and data not shown). In comparison, little to no biofilm was observed on the bottom of the well with wild-type G27. Strains carrying the *arsS* or ArsR-D52N mutation also had a granular appearance indicative of the aforementioned aggregates (Figure 7A).

To quantitate biofilm differences across strains, biofilm formation was monitored using crystal violet staining. This analysis was conducted temporally in order to identify any differences in rate of biofilm formation as well as absolute quantity of biofilm

**Figure 7 (S2). Representative images of the various biofilm phenotypes.**

**A.** Representative images of the various biofilm phenotypes observed after 72 hours of growth in 24 well plates. The strains grown in each well are labeled. Wild-type G27 and *ΔnikR* both exhibit pellicle formation in addition to biofilm formation at the air-liquid interface, *Δfur* and *ΔfurΔnikR* display biofilm formation only at the air-liquid interface, and strains lacking *ΔarsS* or containing the ArsR-D52N mutation appear granular due to auto-aggregation and form biofilm at the air-liquid interface and on the bottom of the well. The *ΔarsS-arsS*$^{C}$ exhibited slight biofilm formation at the liquid-air interface as well as minimal aggregation. **B.** Representative images of G27 and the *ΔarsS strain* after crystal violet staining of 72 hour cultures. Both strains show prominent staining at the air-liquid interface; however, the *ΔarsS* mutant strain also shows significant straining along the side and bottom of the well (arrow). Below the dashed line are representative images from the *ΔarsS-arsS*$^{C}$ biofilm assays; complementation resulted in a loss of the hyper-biofilm phenotype.

**Figure 7 (S2). Representative images of the various biofilm phenotypes.**



formation. For the purposes of this study, the rate of biofilm formation was defined as the first time point that showed significantly increased crystal violet staining as compared to T6, which was the earliest time point for which biofilm was quantified. Based on this definition, a strain that showed significant biofilm formation after 24 hours would have an accelerated rate as compared to a strain that did not show significant biofilm formation until 48 hours. As shown in Figure 8A, wild-type G27 developed a significant biofilm after 48 hours; very little biofilm at the air-liquid interface was visible before 48 hours (data not shown). In comparison, all strains carrying the *ΔarsS* or ArsR-D52N mutation showed significant biofilm development after only 24 hours of growth, indicating that they formed biofilm at an accelerated rate (Figure 8A). Conversely, *Δfur, ΔnikR*, and *ΔfurΔnikR* showed delayed biofilm development and did not show significant biofilm formation until 72 hours of growth (Figure 8A).

In addition to rate analysis, we also directly compared the total amount of biofilm formed at each time point across all of the strains (Table 2). In light of the accelerated rate of biofilm formation observed for *ΔarsS,* it is perhaps not surprising that this strain also formed significantly more biofilm than wild-type G27, *Δfur, ΔnikR, and ΔfurΔnikR* at 24 hours (Table 2). Similarly, *ΔfurΔarsS*, *ΔnikRΔarsS*, and *ΔfurΔnikRΔarsS* all showed increased amounts of biofilm at 24 hours (Table 2). We found that the ArsR-D52N mutant did not form significantly more biofilm than wild-type G27 until after 48 hours, suggesting slight phenotypic differences between the various ArsS mutant strains (Table 2). Further support of this difference is seen by the fact that while *ΔarsS* no longer showed higher level of biofilm than wild-type at 72 hours, the ArsR-D52N mutant

**Figure 8. Biofilm quantification.**

**A.** Absorbance of crystal violet staining of biofilms was read at OD 590nm and plotted over time. (Λ) designates the point at which wild-type G27 displayed significant biofilm formation ($p = 0.041$). (~) indicates the point at which *ΔarsS* ($p < 0.0001$), *ΔfurΔarsS* ($p = 0.0045$), *ΔnikRΔarsS* ($p = 0.0012$), *ΔfurΔnikRΔarsS* ($p = 0.0067$), and ArsR-D52N ($p = 0.0403$) displayed significant biofilm formation. (#) identifies the point at which *Δfur* ($p = 0.0082$), *ΔnikR* ($p = 0.0092$), and *ΔfurΔnikR* ($p = 0.0451$) displayed significant biofilm formation. Data for each strain are plotted as the mean of three biologically independent experiments; error bars show the standard error of the mean. **B.** Absorbance of crystal violet staining of biofilms was read at OD 590nm and plotted over time. Biofilm formation by *ΔarsS-arsS*$^{C}$ is shown in comparison to the *ΔarsS* and wild-type G27 strains. (Λ) designates the point at which wild-type G27 displayed significant biofilm formation ($p = 0.0027$); (~) indicates the point at which *ΔarsS* displayed significant biofilm formation ($p < 0.0001$). Data for wild-type G27 and *ΔarsS* are plotted as the mean of at least three independent experiments. The mean plotted for *ΔarsS-arsS*$^{C}$ represents 5 isolates tested in up to two biologically independent experiments. For all data sets error bars show the standard error of the mean.

**Figure 8. Biofilm quantification.**

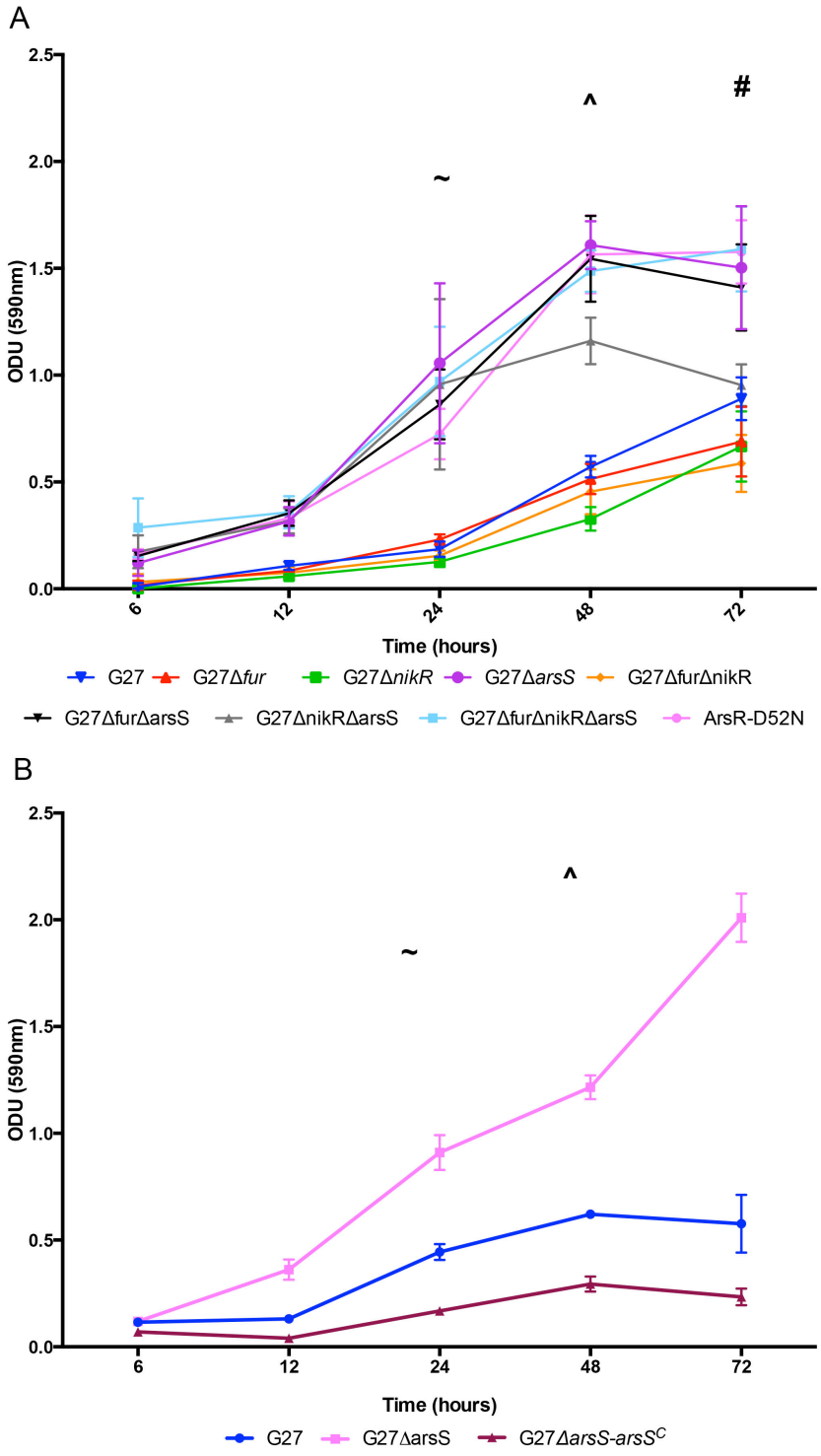


**Table 2 (S1). Comparison of the degree of biofilm formation at each time point**

| T24 | Adjusted P Value[#] | Summary |
|---|---|---|
| G27 vs. G27Δ*arsS* | 0.0008 | **,A |
| G27 vs. G27*ΔfurΔarsS* | 0.0231 | **,A |
| G27 vs. G27*ΔnikRΔarsS* | 0.005 | *,A |
| G27 vs. G27*ΔfurΔnikRΔarsS* | 0.004 | *** |
| G27Δ*fur* vs. G27Δ*arsS* | 0.002 | *** |
| G27Δ*fur* vs. G27*ΔfurΔarsS* | 0.0447 | *** |
| G27Δ*fur* vs. G27*ΔnikRΔarsS* | 0.0107 | **,A |
| G27Δ*fur* vs. *G27ΔfurΔnikRΔarsS* | 0.0086 | *** |
| G27Δ*nikR* vs. G27*ΔarsS* | 0.0003 | *** |
| G27Δ*nikR* vs. G27*ΔfurΔarsS* | 0.0091 | *** |
| G27Δ*nikR* vs. G27*ΔnikRΔarsS* | 0.0018 | **,A |
| G27Δ*nikR* vs. G27*ΔfurΔnikRΔarsS* | 0.0014 | *** |
| G27*ΔfurΔnik*R vs. G27*ΔarsS* | 0.0005 | *** |
| G27*ΔfurΔnikR* vs. G27*ΔfurΔarsS* | 0.0144 | *** |
| G27*ΔfurΔnikR* vs. G27*ΔnikRΔarsS* | 0.003 | **,A |
| G27*ΔfurΔnikR* vs. G27*ΔfurΔnikRΔarsS* | 0.0023 | *** |
| 48 | Adjusted P Value[#] | Summary |
| G27 vs. G27Δ*arsS* | < 0.0001 | **,A |
| G27 vs. G27*ΔfurΔarsS* | 0.0001 | **,A |
| G27 vs. G27*ΔfurΔnikRΔarsS* | 0.0004 | *** |
| G27 vs. ArsR D52N | < 0.0001 | **,B |
| G27Δ*fur* vs. G27Δ*arsS* | < 0.0001 | *** |
| G27Δ*fur* vs. G27*ΔfurΔarsS* | < 0.0001 | *** |
| G27Δ*fur* vs. G27*ΔnikRΔarsS* | 0.0366 | **,A |
| G27Δ*fur* vs. *G27ΔfurΔnikRΔarsS* | 0.0001 | *** |
| G27Δfur vs. ArsR D52N | < 0.0001 | **,B |
| G27Δ*nikR* vs. G27*ΔarsS* | < 0.0001 | *** |
| G27Δ*nikR* vs. G27*ΔfurΔarsS* | < 0.0001 | *** |
| G27Δ*nikR* vs. G27*ΔnikRΔarsS* | 0.0017 | **,A |
| G27Δ*nikR* vs. G27*ΔfurΔnikRΔarsS* | < 0.0001 | *** |
| G27ΔnikR vs. ArsR D52N | < 0.0001 | **,B |
| G27*ΔfurΔnik*R vs. G27*ΔarsS* | < 0.0001 | *** |
| G27*ΔfurΔnikR* vs. G27*ΔfurΔarsS* | < 0.0001 | *** |
| G27*ΔfurΔnikR* vs. G27*ΔnikRΔarsS* | 0.0149 | **,A |

| | | |
|---|---|---|
| G27*ΔfurΔnikR* vs. G27*ΔfurΔnikRΔarsS* | < 0.0001 | *** |
| G27*ΔfurΔnikR* vs. ArsR D52N | < 0.0001 | **,B |
| 72 | Adjusted P Value# | Summary |
| G27 vs. G27*ΔfurΔnikRΔarsS* | 0.0162 | *** |
| G27 vs. ArsR D52N | 0.0199 | **,B |
| G27Δ*fur* vs. G27Δ*arsS* | 0.0024 | *** |
| G27Δ*fur* vs. G27*ΔfurΔarsS* | 0.0115 | *** |
| G27Δ*fur* vs. *G27ΔfurΔnikRΔarsS* | 0.0005 | *** |
| G27Δfur vs. ArsR D52N | 0.0006 | **,B |
| G27Δ*nikR* vs. G27*ΔarsS* | 0.0016 | *** |
| G27Δ*nikR* vs. G27*ΔfurΔarsS* | 0.008 | *** |
| G27Δ*nikR* vs. G27*ΔfurΔnikRΔarsS* | 0.0003 | *** |
| G27ΔnikR vs. ArsR D52N | 0.0004 | **,B |
| G27*ΔfurΔnik*R vs. G27*ΔarsS* | 0.0004 | *** |
| G27*ΔfurΔnikR* vs. G27*ΔfurΔarsS* | 0.002 | *** |
| G27*ΔfurΔnikR* vs. G27*ΔfurΔnikRΔarsS* | < 0.0001 | *** |
| G27*ΔfurΔnikR* vs. ArsR D52N | < 0.0001 | **,B |
| G27*ΔnikRΔarsS* vs. G27*ΔfurΔnikRΔarsS* | 0.0417 | *,B |

# 2-way ANOVA with Tukey's multiple comparisons test

*,A significant 24 hours

*,B significant at 72 hours

**,A significant at 24 and 48 hours

**,B significant at 48 and 72 hours

*** significant 24, 48, and 72 hours

was still significantly higher than G27 at this time point. Also of note, while *ΔnikRΔarsS* formed significantly more biofilm than wild-type G27 at 24 hours, the increase in biofilm formation plateaued by 48 hours and was not significantly different than wild-type G27 at the later time points (Figure 8A and Table 2). Even though the difference was not significant, we note that G27 *Δfur*, *ΔnikR*, and *ΔfurΔnikR* seem to plateau at lower levels of biofilm formation than wild-type (Figure 8A). Finally, while biofilm staining intensity for *ΔarsS*, ArsR-D52N, *ΔfurΔarsS*, *ΔnikRΔarsS*, and *ΔfurΔnikRΔarsS* plateau by 48 hours, wild-type G27 biofilm appeared to steadily increase even at 72 hours (Figure 8A). In fact at 72 hours, only ArsR-D52N and *ΔfurΔnikRΔarsS* have significantly more biofilm than wild-type G27 (Table 2).

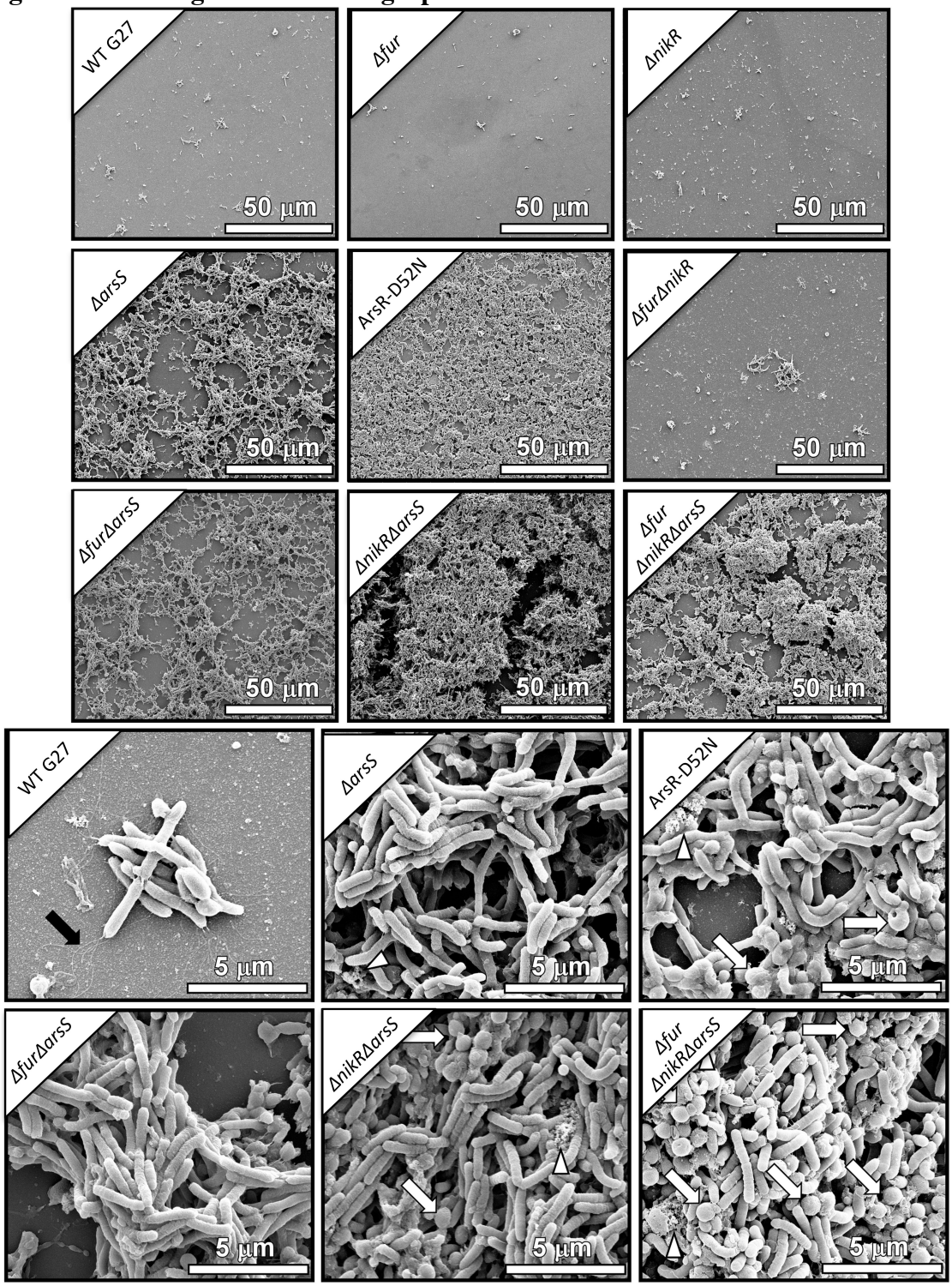
To confirm that mutation of *arsS* was responsible for the hyper-biofilm phenotype, we constructed a strain in which the *ΔarsS* mutation was complemented by *arsS* carried on the low-copy number pTM117 vector (36). Expression of *arsS in trans* completely suppressed the enhanced biofilm phenotype observed in the *ΔarsS* strain (Figure 8B). In fact, the complementation strain exhibited levels of biofilm formation that were less than those seen with the wild-type strain (Figure 8B and Figure 7B). This decrease may indicate that higher levels of ArsS are produced from the plasmid system.

In addition to the quantitative analysis, the parental strain G27 and the isogenic mutants were also analyzed by high resolution scanning electron microscopy (SEM) to evaluate changes in biofilm architecture on an abiotic substrate. SEM analyses revealed numerous cell-to-cell interactions occurring in biofilms and several cell morphologies including rod, curved rod/helix, and coccoid cells (Figure 9). Furthermore, numerous cell-surface features (such as small vesicles, flagella, and branched extracellular matrix

**Figure 9. Scanning electron micrographs of biofilm formation after 48 hours.**

Representative micrographs from all 9 strains are shown at 1000x magnification. In comparison to G27 wild-type, *Δfur*, *ΔnikR,* and *ΔfurΔnikR* which show sporadic adherent cells in small clusters, *ΔarsS*, ArsR-D52N, *ΔfurΔarsS*, *ΔnikRΔarsS*, and *ΔfurΔnikRΔarsS* all contain mats of cells arranged in 3D architectures. The three dimensional structure is especially apparent in the representative image of *ΔnikRΔarsS*. Biofilm forming strains are shown in comparison to wild-type in the lower half of the figure at 10,000X magnification. Of note, black arrows indicate flagella, white triangles show areas of extracellular matrix, and white arrows show coccoid cells. Coccoid cells were more prevalent in ArsR-D52N, *ΔnikRΔarsS*, and *ΔfurΔnikRΔarsS*.

**Figure 9. Scanning electron micrographs of biofilm formation after 48 hours.**



(ECM)) could be appreciated in the biofilm samples (Figure 9). After 48 hours, *H. pylori* G27 exhibited minimal adherence to the abiotic substrate. Conversely, the *ΔarsS* isogenic mutant exhibited enhanced biofilm formation and tertiary architecture as compared to the parental strain (Figure 9). The double mutants, *ΔarsSΔfur, ΔarsSΔnikR,* and the triple mutant, *ΔarsSΔfurΔnikR*, all displayed enhanced biofilm formation compared to the wild-type parental strain (Figure 9). As expected given the crystal violet results (Figure 8A), the ArsR-D52N mutant formed a dense biofilm similar to the *ΔarsS* strains; however, in contrast to the biofilms observed for *ΔarsS*, more coccoid cells were present and there appeared to be an increase in the ECM present for the ArsR-D52N mutant strain. Likewise, an increase in coccoid cells and ECM was also seen for *ΔarsSΔfurΔnikR and ΔarsSΔnikR,* perhaps reflecting slight differences in the fitness of these strains (Figure 9). *En masse*, these results indicate that the ArsRS signaling pathway may play an important role in the regulation of biofilm formation.

**Loss of ArsRS does not Influence *H. pylori* Motility**

In an attempt to gain some insight into the increased biofilm formation exhibited by the ArsRS mutant strains, we next considered mechanisms known to effect biofilm formation. Previous studies suggest that flagellar genes and motility may play a significant role in adherence and/or biofilm formation in *H. pylori* and the closely related *Campylobacter jejuni* (44; 91; 197). Thus, to determine if mutations in the ArsRS system lead to alterations in *H. pylori* motility, swimming motility was assessed for wild-type G27 and the enhanced biofilm formers *ΔarsS* and ArsR-D52N. As a control, a non-

motile mutant (*ΔflgR*) (134), which contains a deletion in the gene encoding the master flagellar regulator FlgR, was also included. At all time points assessed (days 3, 5 and 7 after inoculation in soft agar), there were no overt difference in the area of motility observed for the wild-type and ArsRS mutant strains (Figure 10). Thus, mutation of the ArsRS system does not appear to affect motility in *H. pylori* G27.

**Enhanced Surface Adherence and Auto-Aggregation in Strains with a Mutated ArsRS System**

Two additional factors that have been shown to contribute to biofilm formation are bacterial aggregation and initial attachment to a surface. Given the large bacterial aggregates that were observed during both the growth curve and biofilm assays (Figure 6 and 7), we reasoned that increased aggregation in strains lacking ArsS or containing the non-phosphorylatable ArsR-D52N mutation could contribute to the enhanced biofilm formation seen with these strains. Therefore, we assessed bacterial aggregation using a quantitative aggregation assay (219). Since biofilm formation occurs as early as 24 hours for some strains, 20 hour cultures were used for this assessment. While the observed differences were not statistically significant, we did observe increased auto-aggregation for the *ΔarsS* and ArsR-D52N mutant strains as compared to the wild-type strain (Figure 11A). Thus, mutation of the ArsRS system appears to affect bacterial auto-aggregation.

Another important step in biofilm formation is adherence to surfaces. To determine if changes in initial adherence contribute to increased biofilm formation for the *ΔarsS* and ArsR-D52N mutant strains, bacterial adherence to tissue culture treated polystrene plates were assessed. After a 2 hour of incubation on the plate, very few wild-type G27 were adherent (Figure 11B). In contrast, approximately 20% of the G27 ArsR-

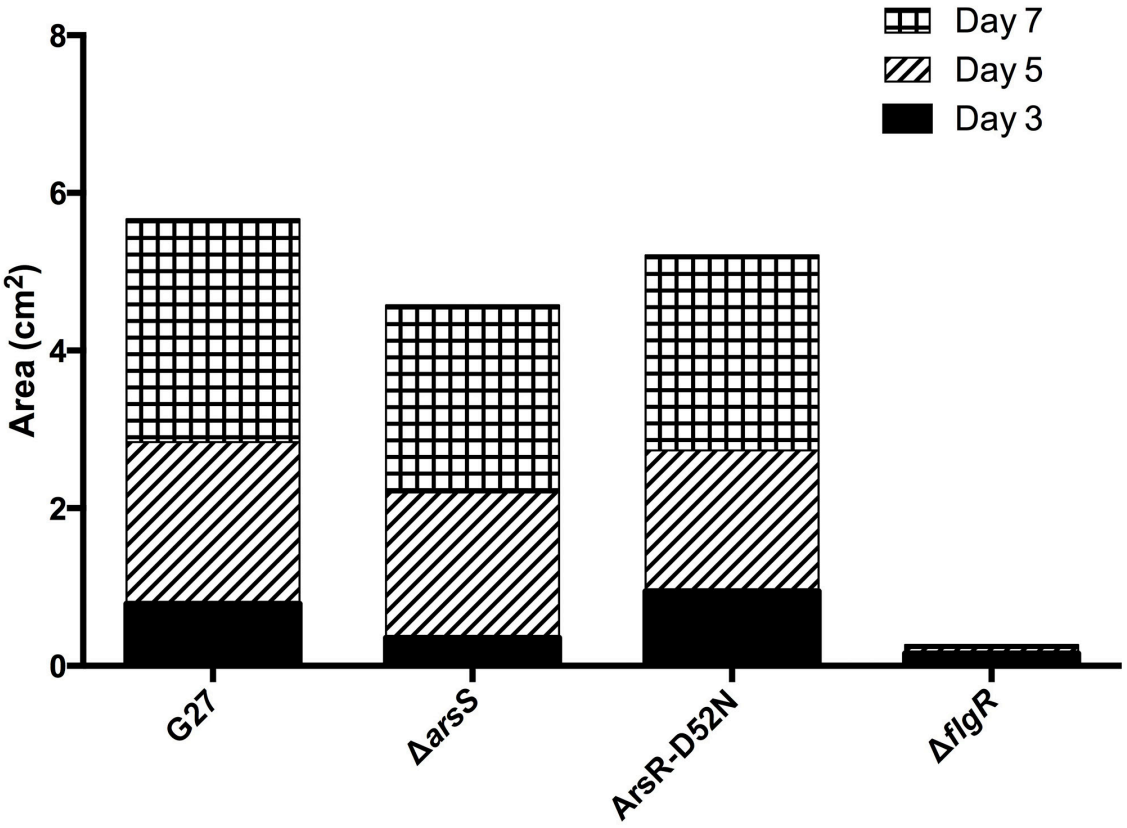



**Figure 10. Swimming motility of wild-type G27, *ΔarsS,* and ArsR-D52N strains.** The motility of wild-type G27 and the *ΔarsS* and ArsR-D52N mutant strains was compared using a soft agar assay. The area of the halo was measured at days 3, 5, and 7. The bar graph shows the cumulative area of motility as measured over 7 days; the solid black section corresponds to growth through day 3, the hatched segment shows additional area covered between days 3 and 5, and the checkered portion represents the increase from day 5 through 7. Bars represent the mean area of three independent biological replicates.

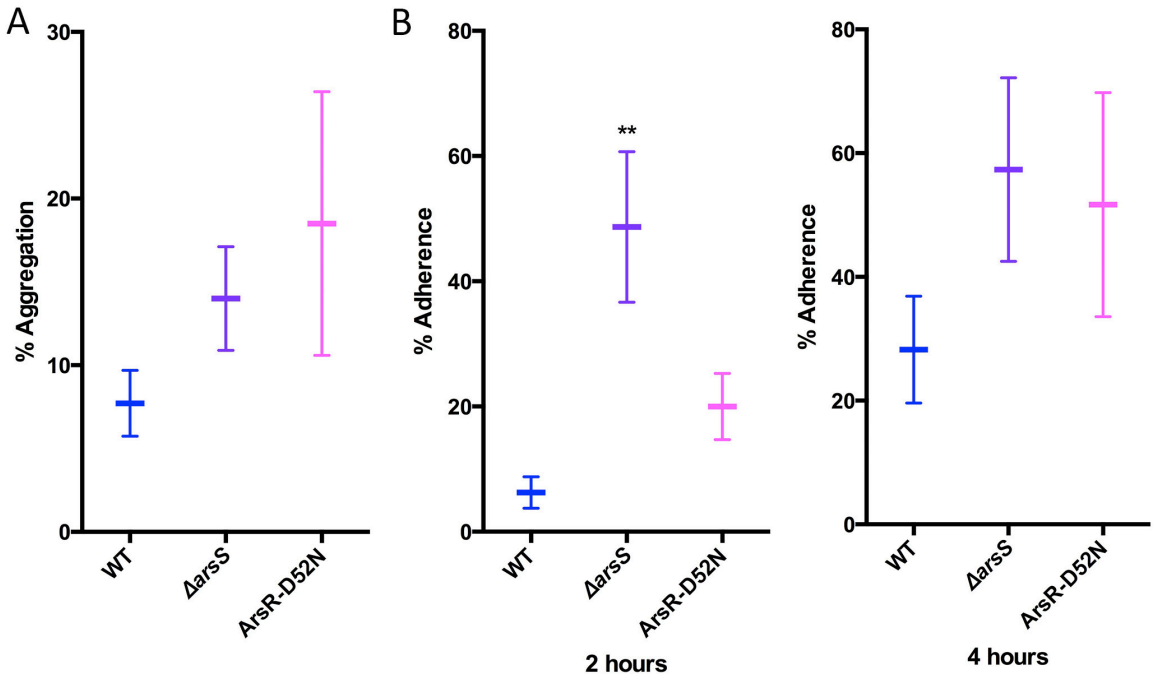



**Figure 11. Auto-aggregation and adherence of the *ΔarsS* and ArsR-D52N mutant strains.**

**A.** Auto-aggregation was assessed in wild-type G27 and the *ΔarsS* and ArsR-D52N mutant strains. Horizontal bars indicate the mean of three biological replicates and error bars represent the standard error of the mean. **B.** Surface adherence was assessed for wild-type G27 and the *ΔarsS* and ArsR-D52N mutants strain after two and four hours of incubation. ** indicates significantly more adherence by the *ΔarsS* mutant strain ($p = 0.0057$) compared to wild-type. Horizontal bars indicate the mean of three biological replicates and error bars represent the standard error of the mean.

D52N inoculum had adhered to the plate by 2 hours. The difference in adherence was even more pronounced with the *ΔarsS* strain; approximately 48% of the inoculum was adherent to the plate (Figure 11B), which was significantly more than wild-type G27 (p=0.0057). After 4 hours of incubation, the wild-type strain began adhering to the plates, with an average of 28% percent adherence (Figure 11B). At the same time point, adherence of the ArsR-D52N and *ΔarsS* mutant strains increased to approximately 50% (Figure 11B). These data suggest that an increase in speed and efficiency of bacterial surface adherence may contribute to the enhanced biofilm formation seen with the strains containing mutations in the ArsRS system.

## DISCUSSION

Given the paucity of regulatory proteins encoded in the *H. pylori* genome (175; 203), numerous studies have sought to define the roles of Fur, NikR, and ArsRS in the biology of *H. pylori*. The vast majority of these studies have focused on resistance to various stressful environments (78; 79; 140; 208), identification of the regulon of genes controlled by each regulatory factor (46; 62; 67; 68; 114; 121; 163), or elucidation of the contribution of each regulator to colonization in an animal model (32; 78; 141; 158). Despite the evidence supporting the fact that the Fur, NikR, and ArsRS regulons show considerable overlap, very few studies (32; 37; 61; 163) have attempted to analyze the combined affect of these regulatory factors in *H. pylori* biology. Herein, we constructed a series of isogenic regulatory knockout strains and evaluated *H. pylori* growth *in vitro*. We found that while differences in growth were minimal over the first 24 hours, at later time

points strains containing a *ΔarsS* mutation began to behave differently than the other strains; both CFU counts and OD 600 readings for strains lacking ArsS began to plateau and even decline after 24 hours (Figure 4A and B). Upon further examination, significant cellular aggregation (Figure 6) and adherence to the flask at the air liquid interface (Figure 5B) were observed for the *ΔarsS* mutant strains. Indeed, it is likely that the aggregation and adherence of these bacterial strains skewed both the CFU counts and OD readings due to loss of planktonic bacteria from the media. SEM analysis verified increased biofilm formation in the *ΔarsS* strains (Figure 9) as well as in the strain carrying the ArsR-D52N mutation. In addition to the biofilm phenotype, the SEM images also showed an increase in the number of coccoid cells in strains carrying the ArsR-D52N mutation or containing a deletion of NikR and ArsS in combination *(*Figure 9). In conjunction with the lower CFU counts observed for these three strains (Figure 4B), it is likely that these strains show a quicker transition to the coccoid form. Indeed, it has been shown that in wild-type *H. pylori* cells, the coccoid phenotype is triggered by environmental stressors such as nutrient deprivation (11). Therefore, one potential explanation for the increase in the coccoid state seen with the ArsR-D52N, *ΔnikRΔarsS,* and *ΔfurΔnikRΔarsS* strains may be the induction of an aberrant stress response due to the loss of key regulatory factors required to sense the environment. Intriguingly, biofilm formation has also been proposed as a mechanism to cope with stressful conditions (80). Together these data suggest that strains lacking a functional ArsRS system display an exacerbated stress response that results in enhanced and accelerated biofilm formation. Combining the *arsS* deletion with the *nikR* deletion further limits the ability of *H. pylori* to function normally and results in an increase in coccoid bacteria in addition to biofilm

formation. These findings not only highlight the benefit of utilizing combined isogenic regulatory mutants to tease out the coordinated regulation of *H. pylori's* stress response, but also suggest a role for the ArsRS system in regulation of biofilm formation.

Of note, a similar phenomenon has been seen in *C. jejuni* upon deletion of the CprS histidine kinase; loss of CprS led to bacterial aggregation and biofilm formation (195). Similar to the ArsRS system in *H. pylori*, CprR, which is the cognate response regulator to CprS, is essential in *C. jejuni* (195). Further investigation of the CprRS system showed that CprS mutants display increased expression of the flagella filament protein (FlaA), the major outer membrane protein (MOMP), and many other cellular envelope components (195; 196). Given the parallels between ΔCprS in *C. jejuni* and *ΔarsS* in *H. pylori*, we investigated motility and other phenotypes that may reflect changes to the outer membrane composition. We found that neither the *ΔarsS* nor ArsR-D52N mutant strains showed changes in motility as compared to the wild-type strain (Figure 10); this finding may not be entirely surprising given that previous studies have suggested that flagella do not contribute directly to *H. pylori*'s adhesive properties (44; 91; 211). In contrast to the motility phenotype, both *ΔarsS* and ArsR-D52N mutant strains showed an increase in aggregation and adherence compared to wild-type G27 (Figure 11). Both adherence and auto-aggregation in *H. pylori* are mediated by numerous cellular components, including LPS and an extensive list of OMPs (reviewed in (156)). Early exploration of the cell surface of *H. pylori* revealed that this bacterium is antigenically unique (64). A major contributor to this finding may be that, as compared to many other Gram- negative bacteria, *H. pylori*'s OMP profile lacks the highly abundant non-specific porins and is comprised of numerous less abundant species (5). Indeed, *H.*

*pylori* encodes over 60 OMPs, representing approximately 4% of the genome (5). Interestingly, transcriptomic studies of *H. pylori* strains lacking ArsS indicate that numerous OMPs are regulated by the ArsRS system; the majority of these OMPs appear to be repressed (121; 163). Therefore, the observed increases in adherence and auto-aggregation seen in the *ΔarsS* mutant strains may be due to de-repression and subsequent over expression of OMP(s) in these strains.

Biofilm formation in *H. pylori* infected individuals is well documented (33; 39; 41; 52), and a recent publication identified biofilm formation in an animal model of infection (17). *H. pylori* colonizes the gastric niche, which is highly acidic; however, *H. pylori* is a neutrophile and employs several mechanism to neutralize the immediate environment (171). Additionally, the pH in the stomach is not uniform; while the pH of the lumen can be as low as 1.0, the protective mucus layer covering the gastric mucosa acts as a buffer and maintains a more neutral environment (168; 208). In order to successfully colonize the stomach, *H. pylori* employs the use of flagella to burrow through the mucus layer, and upon reaching the gastric mucosa uses several proposed adhesins to anchor itself to the epithelium (9). Thus, it seems plausible that *H. pylori* may utilize pH changes as an environmental indicator of proximity to the epithelial layer. Upon exposure to the highly acidic environment in the gastric lumen, ArsS phosphorylates ArsR (208), and then phosphorylated ArsR targets specific promoters for gene regulation (130; 162). Conversely, as *H. pylori* enters the gastric mucus layer, it experiences increasing pH. Under these conditions, ArsRS is no longer phosphorylated, which leads to de-repression of ArsR~P specific targets. In this model, the ArsRS system

could be one mechanism by which *H. pylori* senses it is nearing the gastric epithelium as a way to modify expression of components that promote colonization (Figure 12). Deletion of ArsS or mutation of the phosphate accepting aspartic (D) residue at position 52 nullifies the response to changing pH of the ArsRS two-component system (176). Thus, in either of the mutant strains, ArsR will remain in an unphosphorylated state, which leads to deregulated expression of protein(s) that may contribute to enhanced biofilm formation, aggregation, and adherence observed in this investigation. This study opens several areas for future study; these include identification of the specific target(s) of ArsR~P that lead to the observed phenotypes, determination of the effect of environmental stressors such as acidic pH on biofilm formation, and investigation of biofilm formation in additional *H. pylori* strain backgrounds.

## ACKNOWLEDGEMENTS

We thank Dr. Robert Shanks for helpful suggestions regarding biofilm assays and Christina Ferreira for critical comments on the manuscript.

Research in the laboratory of DSM is supported by funding from the NIH and DoD; SLS is supported by the Henry M. Jackson Foundation for the Advancement of Science Fellowship. JAG and KPH were supported by the Dept. of Veterans Affairs (1IK2BX001701) and NIH Childhood Infectious Diseases Research Program (T32-AI095202. Core services, including use of the Cell Imaging Shared Resource, were performed through both Vanderbilt University Medical Center's Digestive Disease

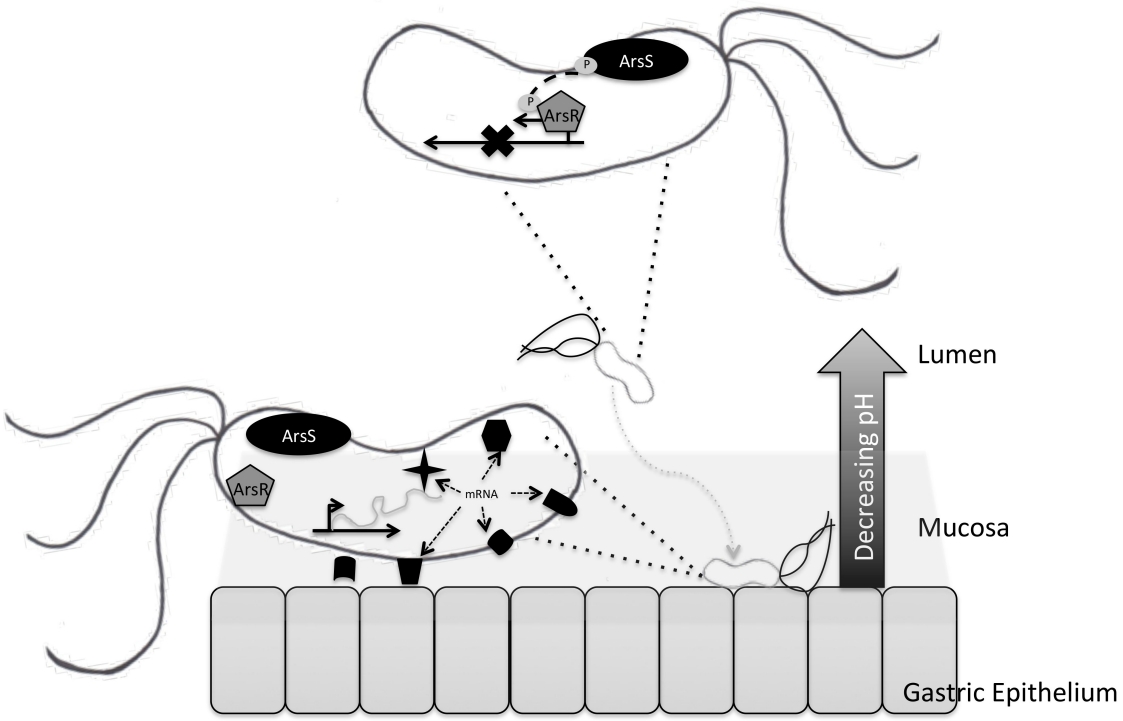



**Figure 12. Model of ArsRS regulation of colonization factors**

When located in the gastric lumen (upper) *H. pylori* is exposed to an acidic environment, which leads to autophosphorlyation of ArsS and phosphotransfer to ArsR. Subsequently, ArsR~P will bind and inhibit ArsR~P target genes. Once *H. pylori* has reached the more neutral gastric epithelium (bottom), ArsRS remains unphosphorylated. This results in alleviation of repression and expression of various factors, depicted here by the various shapes,which aid in colonization and biofilm formation.

Research Center, supported by NIH grant P30DK058404 Core Scholarship. Contents are the sole responsibility of the authors and do not necessarily represent the official views of the DoD, Dept. of Veterans Affairs or the NIH.

# CHAPTER THREE: The ArsRS Two Component System Regulates the Expression of the *H. pylori* Outer Membrane Protein, *homB*

**Stephanie L. Servetas, Ryan S. Doster, Jieun Kang, Jeong-Heon Cha, Jennifer A. Gaddy, D. Scott Merrell**

The work presented in this chapter is the sole work of S.L. Servetas with the following exception: J. Kang and J.H. Cha sequenced *homB* locus A and locus B promoter regions from clinical ioslates; R.S. Doster and J.A. Gaddy conducted the SEM studies.

## ABSTRACT

One elusive area in the *Helicobacter pylori* field is an understanding of why some infections result in gastric cancer, the 3rd leading cause of cancer related deaths, yet others persist unnoticed for the life-span of the individual. Even before the genomic era, the intraspecies diversity of *H. pylori* was well recognized and became an intriguing area of investigation with respect to disease progression. The association between the cytotoxins, CagA and VacA, and disease are well substantiated; however, *H. pylori* also encodes a unique repertoire of over 60 OMPs, several of which have been associated with disease outcome. Of these, the association between HomB and disease outcome was found to vary based on the population being studied. While the molecular roles for some of the disease-associated OMPs have been evaluated, little is known about the role that HomB plays in the *H. pylori* lifecycle. Thus, herein we investigate *homB* expression, regulation,

and contribution to biofilm formation. We show that in wild-type *H. pylori* G27, *homB* is expressed at a relatively low level until stationary phase. Furthermore, *homB* expression is suppressed at pH 5 in an ArsRS-dependent manner; mutation of *arsRS* resulted in an increase in *homB* transcript at all time-points tested over a 48 hour period. Taken together with our previous finding that ArsRS mutations lead to enhanced biofilm formation, we hypothesized that *homB* may contribute to biofilm formation by *H. pylori*. Indeed, subsequent biofilm analysis using a crystal-violet quantification assay and scanning electron microscopy (SEM) revealed that in *H. pylori* G27, *homB* contributes to biofilm formation by *H. pylori*. In summary, these data suggest that the regulation of OMPs via the ArsRS two-component system may be one mechanism by which ArsRS dictates biofilm development in a pH responsive manner.

## INTRODUCTION

*Helicobacter pylori* is an intriguing pathogen. Not only does it colonize the inhospitable gastric niche, but it does so with incredible efficiency. It is estimated that approximately 50% of the world's population is colonized by *H. pylori* (85). In a majority of cases, infection with *H. pylori* remains in a relatively quiescent, asymptomatic state. For those individuals that develop *H. pylori*-associated diseases, symptomatic *H. pylori* infections can result in gastritis, duodenal or gastric ulcers, gastric adenocarcinoma, or mucosa associated lymphoid tissue (MALT) lymphoma (22). Given the broad range of clinical manifestations associated with *H. pylori* infection, one of the most tantalizing avenues of research is the identification of the bacterial factors that

dictate disease progression. Indeed, the extensive heterogeneity of the *H. pylori* genome makes this pathogen an attractive target in the hunt for factors that can predict disease outcome.

Greater than 20% of *H. pylori* 's gene content varies between strains (65). Many of these auxiliary genes encode for proteins of unknown function; however, genes of known function include those encoding for DNA modifying proteins, LPS modifying proteins, outer membrane proteins, and members of the Cag pathogenicity island (PAI) (65). In fact, one of the main predictors of severe disease outcome, particularly gastric cancer, is the presence of the cytotoxin, CagA, which is encoded on the PAI. Interestingly, while CagA-positive strains are associated with an increased risk for development of gastric cancer as compared to CagA-negative strains, the risk among CagA-positive strains is further delineated based on conserved polymorphisms in the CagA C-terminal EPIYA motif (55). Thus, as demonstrated by the previous example, *H. pylori* 's panmictic structure is based not only on the presence or absence of genes, but also on conserved polymorphisms within certain genes.

As a cytotoxin that directly damages host cells, CagA presents an obvious target for study. Additionally, another class of genes that has been of particular interest in terms of an association with disease outcome are those that encode for outer membrane proteins (OMPs) (55). Gram-negative OMPs are unique because, unlike many integral membrane proteins that are comprised of alpha helices, OMPs are beta-barrel proteins (111; 179; 212). OMPs have been shown to function in a variety of roles: molecular transportation, antibiotic resistance, and virulence (111; 179; 212). For a pathogen that resides within a very specific niche, *H. pylori* encodes an extensive array of OMPs. In the majority of

Gram-negative bacteria, two to three percent of the genome is typically dedicated to OMPs (212). For a bacterium such as *Haemophilus influenza*, which has a genome size similar to *H. pylori*, this equates to approximately 40 proteins (212). In comparison, up to 60 OMPs are predicted to be encoded by *H. pylori*'s relatively small genome (1.6Mb); this equals approximately 4% of the genome (5; 7). This number becomes even more impressive when one considers that *H. pylori* is only predicted to encode 17 regulatory proteins (175).

Several *H. pylori* OMPs have been identified as adherence factors (reviewed in (156). While adhesins may not constitute traditional virulence factors, adherence is likely paramount to virulence for a persistent pathogen like *H.pylori*. This is supported by the fact that CagA is delivered to cells in a contact dependent manner. Without an adherent population, CagA would not reach the host epithelium (106). Indeed, in some strains the presence of *hopQ* was found to be required for CagA translocation (26). Additionally, the intimate attachment of *H. pylori* to the gastric epithelium is well documented (178) and believed to be crucial for maintenance of long term infection (29). Finally, given that surface adherence is the initiating step in biofilm formation, OMPs likely contribute to biofilm formation as well (148). In fact, biofilm formation is thought to contribute to many chronic bacterial infections, and the presence of *H. pylori* biofilms *in vivo* has been observed (33; 39; 41; 52).

The majority of *H. pylori* OMP research has focused on the largest family of OMP paralogs, the Hop family. Like many other facets of the *H. pylori* genome, OMP profiles vary greatly between strains. OMP diversity can result from several mechanisms, including intrastrain recombination. The presence of large stretches of homology has

been identified within small sub-families of OMPs. Subsequently, recombination between these OMP loci leads to gene duplication and/or deletion and results in strains with various combinations of OMPs. Intrastrain recombination between genes of the Bab family (*babA*, *babB* and *babC*) and the Sab family (*sabA* and *sabB*) have been reported (187; 198). The presence of multiple loci has also been associated with differential gene expression due to regulation via slip strand mispairing (86; 96; 118; 119; 187; 192). Mono/di nucleotide repeat regions have been found in association with over 30 genes in *H. pylori* and are known to contribute to phase variation in several OMPs (173).

The high level of genetic diversity seen with the OMPs has led several groups to investigate the association between OMP(s) and clinical outcome (reviewed in (156)). Interestingly, one OMP that showed an association with more severe disease outcome, HomB, is not a member of the Hop family (55; 156). *homB*, was identified in a screen for genes associated with formation of duodenal ulcers; the authors compared isolates from a duodenal ulcer and a gastritis patient (157). Further analysis showed that in individuals less than 40 years of age, *homB* was significantly associated with peptic ulcer disease (152; 155; 157). Following this initial publication, several groups have studied the association between *homB* and disease severity in a variety of populations(107; 108; 152; 155; 199).

Despite these initial findings, in comparison to other OMPs that are associated with disease outcome, very little is known about this OMP or the Hom family at a molecular level. From the original genome publication, we know that HomB is a member of a small OMP family, the Helicobacter outer membrane protein or Hom family. There are four members that make up the *hom* family: HomA, HomB, HomC, and HomD.

Based on the few molecular studies that have been conducted, HomB appears to be involved in adherence. Furthermore, serological evidence suggests that HomB is expressed *in vivo* (152). Additionally, two interesting observations that arose from the epidemiological studies were that (1) *homB* could occupy one of two loci, locus A and locus B, and (2) strains can carry a single or double copy of *homB* (108; 153; 154). To date, neither phase variation nor recombination has been demonstrated for *homB*; however, differences in distribution of *homB* location and copy number have been reported. This is particularly intriguing given that the *homB* genotype varies between Western and East Asian isolates and because the association between *homB* and disease severity is dependent on the population being studied (100; 107; 108; 152; 155; 199).

Given the lack of molecular detail concerning HomB, we compared the promoter regions of locus A and locus B, assessed *homB* expression, and examined the role of HomB in *H. pylori* biofilm formation. Our data indicate that both *homB* loci have nearly identical promoter regions, which we reason may maintain the presence of transcriptional binding sequences. Interestingly, subsequent evaluation of *homB* expression revealed that *homB* was suppressed in response to pH 5 shock and that this regulation occurs via the ArsRS two component system. Previous work from our lab identified an enhanced biofilm phenotype in strains in which ArsRS was mutated (182). Herein, we demonstrate that *homB* contributes to ArsRS-induced biofilm formation by *H. pylori*.

## MATERIALS AND METHODS

### Bacterial Strains and Growth Conditions

Strains and plasmids used in this study are listed in Table 3. Unless noted otherwise, *H. pylori* strains were grown as previously described (37). Briefly, *H. pylori* stock cultures were maintained at -80˚C in *H. pylori* freezing media (brain heart infusion broth (BD Biosciences) containing 10% FBS and 20% glycerol (EMD chemicals, Inc)). Freezer stocks were plated on horse blood agar (HBA) plates comprised of 4% Columbia agar (Neogean Corporation), 5% defibrinated horse blood (HemoStat Laboratoris, Dixon, CA), 2 mg/mL β-cyclodextran (Sigma), and an antibiotic/antifungal cocktail (10 μg/ml vancomycin (Amresco), 5 μg/ml cefsulodin (Sigma), 2.5 U/ml polymyxin B (Sigma), 5 μg/ml trimethoprim (Sigma), and 8 μg/ml amphotericin B (Amresco)). Following growth on HBA plates, *H. pylori* strains were grown in liquid culture. Liquid media consisted of brucella broth (Neogen Corporation) supplemented with 10% fetal bovine serum (FBS) (Gibco) and 10ug/mL vancomycin. Where indicated, 25μg/mL Kanamycin (Kan25) was used to supplement the media. All cultures were grown at 37°C, in gas evacuation jars, under microaerobic conditions (5% $O_2$, 10% $CO_2$ , and 85% $N_2$ ) generated with an Anoxomat gas evacuation and replacement system (Advanced Instruments, Inc); in addition, liquid cultures were grown shaking at 100rpm.

**Sequencing of *homB* Promoter Regions**

Twenty-one *homB* positive strains from a collection of South Korean *H. pylori* clinical isolates (K42, K248, K23, K16, K82, K34, K183, K25, K209, K36, K44, K131, K3, K197, K10, K165, K26, K259, K104, K107, K197) were chosen for promoter analysis (Table 3) (108). Hom locus A and locus B were amplified from these 21 clinical isolates and from DSM48 using the AF/AR and BF/BR primer pairs, respectively (Table

**Table 3. Plasmids and strains used in this study**

| | Description | Reference |
|---|---|---|
| Plasmids | | |
| | Description | Reference |
| pDSM 3 | pKSF-II | (47) |
| pDSM 215 | pTM177 | (36) |
| pDSM 463 | p199::G27-HP0073 promoter region | (37) |
| pDSM 920 | pBluescript::Δ*arsS*::kan | (209) |
| pDSM 1530 | pGEM-T Easy::HPG27_667::KanSacB | this study |
| pDSM 1546 | pGEM-T Easy::G27Δ*hom* | this study |
| pDSM 1532 | pGEM-T Easy::pUreA-*homB* | this study |
| pDSM 1533 | pGEM-T Easy::pUreA-*homB* | this study |
| pDSM 1538 | pGEM-T Easy::pFlaA-*homB* | this study |
| Strains | | |
| | Description | Reference |
| DSM 1 | WT G27 | (21) |
| DSM 215 | G27 pTM117, $Kan^R$ | (36) |
| DSM 983 | G27*ΔarsS* markerless | (37) |
| DSM 1442 | G27*Δhom* markerless | this study |
| DSM 1443 | G27 *ΔarsSΔhom* | this study |
| DSM 1446 | G27 ArsRD52N, unphosphorylatable, $Kan^R$ | (182) |
| DSM 1530 | *E.coli top10 pGEM HPG27_667::KanSacB* | this study |
| DSM 1531 | *G27 HPG27_667::KanSacB* | this study |
| DSM 1532 | Top10 pUreA-*homB* | this study |
| DSM 1533 | G27 pUreA-*homB* | this study |
| DSM 1538 | Top10 pFlaA-*homB* | this study |
| DSM 1539 | G27 pFlaA-*homB* | this study |
| DSM 1540 | G27*Δhom* pFlaA-*homB* | this study |
| DSM 1543 | *H. pylori* G27 ArsR-D52N *Δhom* | this study |
| DSM 1547 | *H. pylori* G27*Δhom* pTM117 | this study |
| DSM 1546 | G27*Δhom* pUreA-*homB* | this study |
| Strains for locus A and locus B promoter analysis | | |
| | Description | Reference |
| DSM48 | *H. pylori* strain 7.13 | (74) |
| K42 | Korean clinical isolate | (108) |
| K248 | Korean clinical isolate | (108) |
| K23 | Korean clinical isolate | (108) |
| K16 | Korean clinical isolate | (108) |
| K82 | Korean clinical isolate | (108) |
| K34 | Korean clinical isolate | (108) |
| K183 | Korean clinical isolate | (108) |

| | | |
|---|---|---|
| K25 | Korean clinical isolate | (108) |
| K209 | Korean clinical isolate | (108) |
| K36 | Korean clinical isolate | (108) |
| K44 | Korean clinical isolate | (108) |
| K131 | Korean clinical isolate | (108) |
| K3 | Korean clinical isolate | (108) |
| K197 | Korean clinical isolate | (108) |
| K10 | Korean clinical isolate | (108) |
| K165 | Korean clinical isolate | (108) |
| K26 | Korean clinical isolate | (108) |
| K259 | Korean clinical isolate | (108) |
| K104 | Korean clinical isolate | (108) |
| K107 | Korean clinical isolate | (108) |
| K197 | Korean clinical isolate | (108) |

4). Sanger sequencing of the promoter regions was carried out with the forward primers (Af and Bf) of each locus. The transcriptional start site was identified using the predictive data from Sharma et al 2010. The *homB* promoter regions from a handful of *H. pylori* genomes that are available in Pubmed (SJM180, J99, HUP-B14, Gambia94/24, G27, P12, HPAG1) were also used to supplement the data from the clinical isolates. In total, 30 sequences that extend 180bp upstream from the *homB* start codon were compiled and aligned using the Geneious software 9.1.6 (Biomatters Ltd.). A Weblogo was created based on nucleotide conservation (Figure 13 A and B). In addition, the *homB* promoter region from G27 was also run through the Basic Local Alignment Search Tool (BLAST) to get a more global view of the sequence conservation among *homB* promoters.

**Construction of G27 *ΔhomB***

All isogenic mutant constructs and strains were created using the *H. pylori* G27 parental strain (21). The *homB* gene in locus A (HPG27_667) of G27 was replaced with *kansacB*. The *kansacB* cassette was amplified with K4A and PFA3 from pDSM3 (47); primer sequences used in this study can be found in Table 4. The *kansacB* cassette was then fused to the up-stream region of G27 Locus A (amplified with LocA_UpF/LocA_UpR) and the downstream region of G27 Locus A (amplified with LocA_DnF/LocA_DnR) by SOE PCR. The full construct was then amplified with FidelTaq and cloned into pGEM T-easy. Transformants were detected by blue-white screening and confirmed by PCR and $Kan^R$. One colony was archived as pDSM1530. Wild-type G27 (DSM1) was transformed with pDSM1530, and transformants were selected on HBA $Kan^{25}$, confirmed by PCR, and one colony was archived as DSM1531.

## Table 4. Primers used in this study

| Primer Name | Sequence (5'-3') | Reference |
|---|---|---|
| | *Δhom* construction primers[A] | |
| K4A | GTGCTCGAGCCCGGGCGAACCATTTGAGGTGA | (47) |
| PFA3 | GCTCTAGACCCGGGTATAAGCCCATTTTCATGC | (47) |
| LocA_UpF | GTAAATCGGTGGTATTGAAG | this study |
| LocA_UpR | TCACCTCAAATGGTTCGCCCGGGCTCGAGCACATATCAATATTCCAGTCAGCAC | this study |
| LocA_DnF | GCATGAAAATGGGCTTATACCCGGGTCTAGAGCATCAACGCTTCTTAATAAGC | this study |
| LocA_DnR | ACAAATGATCGTTATTGTGG | this study |
| | ArsR D52N confirmation primers | |
| Up1F | GGGGATTTTTTGAGCGTTGAG | (182) |
| Dn4R | CGCAAACGGCCAATGATCAC | (182) |
| | *ΔarsS* Confirmation primers | |
| HP0165_del_ver_F | TGAAAGCATTGCGATTGAGA | (182) |
| HP0165_del_ver_R | AAAACGGCTTTGATGCCTAA | (182) |
| | *hom* expression vector contruction primers[B] | |
| G27homB_FXbaI | agagag*TCTAGA*ATGAGAAAACTATTCATCCCACTTTTATTATTC | this study |
| G27homB_RPstI | agagag*CTGCAG*CACGCTCAAAACACCCAC | this study |
| G27pflaA_FKpnI | agagag*GGTACC*AGCCCATTTTCATGCTCCTAATT | this study |
| G27pflaA_RXbaI | agagag*TCTAGA*TTATAAAAAACCCAAAGGCATCCTTG | this study |
| | Promoter Sequencing Primers | |
| Af (F1-jhp0648/HP0709) | TAATTTCGCGCAAAAACATC | (157) |
| Ar (R1-jhp0650/HP0711) | ATTCCAGCGCCTAATGGAC | (157) |
| Bf (F1-jhp0869/HP0935) | AAGAGGATTGCGTGGTGGAGTTG | (157) |
| Br (R1-jhp0871/HP0936) | GGGTTGCCTTTGGGCTTGGA | (157) |
| | qRT-PCR primers | |
| G27_16S RT-F | ATGGATGCTAGTTGTTGGAGGGCT | (37) |
| G27_16S RT-R | TTAAACCACATGCTCCACCGCTTG | (37) |
| G27homB FWD Set 4 | CCCGGAGAGGCATTTGATAG | this study |
| G27homB REV Set 4 | ACACCCTTTGCGTGTTAGT | this study |

[A]regions of overlapping sequence for SOE PCR are underlined

[B]restriction sites are italicized; lower case nucleotides are nonspecific nucleotides added to facilitate restriction digest

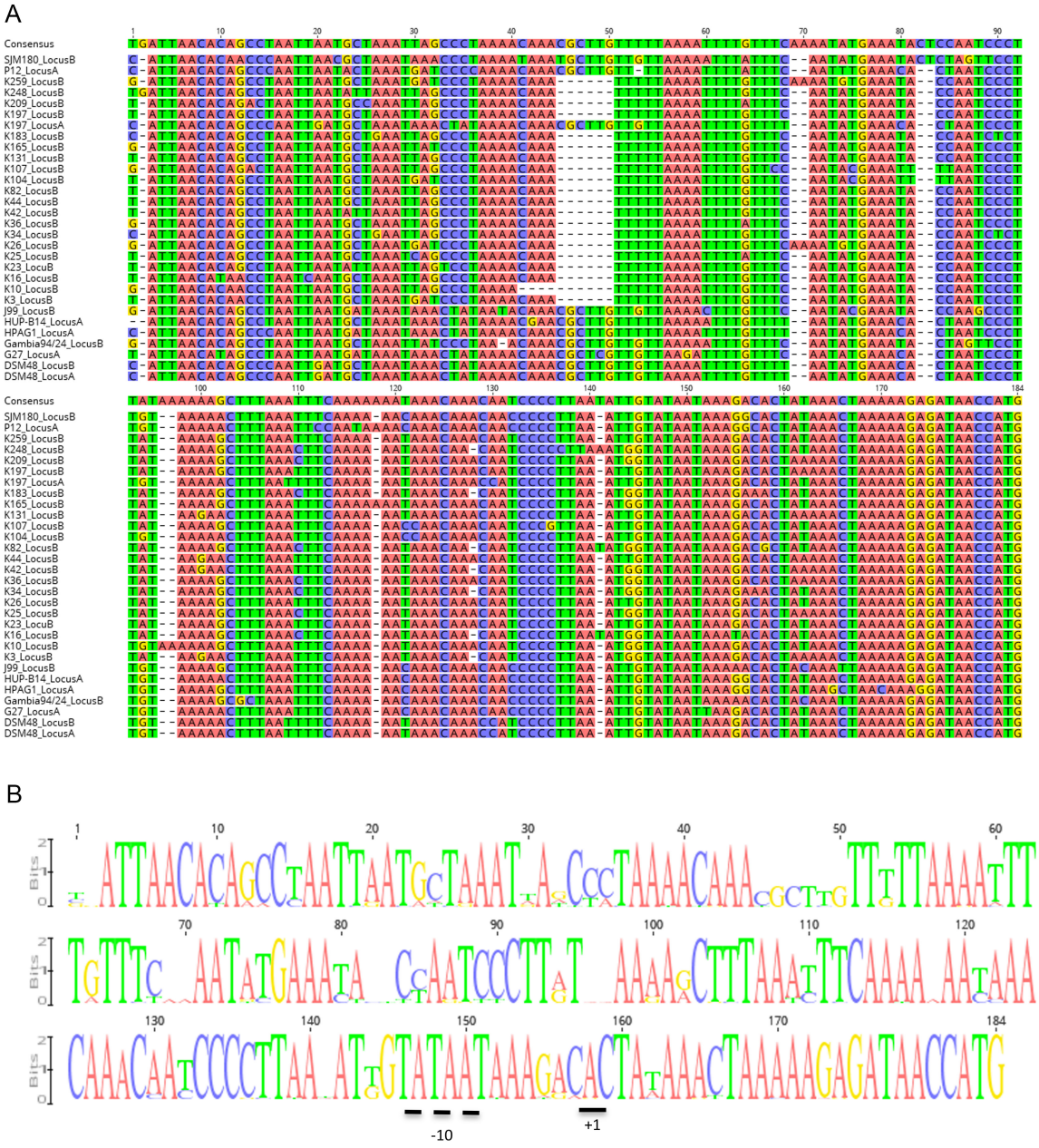



**Figure 13. *homB* locus A and locus B promoter alignment**

**A.** A pairwise alignment and **B.** Weblogo were generated from sequences from 30 strains. The overall conservation of the 180 nucleotides upstream of the *homB* genes from locus A and locus B are shown. The ATG at nucleotide position 168-170 is the start codon for the *homB* gene. The +1 transcriptional start site and the -10 sequence are both indicated. The larger the letter, the more highly conserved the corresponding base pair. To indicate nucleotides where substitutions/deletions are present, the letter(s) at these positions do not reach the full height.

To create a clean deletion for use as the background for multiple knockout mutations, pDSM1530 was digested with XbaI and XhoI to remove *KanSacB* but leave the Locus A flanking sequence. The up and downstream regions of *hom* locus A were then ligated together to create an empty *hom* Locus A. The ligation product was transformed into chemically competent Top10 *E. coli,* and transformants were selected on LB+Amp (100μg/mL). PCR and sequencing were used to confirm the empty locus, and one colony was archived as DSM1546. DSM1531 was transformed with pDSM1546 to create a markerless *hom* deletion in the G27 strain background. Transformants were selected for $Suc^R$ and then screened for $Kan^S$. $Suc^RKan^S$ colonies were then confirmed by PCR. Presence of an approximately 1 kB band following PCR amplification with LocA_UpF and LocA_DnR showed that *kansacB* had been eliminated, and the upstream and downstream regions of locus A were now fused. One *G27ΔhomB* colony was selected and archived as DSM1442.

**Construction of *ΔhomB arsRS* Double Mutant Strains**

Both *ΔarsS ΔhomB* and ArsR-D52N *ΔhomB* were constructed in the DSM1442 strain background. To construct G27*ΔarsS ΔhomB*, DSM1442 was transformed with pDSM920, which carries a *ΔarsS* ::*kan* cassette. Transformants were selected on HBA+ Kan25. The *arsS* deletion was confirmed by PCR and sequencing with HP0165_del_ver_F and HP0165_del_ver_F. In addition, LocA_UpF and LocA_DnR were used to ensure the *ΔhomB* deletion. To demonstrate reproducibility of the *ΔarsS ΔhomB* phenotype, three separate colonies (designated *ΔarsS ΔhomB*) were selected and

were tested in the biofilm assay. Given that all 3 isolates behaved similarly, one isolate was archived as DSM1443.

The G27 ArsR-D52N *ΔhomB* strain was constructed in a similar manner. First, The ArsR-D52N Kan$^R$ SOE cassette from DSM1446 was PCR amplified with Up1F and Dn4R. The PCR product was used to transform DSM1442. Colonies were selected on HBA+ Kan25. The D52N mutation in ArsR was confirmed by PCR amplification of *arsR* followed by ApoI digestion of the amplicon. One colony with the correct digestion pattern was confirmed by sequencing and archived as DSM1543.

**Construction of *homB* Over Expression Strains**

Plasmids containing *homB* fused to either the *ureA* or *flaA* promoters were generated to assess the effect of increased *homB* expression on biofilm formation. First, pDSM463 (P199+*ureA* promoter) was digested with XbaI and PstI to drop out the 730bp GFP gene. *homB* was amplified from G27 with G27*homB*_FXbaI and G27*homB*_RPstI. The *homB* amplicon was digested with PstI and XbaI and was then ligated with the similarly digested p463. The ligation was transformed into Top10 *E.coli* and transformants were selected on LB+Kan25. Kan$^R$ colonies were screened by PCR for the presence of *homB*. Upon confirmation, one colony was archived as DSM1532. pDSM1532 was used to transform DSM1 and transformants were selected on HBA+Kan 25. *homB* over-expression was confirmed by qRT-PCR (primers listed in Table 4) for several Kan$^R$ colonies. Similar levels of *homB* expression were observed among all tested strains (data not shown) and one strain was archived as DSM1533. Following the same procedure, DSM1442 was also transformed with pDSM1533 and the resulting strain was

archived as DSM1546. A lower expressed promoter for *homB* expression was also chosen, the *flaA* promoter. To construct the p*flaA-homB* vector, p*flaA* was amplified first with G27pflaA_FKpnI and G27pflaA_RXbaI from the pDSM3 plasmid. The amplicon and pDSM1532 were digested with KpnI and XbaI. Digestion of pDSM1532 resulted in the release of the *ureA* promoter, which was then replaced with the similarly digested *flaA* promoter. Ligation products were transformed into Top10 *E. coli* and selected for on LB+ Kan25. $Kan^R$ colonies were screened for the correct promoter by PCR and sequencing. One correct colony was archived as DSM1538. Using the same method as described above for construction of DSM1533 and DSM1546. DSM1 and DSM1442 were transformed with the p*flaA-homB* vector, pDSM1538, and the resulting transformants were archived as DSM1539 and DSM1540, respectively.

### *H. pylori* Growth Conditions for RNA Isolation

Briefly, bacterial cultures were patched from freezer stocks onto HBA plates. After 48 hours of plate growth, sterile swabs were used to resuspended bacteria in 5mLs of *H. pylori* liquid media. After 24 hours, cultures were standardized to an OD of 0.05 ODU (OD control culture) in fresh liquid media and allowed to grow for a specified length of time depending on the experiment. To analyze growth dependent expression, one 60mL OD control culture were separated into 2 separate 125mL flasks, each containing 30mLs, and cells were collected from alternating flasks at each indicated time point. At each time point, 5mLs of bacterial culture was removed, captured on 0.45µM cellulose nitrate membrane filter (GE Healthcare), and flash frozen in liquid nitrogen. Flash frozen filters were stored at -80˚C until RNA preparation.

Shock experiments were carried out as previously described (37; 140). Briefly, OD control cultures were prepared in a similar manner as described above. 10mL OD control cultures were collected after 20 hours of growth, and separated into two equal portions. One portion was immediately collected on a filter and represented the $T_0$ pre-shock sample. The second portion was resuspended in 5mLs of media, which represented one of three environmental shocks: low pH (pH 5), excess nickel (10 μM $NiSO_4$ (Sigma-Aldrich)), or limited iron (200μM 2-2'-dipyridyl (DPP)). pH shock and excess nickel were maintained for 90min, and iron limitation was carried out for 60min. All cultures were incubated under microaerobic conditions with shaking at 100rpm as described under growth conditions. After the respective incubation period, cultures were collected by the filter method, flash frozen and stored at -80˚C.

**RNA Isolation, cDNA Synthesis and Quantitative Real-Time PCR**

RNA isolation was conducted as previously described (201). RNA integrity was determined by visualization on a 2% agarose gel. cDNA synthesis and Quantitative Real-Time PCR (qRT-PCR) were conducted as previously described with a few modifications (Gilbreath e al 2012). Each cDNA synthesis reaction was carried out using the Quantitect reverse transcriptase kit (qiagen) and 500ng of RNA. A control reaction, to ensure there was no genomic DNA contamination, was carried out for each sample by excluding the reverse transcriptase enzyme (NoRT). Following cDNA synthesis, qRT-PCR was performed using an intercalating dye, SYBR green, for *homB* expression and the 16S rRNA internal reference gene. qRT-PCR primers can be found in Table 4. The qRT-PCR reaction mixture was comprised of 1x SYBR green RT-PCR maxter mix, 3pmol of each

forward and reverse primer, and 1$\mu$l of template (cDNA or the NoRT control). Reactions were brought to a final volume of 20$\mu$l with DNase free water (IDT). The following 3-step cycling protocol was used: a 10 min hold at 95˚C (initial denature) was followed by 40 cycles of 10s at 95˚C (denature), 10s at 50˚C (annealing), and 10s at 72˚C (extension). SYBR green fluorescence was measured at the end of the extension step. Relative gene expression of each gene compared to the 16S internal control was calculated by the $2^{-\Delta CT}$ method; relative fold difference in gene expression was calculated using the $2^{-\Delta\Delta CT}$ method. Gene expression studies represent a minimum of three biologically independent experiments.

**Biofilm Analysis**

Biofilm assays were conducted as previously described (Servetas 2016). Briefly, biofilms were grown in a 24-well tissue culture treated plate. Cultures were added to each well to an $OD_{600}$ of 0.1 ODU in 1mL of HBA liquid media. Biofilms were allowed to develop for 24, 48, and 72 hours. At each time point, planktonic cells were aspirated, wells were washed 2x with PBS, biofilms were fixed with methanol, and then stained with crystal violet. Crystal violet was rinsed away 3x with distilled water and then solubilized to assess relative quantity of biofilm. Solubilized crystal violet was read at $OD_{590}$. Data shown represents a minimum of 3 biological replicates. Scanning electron microscopy (SEM) assays were preformed as previously described (182). Fresh liquid media in 12 or 24 well plates was inoculated with a 1:10 from overnight cultures (0.3-0.7 OD 600). Cell-cultured treated cover-slips were placed at the bottom of each well to capture biofilm formation. Samples were treated and visualized by SEM as previously

described (77). SEM images selected are representative of three biological replicates.

**Cellular Adherence**

Cellular adherence was assessed as previously described (Servetas 2016). Briefly, 1 ODU of an overnight liquid culture was resuspended in 1mL of *H. pylori* liquid media. 100µL of each strain was inoculated into 6 separate wells of a 96 well tissue culture treated plate (Corning). The 96 well plate was incubated for 2 hours under microaerobic conditions (5% O, 10% $CO_2$, and 85% $N_2$) with shaking (110 rpm). Following a 2 hour incubation, liquid was aspirated from 3 wells/strain, and the same wells were washed 2x with PBS, after which 5% alamarBlue (Invitrogen) that was diluted in *H. pylori* liquid medium was added to each well. In addition, 5% alamarBlue solution was also added to three liquid media only wells to serve as the blank. The 3 wells/strain that remained untouched were treated with 100µL of 10% alamarBlue that was likewise diluted in *H. pylori* liquid media. Plates were wrapped in foil to prevent light exposure and were incubated for 2 hours at 37°C, under microaerobic conditions, in a static incubator. Following the 2 hour incubation, the plate was read at an excitation wavelength of 544 nm and emissions wavelength of 590nm. Relative fluorescence units (RFU) for each well were calculated and three wells per biological replicate were averaged. The averaged values were used to calculate percent adherence per strain using the following calculation: (RFU of the washed wells/RFU inoculum) × 100. The data shown represent a minimum of three biologically independent experiments.

**Aggregation Assays**

Flow cytometry was used to analyze bacterial aggregation. Bacterial cultures were inoculated into 24 well tissue culture treated plates as described in the biofilm assays. Following 20 hours of growth, cultures were transferred to 1.5mL microfuge tubes for analysis. As a control for identification of single bacterial cells, half of the wild-type culture was filtered through a 1.2-$\mu$m Acrodisc PSF syringe filter (Pall) to disrupt bacterial clumps and remove debris. The presence of single *H. pylori* cells was confirmed by microscopy (Figure 14 A). Samples were analyzed using the Micro Apogee flow system (Hemel Hempstead, England) with the following parameters: flow rate 1.5ul/min, SALS PMT voltage of 249, and LALS PMT voltage of 247. 500,000 events were analyzed per sample with the SALS and LALS threshold set to 150 based on the filtered wild-type sample. Events were viewed on a scatter plot displaying small angle light scatter (SALS) (more commonly known as forward scatter) verses large angle light scatter (LALS) (more commonly known as side scatter). Data was analyzed using FlowJo 10.2. The first gate was drawn to encompass a majority of the filtered events (>90%) and corresponds to single cell bacteria. Subsequent gates were drawn to encompass regions of increasing SALS values, which correspond to increasing size, which has been previously shown to correspond to bacterial aggregates (188).

**Statistical Analysis**

*homB* expression over time was analyzed by two-way ANOVA with Tukey's correction. The rate of biofilm formation, which is defined as the first time point where the crystal violet quantification was significantly greater than the blank, was determined by a two-way ANOVA with Dunnett's correction. Differences in amount of biofilm

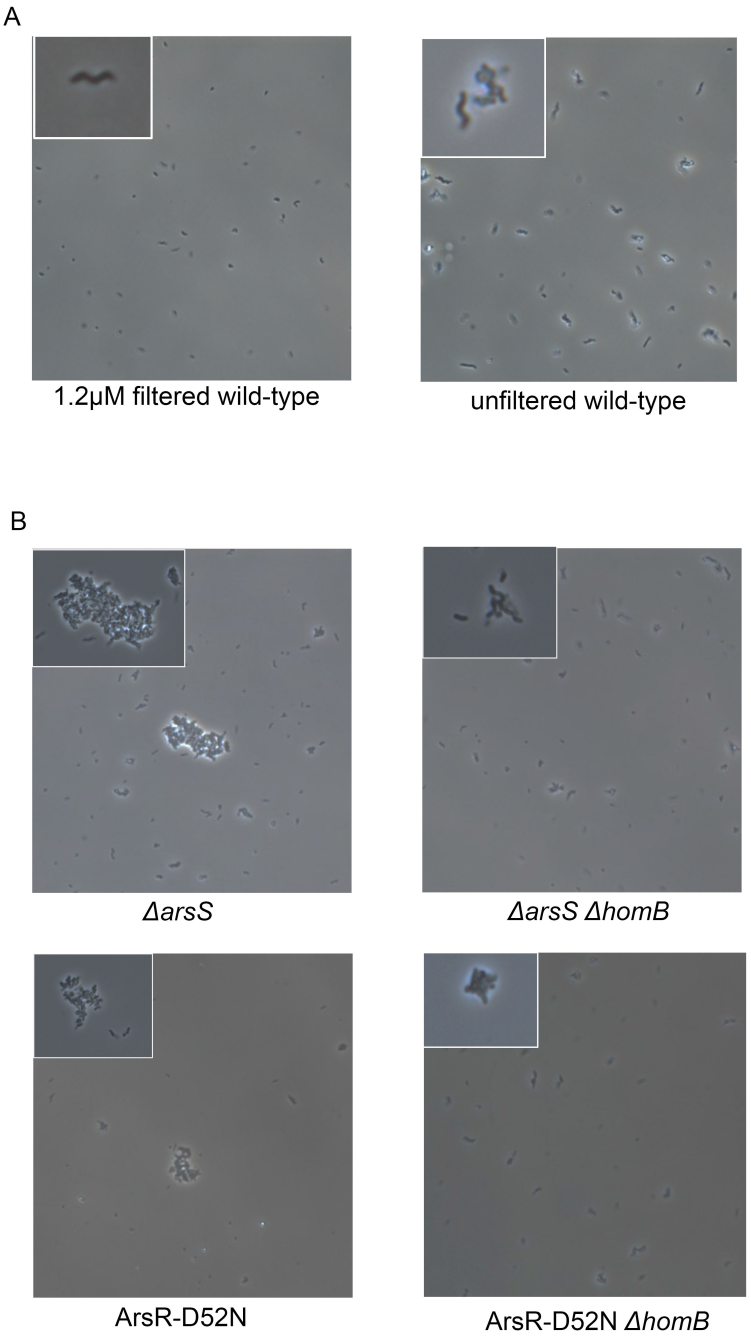



**Figure 14 (S1). Bacterial aggregation observed at 40X (100X inset)**
**A.** A representative image of the single cell bacteria observed following passage of wild-type bacteria through a 1.2μM filter; for comparison, small bacterial clumps are visible in the unfiltered wild-type sample. **B.** Representative images of *ΔarsS* and ArsR-D52N show small bacterial clumps alongside large bacterial aggregates. In comparison, it is difficult to find large bacterial aggregates in the *ΔarsS ΔhomB* and ArsR-D52N *ΔhomB* strains.

formation between strains at each time point were analyzed using a one-way ANOVA with Tukey's correction for multiple comparisons. Aggregation and Adherence data were analyzed by one-way ANOVA with Sidak's test of multiple comparisons. All analyses were conducted with GraphPad Prism 6.

## RESULTS

### The *homB* Promoter Region is Highly Conserved Regardless of Locus

HomB can be found at two different chromosomal loci. To assess *homB* promoter variability based on the chromosomal locus, we first used BLAST to determine the overall conservation between *homB* promoter regions available in Genebank. We found that the promoter regions found in both loci were highly conserved. For instance, the BLAST analysis using the entire intergenic region upstream of *homB* (locus A) from G27 returned matches with complete coverage (228/228) that showed greater than 90% identity (range from 100 (228/228) to 91% (207/228)) to intergenic DNA from strains that carried a *homB* gene at locus A. When this comparison was extended to include strains that carried a *homB* gene at locus B, approximately 85% coverage (194/228) with identity of approximately 90% (ranging from 99% (193/194) to 88% (169/193)) was seen. In some rare cases, even when the locus was empty, parts of the promoter region could still be found. In at least one strain (sequence ID: CP006610.2) although locus B was empty, there was a 136bp fragment present that showed 91.9% identity to the *homB* promoter region. To better observe this high level of conservation, the sequences from 30

*homB* promoter regions (approximately 180bp upstream from the *homB* translational start site) were aligned. The pairwise alignment found a range of 80-100% identity between individual promoters (Figure 13A). As evident in the weblogo, in addition to regions such as the -10, which is 100% conserved among the 30 loci, there were several other highly conserved regions (Figure 13B). This high degree of conservation in the intergenic promoter region regardless of chromosomal locus may indicate that this area is conserved to maintain binding sequences for regulatory proteins.

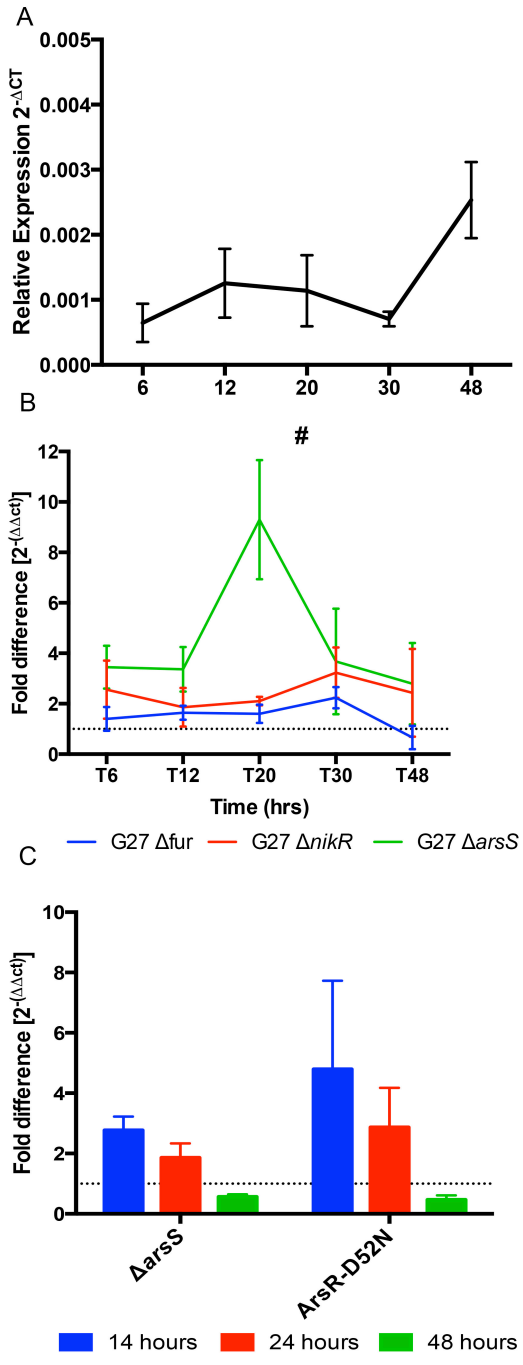
### *homB* Expression is Modulated by Growth Phase and Environmental Cues

There is currently little data regarding the expression of *homB*; however, immunologic data suggests that *homB* is expressed during infection (152). To explore *homB* expression, we performed qRT-PCR to quantify *homB* transcription levels over a standard 48 hour *in vitro* growth curve of *H. pylori* G27 (Figure 15A). We observed relatively low and consistent levels of expression over the first 30 hours of growth, followed by an approximately 3-fold increase in expression at 48 hours. This suggests that *homB* expression is up-regulated in stationary phase (Figure 15A). Next, we sought to determine what regulatory factor(s) were responsible for modulation of *homB* expression. To this end, we assessed *homB* expression in the absence of each of the following well characterized transcriptional regulators: Fur, NikR, and ArsS (Figure 15B). While the pattern of *homB* expression observed in the *Δfur* background was similar to the wild-type strain (Figure 15B), *homB* expression was increased in both the *ΔnikR* and *ΔarsS* strains throughout the growth curve. Specifically, in the *ΔnikR* strain background, we observed an average of a 2-fold increase in *homB* expression at all time-

**Figure 15. *homB* expression in wild-type and regulator-deficient mutant strains**

**A.** *homB* expression profile in wild-type *Helicobacter* strain G27. The relative expression of *homB* as compared to the 16S gene was calculated for each time point shown on the x-axis. The solid line is plotted at the mean of 3 biologically independent experiments; errors bars represent the standard error of the mean. **B.** *homB* expression was monitored over time from strains lacking Fur, NikR and ArsR. The fold difference between wild-type and each regulatory knockout is plotted over time. The solid lines are plotted at the mean of three biological replicates and errors bars show the standard error of the mean. # indicates significant p-values for T20 vs T6 ($p = 0.0086$), vs T12 ($p = 0.0075$), vs T30 ($p = 0.0295$), vs T48 ($p = 0.003$). **C.** *homB* expression in *ΔarsS* and Ars-D52N mutant strains was assessed at 14, 24, and 48 hour. Fold difference at each time point was calculated between the mutants and the wild-type. Each bar represents the mean of three biologically independent experiments and errors bars show the standard error of the mean.

Figure 15. ***homB*** **expression in wild-type and regulator-deficient mutant strains**



points as compared to wild-type. Similarly, the *ΔarsS* strain showed a dramatic difference in *homB* expression as compared to wild-type. For example, in the Δ*arsS* strain, transcript levels of *homB* were elevated between four- and ten-fold in comparison to wild-type over the course of the experiment. Furthermore, where wild-type *homB* expression peaked at 48 hours, the peak expression of *homB* was seen at 20 hours in the Δ*arsS* strain (Figure 15B). Together, these data suggest that the ArsRS two-component system and NikR may play a role in regulation of *homB* expression. Additionally, deletion of *arsS* results in a shift in maximal *homB* expression from the stationary phase to mid-log phase (Figure 15B).

### *homB* Expression is Modulated in Response to pH in an ArsRS Dependent Manner

Given the increase in *homB* expression observed in the *ΔnikR* and *ΔarsS* strain backgrounds, we next asked how *homB* expression changed in response to the environmental cues that are sensed by these regulators, nickel and pH, respectively. In wild-type *H. pylori*, we observed an approximately five-fold decrease in *homB* expression following a 90min shock with 10μM $Ni^{2+}$(Figure 16A). Unexpectedly, we observed a similar decrease in the *ΔnikR* strain background. These data suggest that the decrease in *homB* expression in response to excess nickel is NikR-independent. When wild-type *H. pylori* was subjected to a pH of 5 for 90 min, we saw a 10-fold decrease in *homB* expression as compared to pH 7 (Figure 16B). When the *ΔarsS* strain was similarly exposed to pH 5, we saw little to no decrease in *homB* expression. This finding suggests that *homB* repression at low pH is ArsS-dependent (Figure 16B).

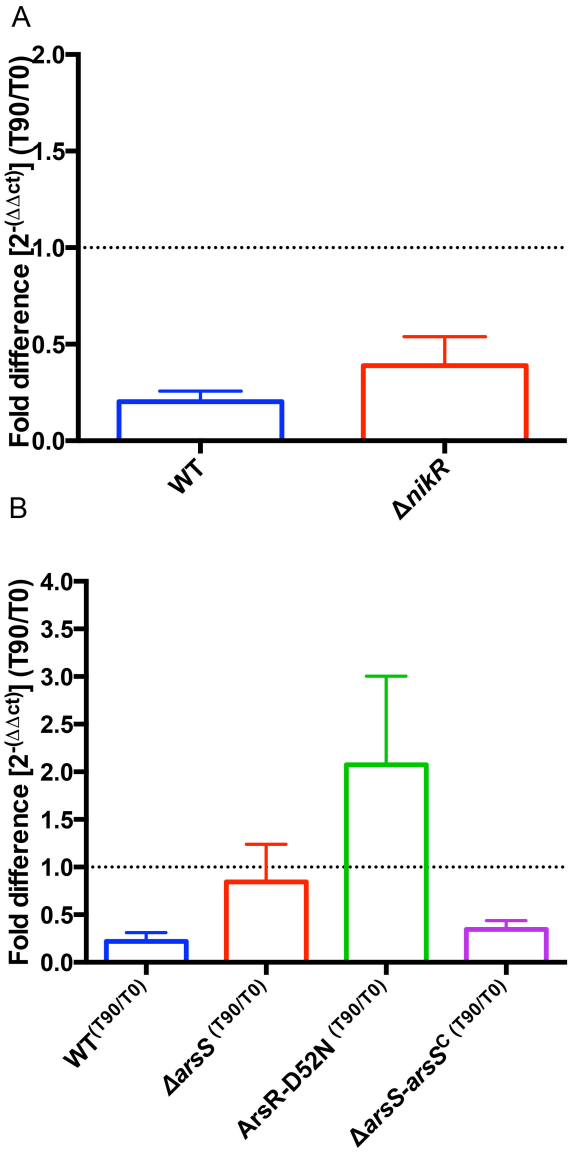



**Figure 16. *homB* expression in response to environmental stimuli**

**A.** The expression of *homB* was evaluated in wild-type and Δ*nikR* following a 90 minute exposure to 10$\mu$M $Ni^{2+}$. The fold difference between pre-shock ($T_0$) and post-shock ($T_{90}$) was calculated, and the mean fold difference of three biological replicates was plotted. The error bars show the standard error of the mean. **B.** The expression of *homB* was evaluated in response to a 90 min shock at pH 5 in wild-type, Δ*arsS*, ArsR-D52N, and Δ*arsR-arsS*$^{C}$. The fold difference between pre-shock ($T_0$) and post-shock ($T_{90}$) expression was calculated and plotted. Each bar represents the mean of a minimum of three biological replicates and error bars display the standard error of the mean.

As mentioned, ArsS is the sensor kinase portion of the two component system, ArsRS. In response to decreasing pH, ArsS phosphorylates ArsR and ArsR~P can then regulate gene expression. Ideally, this regulatory cascade would be assessed by knocking out *arsR*; however, this gene is essential and deletion of *arsS* is often used as a proxy. Another means of assessing this regulatory pathway is via the use of a non-phosphorylatable ArsR-D52N mutan strain; this strain is still viable but ArsR is unable to be phosphorylated in response to low pH. Given the dramatic effect that loss of ArsS had on *homB* expression, we next tested the same conditions using an ArsR-D52N mutant strain. In agreement with the increase in *homB* expression seen at early time points in the *ΔarsS* strain background, we also detected an increase in *homB* expression in the ArsR-D52N mutant strain during log phase (Figure 15C). Additionally, when ArsR-D52N was exposed to pH 5 shock, there was no decrease in *homB* expression (Figure 16B). These data further support the hypothesis that *homB* expression is repressed by ArsRS in response to low pH. Also of note, when *arsS* was expressed *in trans* in the *ΔarsS* strain background (*ΔarsS-arsS*$^{c}$), decreased *homB* expression in response to low pH was restored.

### *homB* Contributes to Hyper Biofilm Formation in ArsRS Mutant Strains

We recently showed that strains with a mutation in either portion of the ArsRS two-component system (*ΔarsS* or ArsR-D52N) form significantly more biofilm and do so at an enhanced rate as compared to wild-type *H. pylori* (182). Given that *homB* was highly over expressed in both the *ΔarsS* and ArsR-D52N strains, we reasoned that increased expression of this OMP could contribute to the previously observed biofilm

phenotype. Crystal violet assay quantification of biofilm formation, showed similar results as in our previous studies (182); significant biofilm formation occurred earlier in the *ΔarsS* and ArsR-D52N strains (24 hours) as compared to wild-type (72 hours) (Figure 17). Deletion of *homB* from the *ΔarsS* or ArsR-D52N strains resulted in a delay in significant biofilm formation (48 hours) (Figure 17). In addition, at all time points tested, *ΔarsS* and ArsR-D52N formed significantly more biofilm than wild-type, whereas *ΔarsS ΔhomB* and ArsR-D52N*ΔhomB* did not (Figure17, Table 5). Overall there was less biofilm formation in the *ΔhomB* strains as compared to their cognate *homB* positive strains; however, differences were not statistically significant (Figure17, Table 5). Despite the subtle differences observed in crystal violet quantification, we did note a dramatic decrease in crystal violet staining on the bottom of the well in all of the strains lacking *hom* (Figure 18).

**Deletion of *homB* Decreases Bacterial Adherence and Auto-Aggregation**

Given that deletion of *homB* decreased biofilm formation, we next assessed whether deletion of *homB* affected surface adherence and/or bacterial aggregation, each of which contribute to biofilm formation. Following a 2 hour incubation period, approximately 20% of the wild-type inoculum adhered to the surface. As observed previously (182), mutation of ArsRS resulted in an increase in surface adherence; 50% and 40% of the inoculum from *ΔarsS* and ArsR-D52N adhered, respectively (Figure 19). Although not statistically significant, comparison of wild-type to *ΔhomB* or ArsR-D52N to ArsR-D52N*ΔhomB*, revealed a decrease in adherence following *homB* deletion (Figure 19). These data suggest that *homB* contributes to surface adherence; however, this same

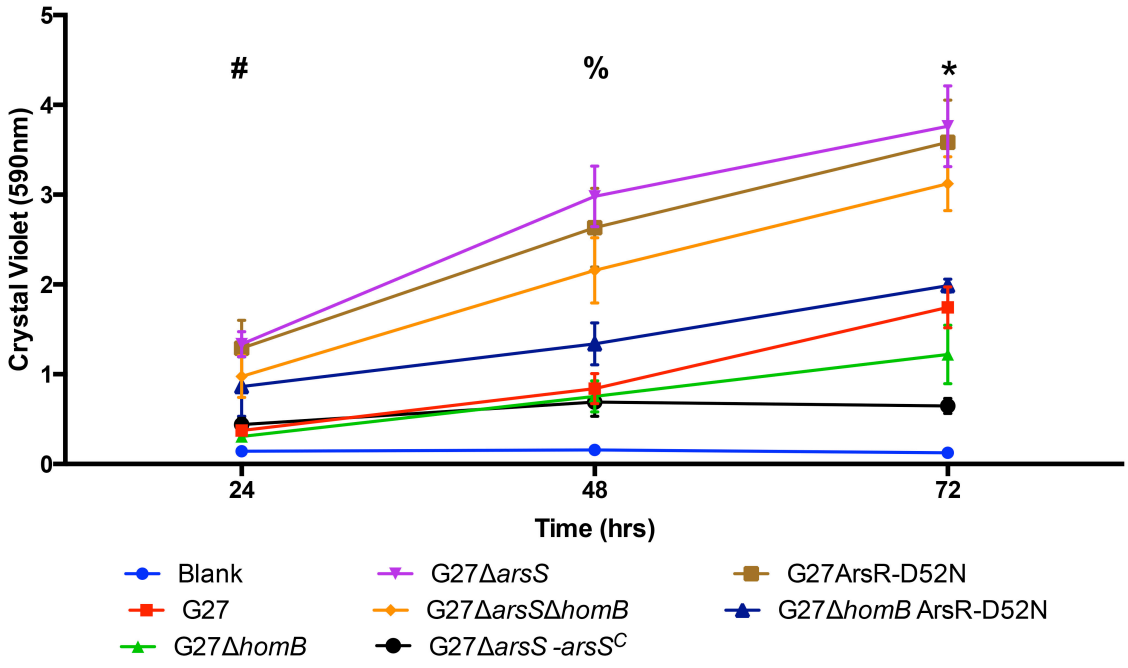



**Figure 17. Biofilm quantification**

Solublized crystal violet stain was read at an $OD_{590}$ and plotted over time. With the exception of *ΔarsS ΔhomB* the mean absorbance for a minimum of three biologically independent replicates is indicated by the solid line. For *ΔarsS ΔhomB*, the plotted mean represents 3 isolates tested in two biological replicates each. For all data sets, error bars show the standard error of the mean**.** #Significant biofilm formation was observed by 24 hours for *ΔarsS* verses blank ($p = 0.0037$) and ArsR-D52N verse blank ($p = 0.0211$); % Significant biofilm formation was observed by 48 hours for *ΔarsS ΔhomB* verse blank ($p < 0.0001$) and ArsR-D52N *ΔhomB* verses blank ($p = 0.0164$); * Significant biofilm formation was observed by 72 hours for G27 wild-type verses blank ($p < 0.0001$) and *ΔhomB* verses blank ($p = 0.0309$)

**Table 5. Statistical analysis of biofilm quantification**

| Comparison | P-value | Summary |
|---|---|---|
| 24 hours | | |
| G27 vs. G27Δ*hom* | > 0.9999 | ns |
| G27 vs. G27Δ*arsS* | 0.0006 | *** |
| G27 vs. G27ArsR-D52N | 0.0066 | ** |
| G27 vs. G27Δ*arsS*Δ*hom* | 0.2555 | ns |
| G27 vs. G27Δ*hom* ArsR-D52N | 0.3957 | ns |
| G27Δ*arsS* vs. G27*ΔarsS*Δ*hom* | 0.862 | ns |
| G27ArsR-D52N vs. G27*Δhom* ArsR-D52N | 0.7294 | ns |
| G27Δ*hom* vs. G27Δ*arsS*Δ*hom* | 0.2518 | ns |
| G27Δ*hom* vs. G27Δ*hom* ArsR-D52N | 0.3864 | ns |
| | | |
| 48 hours | | |
| G27 vs. G27Δ*hom* | > 0.9999 | ns |
| G27 vs. G27Δ*arsS* | < 0.0001 | **** |
| G27 vs. G27ArsR-D52N | 0.0012 | ** |
| G27 vs. G27Δ*arsS*Δ*hom* | 0.0556 | ns |
| G27 vs. G27Δ*hom* ArsR-D52N | 0.9184 | ns |
| G27Δ*arsS* vs. G27Δ*arsS*Δ*hom* | 0.5208 | ns |
| G27ArsR-D52N vs. G27Δ*hom* ArsR-D52N | 0.0736 | ns |
| G27Δ*hom* vs. G27Δ*arsS*Δ*hom* | 0.0721 | ns |
| G27Δ*hom* vs. G27Δ*hom* ArsR-D52N | 0.8974 | ns |
| | | |
| 72 hours | | |
| G27 vs. G27Δ*hom* | 0.9654 | ns |
| G27 vs. G27Δ*arsS*Δ*hom* | 0.1413 | ns |
| G27 vs. G27Δ*hom* ArsR-D52N | > 0.9999 | ns |
| G27 vs. G27Δ*arsS* | 0.0004 | *** |
| G27 vs. G27ArsR-D52N | 0.0074 | ** |
| G27Δ*arsS* vs. G27Δ*arsS*Δ*hom* | 0.9344 | ns |
| G27ArsR-D52N vs. G27Δ*hom* ArsR-D52N | 0.0659 | ns |
| G27Δ*hom* vs. G27Δ*arsS*Δ*hom* | 0.0318 | * |
| G27Δ*hom* vs. G27Δ*hom* ArsR-D52N | 0.8443 | ns |

ns: p>0.05; *: p≤0.05; **: p≤ 0.01; ***:p ≤0.001; ****: p≤ 0.0001

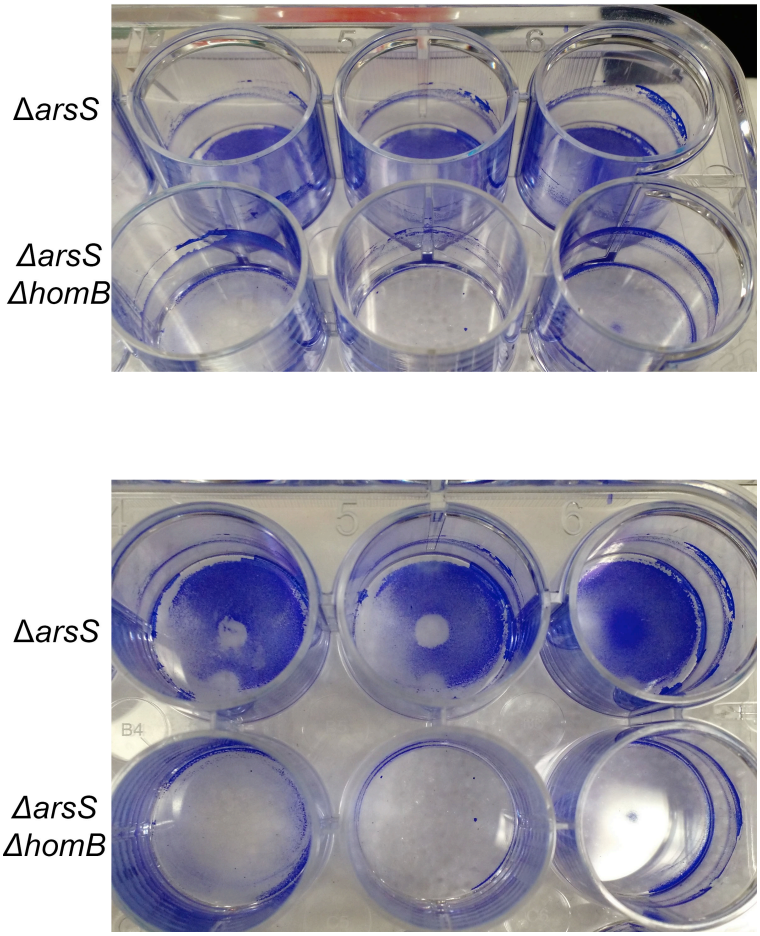



**Figure 18 (S2). Crystal violet staining of biofilm formation**
Following 72 hours of growth, *H. pylori* biofilm for Δ*arsS* and Δ*arsS* Δ*homB* were visualized by crystal violet staining. Each strain is shown in triplicate across the row. The top panel shows formation of biofilm at the liquid-air interface; the bottom panel displays biofilm formation adherent to the bottom of the well.

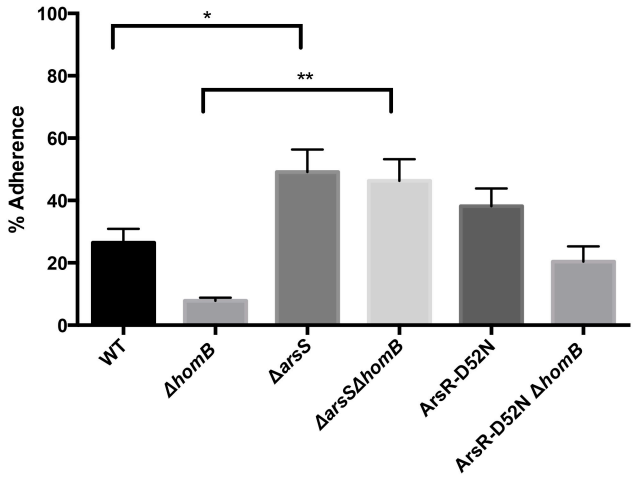



**Figure 19. Bacterial adherence**

Surface adherence was approximated using alamarBlue following a 2 hours incubation. Horizontal bars indicate the mean of the results from three biological replicates, and error bars show the standard error of the mean. * p =0.0226; **p=0.0041

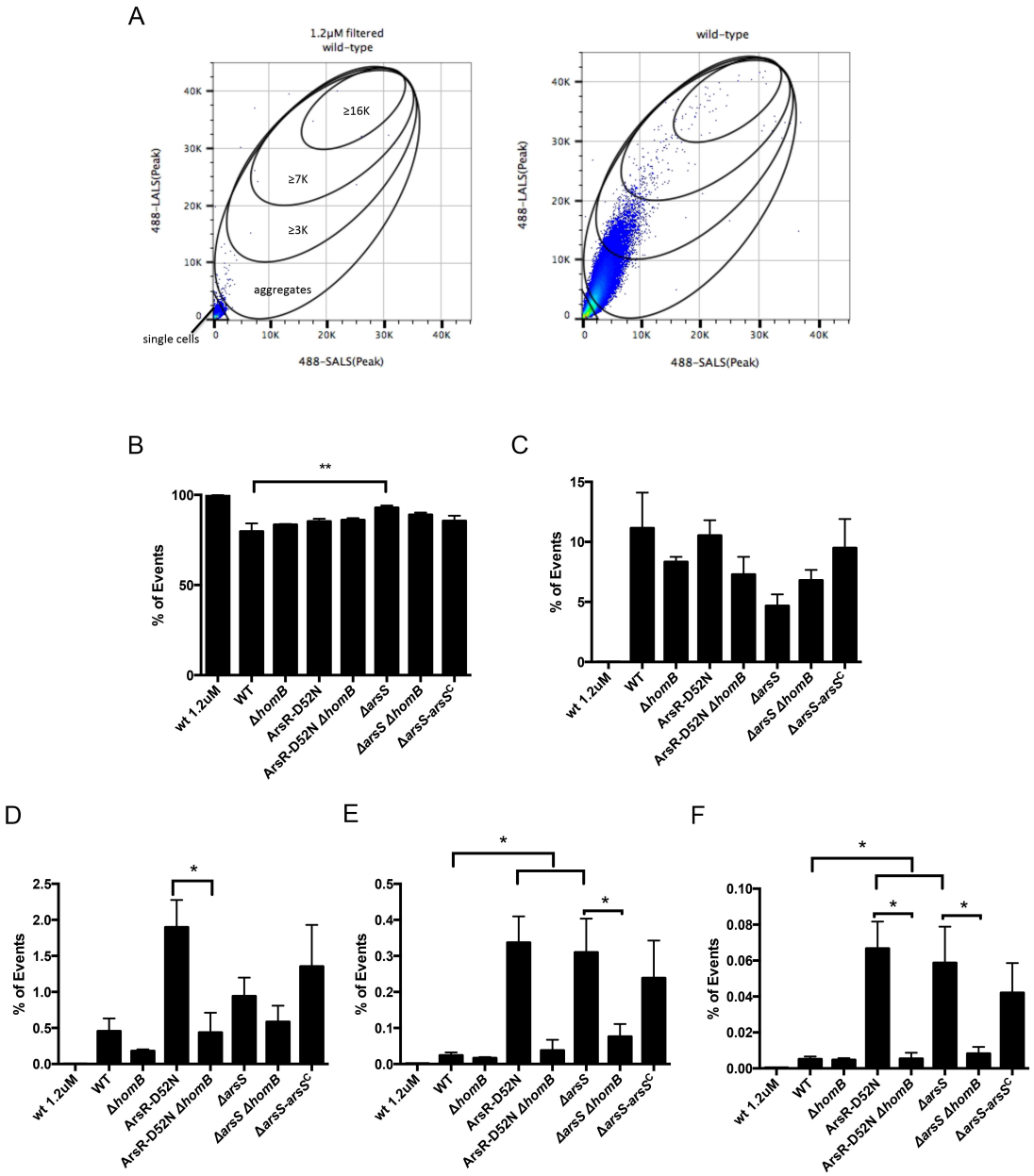
trend was not observed when *ΔarsS* was compared to *ΔarsSΔhomB* (Figure 19). However, a statistically significant increase in adherence between *ΔarsS* and wild-type (p =0.0226) and *ΔarsS ΔhomB* and *ΔhomB* (p =0.0041) was observed. These findings suggest that there may be some phenotypic differences between the *ΔarsS* and ArsR-D52N mutations, and perhaps deletion of ArsS leads to increased adherence in a *homB* independent manner.

We previously reported an increase in bacterial aggregation in *ΔarsS* and ArsR-D52N mutant strains(182). Similarly, we observed large bacterial aggregates in the *ΔarsS* and ArsR-D52N mutant strains in the current study (Figure 14). In order to quantify these observed differences in bacterial aggregation, we used flow cytometry. Filtration of *H. pylori* through a 1.2$\mu$M filter creates a population of almost exclusively single-celled bacteria (Figure 14). Thus, to compare between strains, we drew one gate to encompass a majority of the filtered wild-type population and then drew each subsequent gate to correspond to events with increasing SALS (Figure 20, 21); the SALS value increases with the size of the event, and increasing event size corresponds to aggregates of increasing size. As expected, nearly 100% of bacteria from filtered wild-type samples fell within the smallest gate (range 99.3%-99.9%) (Figure 20 A and B). In comparison, 74.8%-91.8% of unfiltered wild-type bacteria fell within the single cell gate (Figure 20 A and B). Similar percentages to wild-type were observed for *ΔhomB* (82.8%-84.3%), ArsR-D52N (83.7%-88.3%), ArsR-D52N *ΔhomB* (84.2%-88.1%)), *ΔarsS ΔhomB* (85.6%-92.7%), and *ΔarsS-arsS*$^{C}$ (81.8%-91.3%) (Figure 20 B). However, we observed a significant increase in the number of single cell bacteria in the *ΔarsS* strain as compared to wild-type (p =0.0021). To assess the aggregate population, samples were

**Figure 20. Bacterial aggregation**

Bacterial aggregation was assessed by flow cytometry. Gates were drawn to identify the singlet population of bacteria as well as populations with increasing sizes of aggregates. **A.** Representative scatter plots demonstrating the gate placements are given for filtered wild-type and unfiltered wild-type bacteria. SALS is plotted on the x-axis, LALS is plotted along the y-axis. The gate names correspond to the SALS values and are indicated on the filtered wild-type scatter plot **B.** single cell *H. pylori* population; ** $p<0.01$. **C.** total bacterial aggregates (event ≥2.5K SALS); * $p< 0.05$. **D.** bacterial aggregates ≥ 3K SALS; *$p< 0.05$ **(E)** bacterial aggregates ≥7K SALS; *$p<0.05$. **F.** bacterial aggregates ≥16K SALS; *$p<0.05$.

**Figure 20. Bacterial aggregation**



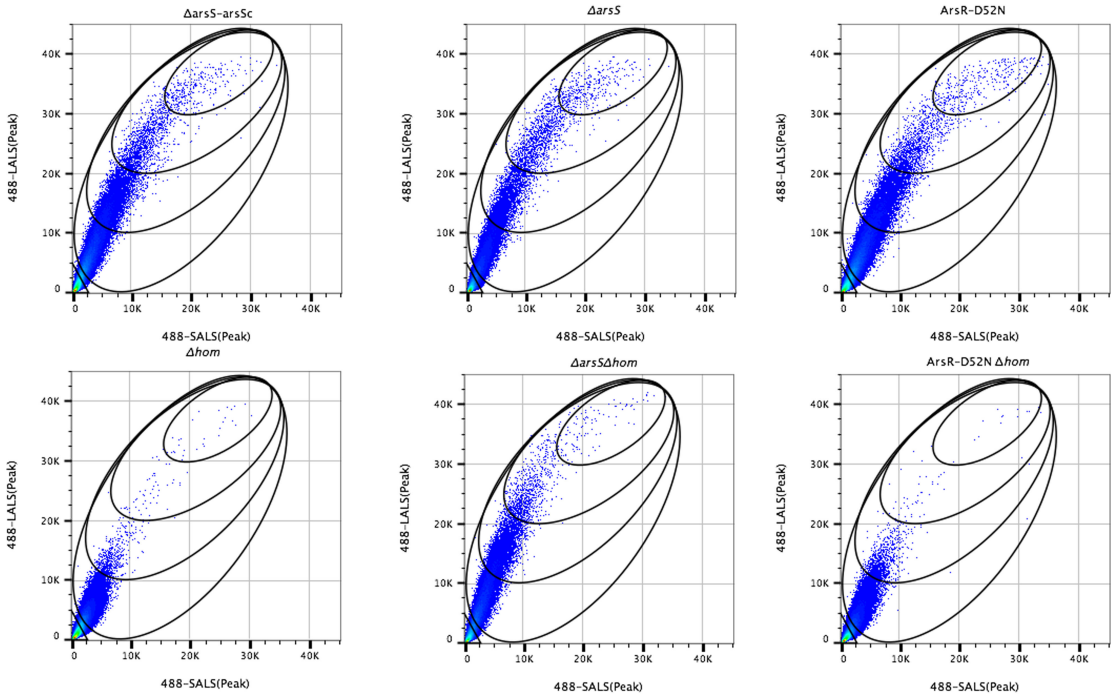



**Figure 21 (S3). Representative scatter plots and gates**
Representative scatter plots are shown for all strains except filtered and unfiltered wild-type (Figure 20). The SALS is plotted on the x-axis and LALS is on the y-axis. Oval gates were drawn to encompass different size aggregates; the smallest gate corresponds to only the largest aggregates.

gated as previously described (188). We observe that 10% of the unfiltered wild-type events fell within the largest gate, which includes all events larger than single cells (Figure 20 B). Overall, a similar number of events were included in the total aggregate gate among all other strains (Figure 20 B); however, there appeared to be a slight reduction in this population for the *ΔarsS* strain ($p=0.0539$ compared to wild-type). This reduction is likely reflective of the increase in single cell bacteria that were observed for this strain. Of note, the ArsR-D52N mutant, which is genetically more similar to *ΔarsS* was phenotypically more similar to wild-type in terms of total aggregation (Figure 20 B). When the effect of the *homB* deletion was considered, there was a slight, but non-significant, decrease in the aggregate *ΔhomB* population as compared to wild-type (Figure 20 C). The same was true for ArsR-D52N *ΔhomB as* compared to ArsR-D52N (Figure 20 C). In contrast, the percent of *ΔarsS ΔhomB* events in this large gate was slightly increased as compared to *ΔarsS*, but was similar to *ΔhomB* and ArsR-D52N *ΔhomB* (Figure 20 C). Subsequent analysis examined the number of events present in gates that encompass smaller areas that correspond to the highest SALS and LALS values; these gates capture the larger aggregates that were observed by microscopy (Figure 14). For unfiltered wild-type, very few events were observed at the high values (<0.5% were greater than SALS 3K) (Figure 20 D,E,F). However, there was a significant increase in large events for both *ΔarsS* ($p\leq0.03$) and ArsR-D52N ($p\leq0.03$) as compared to wild-type for gates encircling aggregates >7K SALS. Within these same gates (>3K SALS) we also observed a decrease in events for strains lacking *homB* as compared to their cognate *homB* positive strain (Figure 20 D,E,F). For aggregates that fall within gate >3K SALS there was a significant difference between ArsR-D52N and ArsR-D52N

*ΔhomB* (p=0.0418); a similar trend was observed for the gate >7K SALS (p =0.0674) and >16K SALS (p =0.0438). A significant decrease was observed between *ΔarsS* and *ΔarsS ΔhomB* for gates >7K SALS (0.0479) and >16K SALS (0.0202). Intriguingly, while the *ΔarsS-arsS*$^{C}$ phenotypically resembled wild-type in terms of the single cell population, this strain still produced more large aggregates (>7K SALS) than wild-type and was more similar to the *ΔarsS* strain (Figure 20 B and C as compared to Figure 20 D,E,F). Overall, loss of *homB* was associated with a loss of large aggregate formation in the *ΔarsS* and ArsR-D52N mutant strains.

### *homB* Over Expression Enhances Biofilm Formation

Through the course of this study, we found that *homB* was regulated by the ArsRS two component system and that deletion of *homB* led to a reduction is biofilm formation. This finding was particularly true in the ArsR-D52N and *ΔarsS* backgrounds. Given this observation, we next asked if increased *homB* expression in a wild-type strain was sufficient to induce the enhanced biofilm formation that was observed in the ArsR-D52N and *ΔarsS* mutant strains. To accomplish this goal, constructs where *homB* expression was driven by either the *ureA* or *flaA* promoters were generated. Increased expression of *homB* was confirmed by qRT-PCR (Figure 22). As expected, in comparison to the wild-type background, the *flaA* promoter led to a modest increase in *homB* expression (G27 pFlaA-*homB* Figure 22), whereas the *ureA* promoter (G27 pUreA-*homB* figure 22) led to a large increase in *homB* expression of approximately 100 fold (Figure 22). In the *ΔhomB* strains transformed with these constructs, we observed between a two- and four- fold

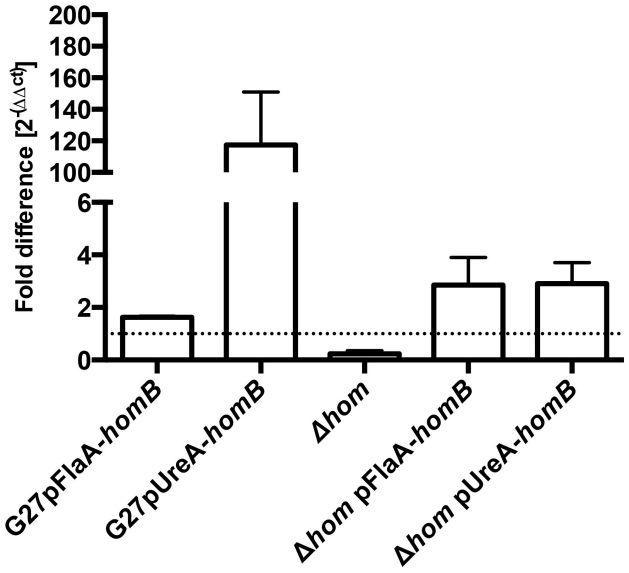



**Figure 22. Evaluation of *homB* expression driven by pFlaA and pUreA**
*homB* expression was assessed by qRT-PCR. For each strain, the fold difference was calculated in relation to the expression of *homB* in wild-type G27 carrying the empty vector. The mean *homB* expression for each strain is indicated by the bar; error bars show the standard error of the mean.

increase in *homB* expression regardless of which promoter was driving expression (Figure 22). Testing of this panel of strains for biofilm formation revealed that the same increased rate of biofilm formation that was observed in the ArsR-D52N or *ΔarsS* strains was not observed (Figure 23). However, there was a trend towards increased biofilm formation between *ΔhomB* and both over expressing strains at 48 and 72 hours, and between G27pFlaA-*homB* and wild-type at 72 hours (Figure 23). Of note, G27pFlaA-*homB*, Δ*homB* pFlaA-*homB*, and Δ*homB* pUreA-*homB* all formed comparable levels of biofilm to *ΔarsS* at 72 hours (Figure 23). These data suggest that over expression of *homB* can enhance the overall amount of biofilm formed; however, this alone it is not sufficient to recapitulate the enhanced biofilm formation observed when the ArsRS system is mutated.

**Biofilm SEMs Reveal Dramatic *homB*-Dependent Differences in Biofilm Formation**

Given the variability observed in the crystal violet straining, adherence, and aggregation assays, we next uses SEM imaging as a qualitative and semi-quantitative approach to directly visual biofilm formation of the isogenic mutant strains (Figure 24 A and B). When wild-type *H. pylori* were viewed at the lower magnification, we observed slight biofilm formation and numerous adherent bacteria (Figure 24A). In comparison, less biofilm and adherent bacteria were observed for the *ΔhomB* strain (Figure 24A). As expected, striking biofilm formation was observed for both of the ArsRS mutant strains. Deletion of *homB* in both strains drastically reduced biofilm formation (Figure 24A). Furthermore, a decrease in biofilm formation was also observed for *ΔarsS-arsS*$^{C}$ (figure A). SEM analysis also revealed a dramatic difference between the empty vector control

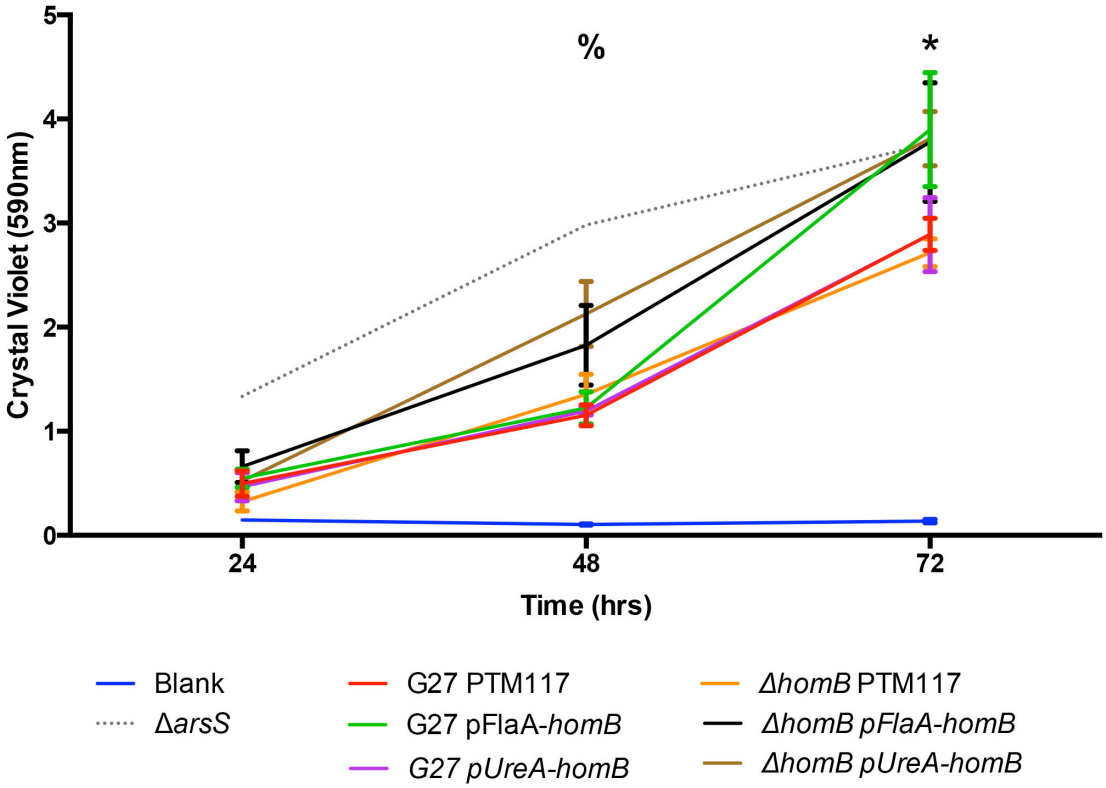



**Figure 23. Biofilm quantification of pHomB strains**

Solublized crystal violet stain was read at an $OD_{590}$ and plotted over time. The mean absorbance of three biologically independent replicates is indicated by the solid lines and error bars show the standard error of the mean. As a reference, the dotted line indicates the mean absorbance value for *ΔarsS* from Figure 17. % Significant biofilm formation was observed by 48 hours for *ΔhomB*PTM117 vs blank ($p = 0.0251$), *ΔhomB*pFlaA-*homB* vs blank ($p = 0.0008$), and *ΔhomB*pFlaA-*homB* vs blank ($p < 0.0001$); Significant biofilm formation was observed by 72 hours for G27 PTM117 vs blank ($p < 0.0001$), G27pFlaA-*homB* vs blank ($p < 0.0001$), and G27pFlaA-*homB* vs blank ($p < 0.0001$)

**Figure 24. SEMs of biofilm formation after 48 hours demonstrate a role of HomB in biofilm formation.**

**A.** Representative SEM images from all 13 isogenic mutant strains. Biofilm-like mats with 3D architecture can be seen for ArsR-D52N, Δ*arsS*, G27 pFlaA-*homB*, G27 pUreA-*homB*, Δ*homB* pFlaA-*homB*, and Δ*homB* pUreA-*homB*. **B.** Biofilms viewed at a closer magnification reveal numerous structures including: flagella (solid red triangle), coccoid *H. pylori* (open red triangle), OMVs (red arrow), and extracellular matrix (red star).

**Figure 24. SEMs of biofilm formation after 48 hours demonstrate a role of HomB in biofilm formation.**

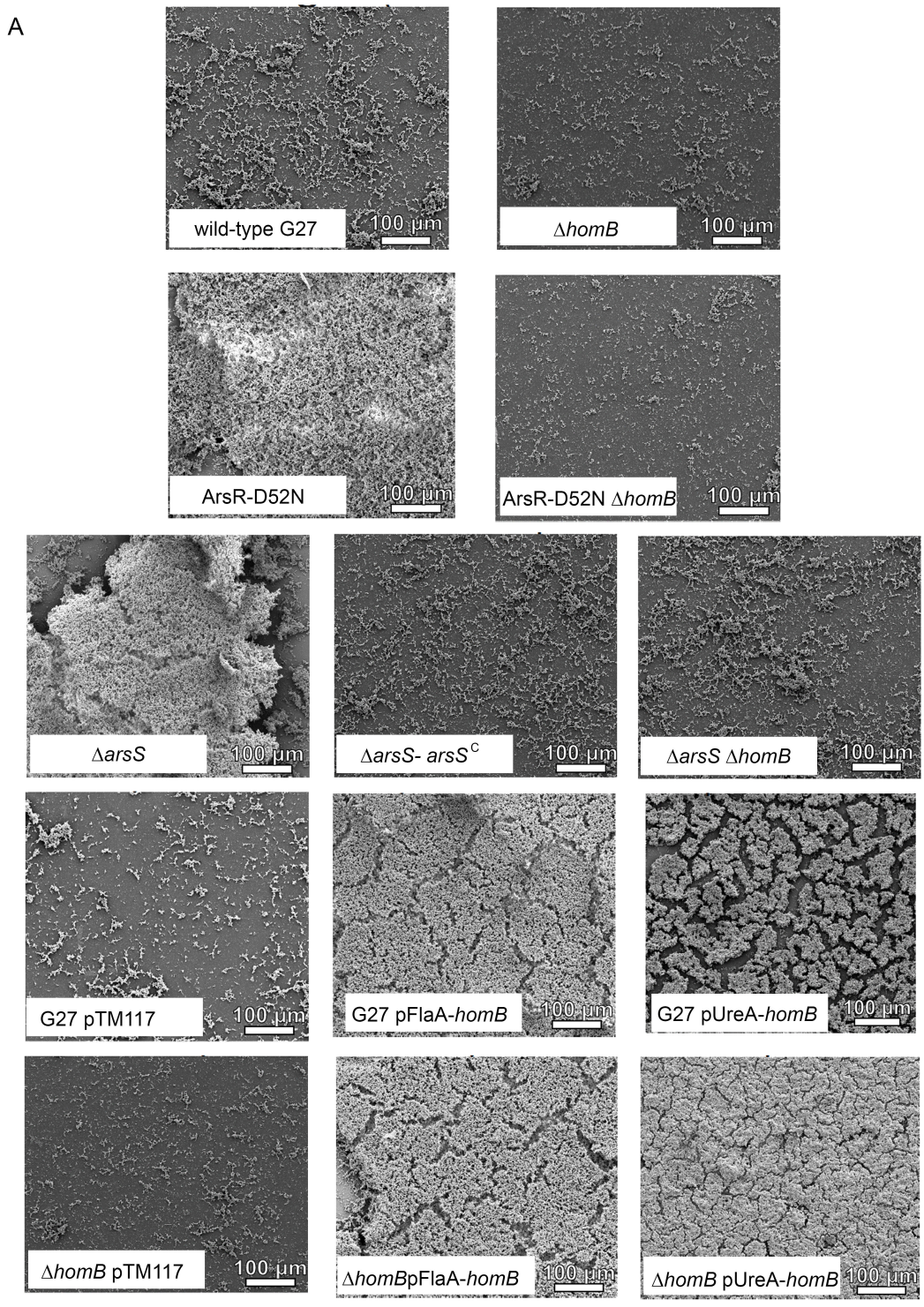
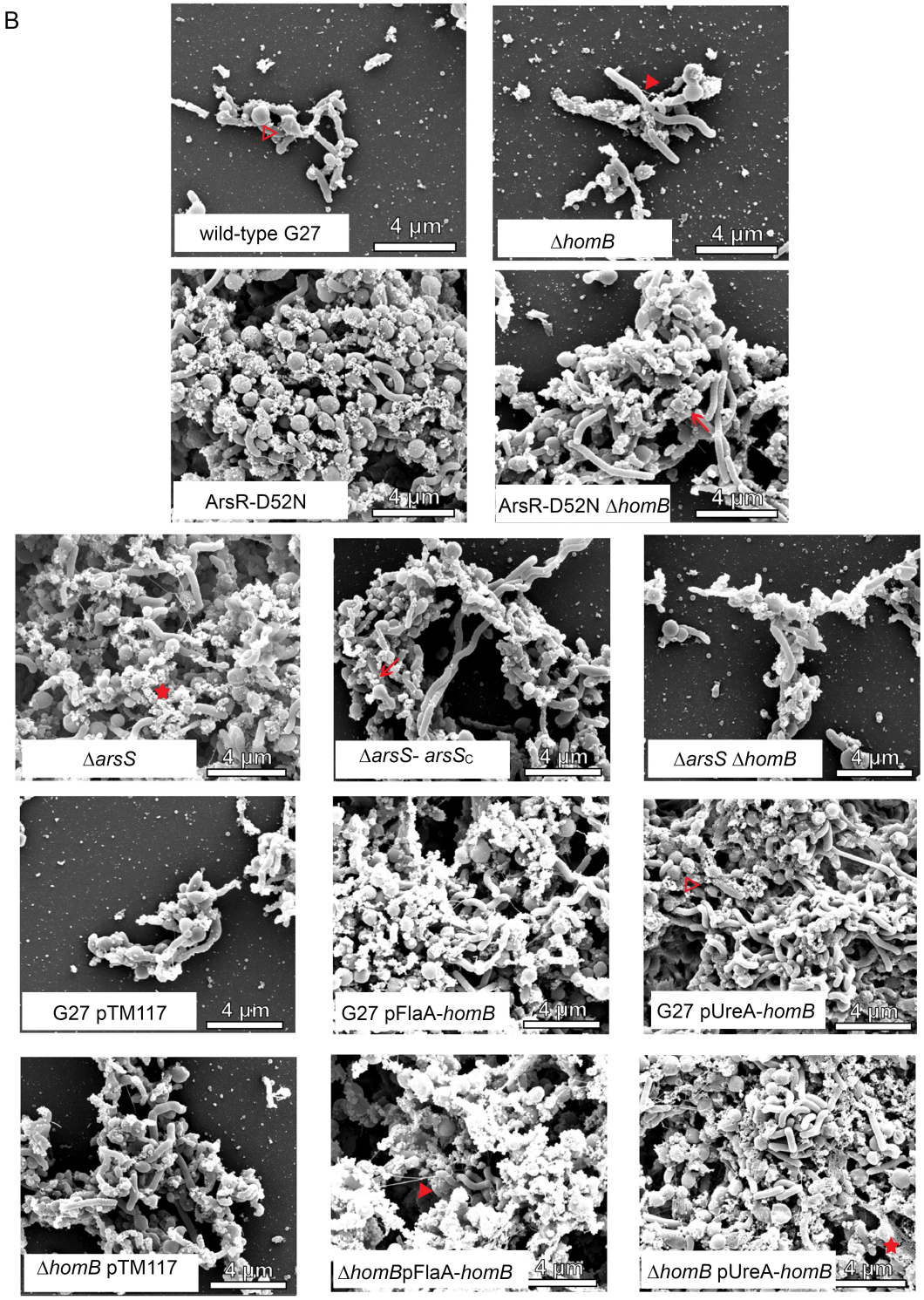


**Figure 24. SEMs of biofilm formation after 48 hours demonstrate a role of HomB in biofilm formation.**



strains (G27 pTM117 and *ΔhomB* pTM117) as compared to the *homB* over expression strains (Figure 24A). Introduction of either vector that expressed *homB* under the control of the *flaA* or *ureA* promoter, lead to biofilm formation similar to what was seen in ArsRS mutant strains (Figure 24A). There did appear to be slight architectural differences between the ArsRS mutant strain biofilms and the biofilms formed when *homB* was expressed *in trans* (Figure 24A). Biofilm matrix, flagella, coccoid *H. pylori*, and potentially OMVs were visualized upon increased magnification of the biofilms (Figure 24B).

## DISCUSSION

*H. pylori* persistently colonizes approximately one half of the world's population. In order to persist within the gastric niche, *H. pylori* must adapt to its host and to any changes that occur within that host throughout the course of infection. Given their location, components that make up the cellular envelope, including OMPs, provide key host-pathogen interactions. Comparative studies of sequenced *H. pylori* genomes have identified at least 60 OMPs. Many of these OMPs are among the auxiliary genes that are present in only a subsets of strains. Furthermore, many OMPs are differenitally expressed in clinical isolates (150), and OMP variations are associated with differential disease outcomes (55).

HomB, is one of the auxiliary OMPs that is associated with severe disease. Initial investigations of HomB found that *homB* could be present at two possible loci, and that strains could harbor zero, one, or two copies of *homB* (108; 110; 152-154). Based on sera from *H. pylori* positive pateints, HomB is likely expressed *in vivo* and may contribute to

adherence (152). While those earlier studies provided a brief glance into the molecular characteristics of HomB, more detailed studies have yet to be conducted. Herein, we endeavored to expand our knowledge of HomB by evaluation of the expression patterns of this OMP, which may give clues about its possible role during *H. pylori* infection.

Since, *homB* can reside in two possible loci; the first step was to identify any differences between the promoter regions of the two loci. Interestingly, the promoter regions at both of these loci are nearly identical (Figure 13 A and B). Unlike many other OMPs, examination of the promoter and coding regions of *homB* revealed no mono/dinucleotide repeats that are typically indicative of slip-strand mispairing and *homB* phase variation (salaun 2004) (Figure 13 A and B, data not shown). The high level of conservation of the *homB* promoter and surrounding intergenic DNA is particularly intriguing given that nucleotide diversity is reported even within conserved coding regions (194). One reason intergenic DNA may be persevered is to maintain the integrity of protein binding sequences that are important for gene regulation. Given that the promoters at both loci were indistinguishable, we chose to begin the evaluation of *homB* expression using G27, which carries a single *homB* gene at locus A.

When *homB* expression was evaluated across the growth curve, relatively low and steady *homB* expression was observed through 30 hours. However, as wild-type cultures approached late stationary phase (48 hours) approximately a 3-fold increase in *homB* expression was observed compared to exponential phase (Figure 15A). An increase in gene expression during the log to stationary transition period is not specific to *homB* and this transition in gene expression has been coined as the "log-stat switch" (201). Interestingly, numerous virulence factors are among the genes shown to increase during

the growth phase transition, suggesting that perhaps *homB* is up-regulated as part of *H. pylori* 's virulence arsenal (201).

To determine which regulatory proteins may contribute to modulation of *homB* expression, we utilized strains carrying single deletions of Fur, NikR and ArsS, which are three hubs of *H. pylori* regulation (network paper). Deletion of Fur had little effect on *homB* expression (Figure 15B). In contrast, deletion of either NikR or ArsS led to an increase in *homB* expression. The most striking differences were observed in the *ΔarsS* strain; there was an increase in *homB* expression at all examined time points, with the largest increase observed at 20 hours (Figure 15B). Given that ArsS is known to affect transcription in response to low pH, we investigated this regulatory pathway further to determine if *homB* expression was pH responsive. Indeed, *homB* was suppressed almost 10-fold following shock at pH 5. However, when ArsS was deleted or the phospho-accepting residue of ArsR was mutated, both of which abrogate the acid responsiveness of the ArsRS system, little difference in *homB* expression was observed in response to pH shock. Futhermore, when *arsS* was complemented *in trans*, the acid-responsive decrease in *homB* expression was restored. Together, these data suggest that the ArsRS two component system represses *homB* transcription in response to decreasing pH.

We further examined the increase in *homB* expression following deletion of NikR, and found a 2-fold increase in *homB* expression as compared to wild-type at all time-points tested. Exposure of *H. pylori* to excess nickel, which mimics the conditions under which NikR binds DNA, led to a nickel responsive decrease in *homB* in both the wild-type and the Δ*nikR* strains. These data suggest that *homB* is repressed under conditions of excess nickel but that NikR does not directly mediate this change. One

possible explanation for the decrease in *homB* expression in response to nickel may be the coordinated regulation of expression of *arsR* by Fur and NikR that was recently reported (170). That study found that in the presence of iron, Fur can bind to the promoter and repress *arsR* expression; however, in excess nickel conditions, NikR interferes with Fur repression of ArsR, leading to increased ArsR expression (170). Thus, it is plausible that the increase in *homB* expression observed in the *nikR* deletion strains is due to the role that NikR plays in ArsR regulation. The effect of this interaction on *homB* expression remains to be investigated.

It is also of note that mutations in the ArsRS system led to changes in the pattern of *homB* expression. For example, in the wild-type strain, an increase in *homB* expression was observed during the transition from late-log to stationary phase, whereas *homB* expression in both of the ArsRS mutant strains increased during log phase. While the timing is different, there was still a significant increase in *homB* expression at one time-point (Figure 15B). Together, these data suggest that while ArsRS is responsible for suppression of *homB* expression, another regulatory factor may be involved in the increase in *homB* expression that is observed at T48 and T20 in the wild-type and *ΔarsS* strains, respectively.

We recently showed that deletion of ArsS or mutation of ArsR leads to an increase in biofilm formation by *H. pylori* (182). Given the location of HomB in *H. pylori*'s outer membrane, and the increase in *homB* expression observed following ArsRS mutation, we speculated that HomB may contributed to biofilm formation. Indeed, deletion of *homB* in combination with either *ΔarsS* or ArsR-D52N results in a delay in significant biofilm formation as compared to *ΔarsS* or ArsR-D52N (Figure 17). While

not statistically significant, we also noted a decrease in the amount of biofilm formed at 72 hours by the *ΔhomB* strain as compared to wild-type (Figure 17). In addition, while the crystal violet quantification showed a trend towards decreased biofilm formation, a striking difference in crystal violet staining at the bottom of the well was evident (Figure 18). Thus, data from the crystal violet staining, adherence, and aggregation assays clearly suggest that *homB* contributes to biofilm formation in *H. pylori* G27. Moreover, SEM analyses demonstrate that *homB* is not only necessary but also sufficient for biofilm formation by this strain (Figure 24A and B).

Interestingly, a phenotypic difference between the *ΔarsS ΔhomB* and ArsR-D52N *ΔhomB* strains became increasingly more apparent during these assays. Slight variations between *ΔarsS* and ArsR-D52N were previously observed (182); however, these differences are more apparent when combined with the *ΔhomB* mutation. Most notably, while *ΔhomB* and ArsR-D52N *ΔhomB* behave more similarly in the aggregation and adherence assays in comparison to wild-type and ArsR-D52N, respectively (Figures 19, 20), in both of those assays *ΔarsS* and/or *ΔarsS ΔhomB* show unique phenotypes. Specifically, in the adherence assay, deletion of *arsS* in the wild-type or *ΔhomB* background leads to a significant increase in adherence in comparison to wild-type or *ΔhomB* respectively (Figure 19). In contrast, the ArsR-D52N mutation leads to slight, although non-significant, increase in adherence as compared to wild-type (Figure 19). Furthermore, both the wild-type and ArsR-D52N strains adhere to a greater extent than *ΔhomB* and ArsR-D52N Δ*homB*, respectively (Figure 19).

When considering what may lead to these differences, it is important to remember that the D52N mutation prevents all phosphorylation of ArsR. Conversely, deletion of

ArsS eliminates the cognate kinase, but leaves the phospho-accepting residue of ArsR intact. Although this distinction may seem slight, phosphorylation of recombinant ArsR protein has been observed in *E. coli* that does not express ArsS (208), and the ability of ArsR to be phosphorylated by other kinases, such as FlgS, has been proposed (121; 208). Additionally, while the initial ArsR regulon was proposed to consist of 2 target groups (1. essential ArsR phosphorylation independent targets and 2. non-essential ArsR phosphorylation dependent targets), the phosphorylation dependent targets have since been subdivided (208). The three sub-categories of the non-essential ArsR regulon now include: (I) ArsS-dependent regulation responding to periplasmic acidification, (II) ArsS-independent regulation responding to cytoplasmic acidification (pH <6.5), and (III) ArsS-independent regulation responding to cytoplasmic acidification (pH <5.3). Given the complexity of the ArsRS regulon, it seems likely that there may be slight differences between the ArsR-D52N and *arsS* mutations. Perhaps these differences are partially masked by *homB* over expression in these strains and become more apparent when *homB* is lost.

Another interesting observation that was specific to the *ΔarsS* strain was the increase in single cell *H. pylori* as compared to the wild-type strain (Figure 20 B). This increase lead to a slightly lower percentage of aggregates in the *ΔarsS* strain as compared to the wild-type and *ΔarsS ΔhomB* strains (Figure 20 C). Intriguingly, despite the increase in single bacteria, there was still an elevated percentage of large aggregates (>3K) in the *ΔarsS* population as compared to both wild-type and *ΔarsS ΔhomB* (Figure 20 D, E, F). It is possible that this difference is the result of the biofilm development cycle. Our previous study (182) showed a slight increase in biofilm formation, as well as

differential biofilm architecture, between *ΔarsS* and ArsR-D52N mutant strain. Following formation of a mature biofilm, cells are dispersed to seed new environments. In fact, the *H. pylori* quorum sensing molecule AI-2 of *H. pylori* has been shown to induce cellular dispersal (12). Although aggregation was assessed at 20 hours, a low level of biofilm is observed at this point in the *ΔarsS* strain. Thus, it is possible that the increase in single cell bacteria in this strain is a result of early cell dispersal from the biofilm. The dispersal phenomenon offers an explanation for simultaneous phenotypes of large aggregates and increased single cells.

One recurrent problem faced throughout the course of this study was the variation seen between biological replicates. The nature of the ArsRS mutations that results in biofilm formation and bacterial aggregation may influence the methods used to standardize our starting population for each assay. Additionally, with the exception of the biofilm assays, the studies presented here are only a snapshot of a single time point. Biofilm development is a dynamic process and perhaps observance of adherence and aggregation across a time course would better demonstrate phenotypic differences.

The *ΔarsS-arsS*$^{C}$ strain also displayed an unexpected phenotype in the aggregation assay. While complementation restored numbers of single bacteria to wild-type levels, (Figures 20 B and C), there was still an increase in larger aggregate; this was comparable to the *ΔarsS* strain (Figures 20 E and F). Intriguingly, the *ΔarsS-arsS*$^{C}$ strain does not form a biofilm; biofilm production by this strain is reduced even in comparison to wild-type (Figure 17, Figure 24). Thus, it is possible that complementation of *arsS in trans* does not recapitulate native *arsS* expression. This is turn leads to the variations in the

conditions tested. To this end, we did observe that expression of *arsS* episomally was at least 10-fold higher than in the wild-type strain (data not shown).

The ArsRS mutations are clearly pleotropic, which may make it difficult to tease out the effect of a single gene. In an attempt to directly assess if increased *homB* expression was sufficient to accelerate biofilm formation, we constructed vectors that would drive *homB* expression from either the *flaA* or *ureA* promoter. As expected, we saw an incredible increase in *homB* expression from the *ureA* promoter; however, this increase did not lead to significantly enhanced biofilm formation. Thus, it is possible that the 100-fold increase in OMP gene expression overwhelms the cell and does not result in a functional increase in HomB expression. A major limitation of these studies is the lack of specific-antibodies to accurately assess HomB protein expression. Therefore, it is still unclear to what extent *homB* contributes to biofilm formation. Given that expression of *homB* from either promoter in *ΔhomB* resulted in an observable increase in biofilm formation at both 48 and 72 hours, future investigation into the role of *homB* in biofilm formation is warranted (Figure 23 and Figure 24).

*En masse*, these data support a role for *homB* in biofilm formation by *H. pylori*. We previously proposed a model in which *H. pylori* used pH as an indicator of proximity to its colonization niche at the gastric epithelium. In that model, ArsRS responds to the increased pH at the gastric epithelium and modulates gene expression to promote colonization and biofilm formation. Herein, we show that *homB* is one of the factors regulated by ArsRS. The findings presented in this chapter suggest that *homB* contributes to biofilm formation; however, future work that directly assesses protein expression directly will help to better define its role. Additionally, this study has opened future

investigations into the role that other OMPs may play in biofilm development, as well as into the consequences that variable OMP genotypes may have on biofilm formation.

## ACKNOWLEDGEMENTS

We thank Drs. Ann Jerse and Isabelle Leduc for allowing us to use the Apogee flow cytometer and for technical assistance with flow cytometry. We thank Dr. Beth Carpenter and Yuliya Seldina for critical reading and comments on the manuscript.

# CHAPTER FOUR: Discussion

## SUMMARY OF FINDINGS AND RELEVANCE

Infection with *H. pylori* remains the most significant risk factor for development of gastric cancer (55; 165). While treatment of *H. pylori* has been shown to alleviate, and even lead to regression of disease, in some cases, there appears to be a "point of no return" after which treatment of *H. pylori* is unable to reverse the pathological effects of infection (reviewed in (113)). Given these observations, early identification of those individuals who will progress to symptomatic disease is paramount to successful treatment. Unfortunately, even if these individuals are identified, *H. pylori* treatment failure rates are around 20% (202; 206; 222). Thus, a better understanding of the genetic mechanisms controlling *H. pylori* pathogenesis is needed in order to solve these problems.

In chapter 2, we took a systematic approach to better understand the idea of coordinated regulation in *H. pylori*. Numerous transcriptional studies highlight the overlapping regulons of three prominent *H. pylori* transcriptional regulators: Fur, NikR, and ArsRS. However, few studies have sought to directly assess the effect of the combined loss of these regulatory proteins. While deletion of all three regulators was tolerated *in vitro*, a previously unreported biofilm phenotype was observed in all strains lacking the ArsS sensor kinase. This phenotype was recapitulated in the ArsR-D52N non-phosphorylatable strain, which suggests that biofilm formation in *H. pylori* is regulated by ArsR~P. Not only is this the first description of a transcriptional regulator involved in

*H. pylori* biofilm formation, but it also hints at the environmental cues that influence *H. pylori* to transition to a biofilm-associated lifestyle.

In parallel to the top-down approach detailed in chapter 2, chapter 3 focused on a single gene, *homB*, which has been associated with the development of gastric disease in some populations. *H. pylori* has a variable repertoire of OMPs that contribute to colonization and adaptation within the human host. To date, a majority of the focus on OMPs has been dedicated to the study of the major OMP family, Hop, leaving unanswered questions regarding OMPs from the smaller families. To this end, we found that expression of *homB,* a member of the Hom family, was suppressed under conditions of low pH in an ArsRS-dependent manner. We also discovered that *homB* contributes to biofilm formation in the *H. pylori* strain G27; however, the precise role of *homB* in biofilm formation has yet to be elucidated.

Through the course of this dissertation work, we described a previously uncharacterized role for ArsRS in biofilm formation and identified a factor in the ArsRS regulatory cascade that contribute to biofilm formation. The refractory nature of biofilms to host defenses and antibiotic treatment is well established, and a greater understanding of the mechanisms of biofilm formation by *H. pylori* may help to improve the current treatment regimen.

## BIOFILM FORMATION BY *H. PYLORI*

Biofilm formation by *H. pylori* was first described in the late 1990s, and has since been demonstrated during *in vitro* growth (45; 189; 219), during animal infection (17),

and in natural human infection (33; 39; 41; 52). Several studies have taken a broad approach and sought to identify the *in vitro* conditions that affect biofilm formation (27; 45; 183). Indeed, the effects of different types of media (27), sera concentrations (183), particular types of metal availability (76), sub-inhibitory concentrations of antimicrobials (27; 217), and the presence of mucin (45) have each been evaluated to varying degrees. Most notably, researchers demonstrated that *in vitro* biofilm formation by *H. pylori* could be induced by sub-inhibitory doses of antibiotics or antimicrobial peptides (27; 76; 217). Additional environmental cues such as oxidative stress, ion availability, and pH have yet to be evaluated; however, the role of ArsRS in biofilm formation that is presented in chapter 2 suggests that pH changes may be an important signal for biofilm formation.

Through the sensing of environmental cues and by coordination of the proper transcriptional response two-component systems play an integral role in bacterial growth and development. Thus, it is not surprising that two-component systems have been associated with stress response, and in some cases, biofilm development in several pathogens (reviewed in (25; 88)). For example, in *C. jejuni*, deletion of the sensor kinase from the CprRS two-component system was shown to accelerate biofilm formation, but to decrease tolerance to osmotic and oxidative stress (195). Similarly, we identified the acid responsive two-component system, ArsRS, as a regulator of biofilm formation by *H. pylori* (182). Deletion of *arsS* or mutation of *arsR* resulted in increased biofilm formation at neutral pH. These observations led to the hypothesis that biofilm formation by *H. pylori* is repressed at low pH, where ArsR would exist in its phosphorylated state. Upon subsequent relocation to a more neutral pH, ArsR~P levels would decrease and biofilm suppression would be alleviated (Figure 12). Though an attractive model, future studies

that compare the effect of pH on biofilm development are required to test this hypothesis. Interestingly, previous work following 12-hours of serum starvation, showed a decrease in ArsR expression was observed in biofilm-associated *H. pylori* as compared to planktonic bacteria (183). Thus, perhaps ArsRS regulates biofilm formation in response to multiple factors. Furthermore, recent data found that transcriptional regulation of ArsR is mediated not only by ArsR, but also by Fur and NikR (170). This finding suggests the potential for further integration of upstream signals on biofilm formation.

Through the course of this work, we focused on the enhanced biofilm phenotype observed in the Δ*arsS* strain; however, as described in chapter 2, characterization of the biofilm formation by Δ*fur*, Δ*nikR*, and Δ*fur*Δ*nikR* revealed a delay in biofilm formation as compared to wild-type. This finding suggests that these metalloregulators may also contribute to biofilm formation. Metal responsive regulation of biofilm formation presents an intriguing line of inquiry. To this end, one of the key mechanisms of action of the antimicrobial peptide, Calprotectin (Cp), is nutritional immunity via sequestration of zinc and manganese (76). Use of the metal chelator, N,N,N',N'-tetrakis(2-pyridylmethyl)ethane-1,2-diamine (TPEN) to mimic zinc sequestration was shown to induce biofilm formation similar to treatment with Cp (76). Together, those prior studies suggest that metal availability may also be a signal for biofilm formation. Future studies that implement the use of specific chelators, excess metal, or a combination of the two should be completed to better define the role of metalloregulation on biofilm formation.

## COMPONENTS OF THE *H. PYLORI* BIOFILM

One focus of biofilm research is to identify proteins and other molecules that contribute to the biofilm matrix. Key components of the biofilm matrix can then be used as targets for therapeutics; blocking or inhibiting the matrix components could enable the destruction of the biofilm and enhance infection clearance. To date, components of the *H. pylori* biofilm matrix that have been identified include mannose containing proteoglycans and OMVs (92; 216; 218). In addition, the eDNA associated with OMVs is also thought to contribute to biofilm formation (92).

While polysaccharides and OMVs comprise key components of the biofilm matrix, evidence also suggests that the composition of the cell envelope also influences biofilm formation. For example, the deletion of *cagE* or *napA* both result in modification of the cell surface and concomitant changes in biofilm formation (45; 216). As previously noted, the presence of alternative lipid A structures not only enhanced Cp resistance, but also enhanced biofilm formation (76). In the current work, we found that deletion of the ArsRS-regulated OMP, HomB, also resulted in decreased biofilm formation (chapter 3). It is worth noting that changes in the cell envelope may also be reflected in the composition of OMVs. In fact, difference in the OMV protein profiles between biofilm forming strains and non-biofilm forming strains have been noted (220). Perhaps a more in-depth evaluation of the biofilm-associated OMV OMP profiles as compared to whole cell OMP profiles would provide interesting insight into the role of OMV OMPs in biofilm formation. In addition, a potential experiment to assess the role of OMVs in biofilm formation could involve the adoptive transfer of OMVs from high biofilm forming strains to strains with low-biofilm forming capacity. One could then determine if presence of the OMVs is sufficient to induce biofilm formation.

Overall, many gaps remain in our understanding of biofilm formation by *H. pylori*. One confounding problem in the ability to draw broad conclusions from the current body of literature on this topic arises from the discordance between the various methods used to culture the biofilms. For example, in the data presented here and by some others, biofilm formation occurs in the presence of sera (chapter 2 , chapter 3, (219). In comparison, other studies suggest that sera inhibits biofilm formation; one study went as far as to show that serum starvation induces biofilm formation in as little as 12 hours (27; 183). Additionally, similar to our biofilm assays, some studies are conducted using tissue culture treated polystyrene plates, whereas other use a variety of surfaces that include glass and untreated polystyrene plates and/or coverslips. The inconsistencies between these study conditions may explain why some studies report biofilm formation by a particular strain, which are not reproducible when the same strain is tested under other conditions (45; 219). One way to enable the field to better characterize biofilm formation would be the development of a standardized, reproducible, high-throughput biofilm assay. Implementation of a defined medium and culture surface could aid our understanding of the biofilm proclivity of individual strains. This type of assay might also enable us to ascertain whether or not all *H. pylori* strains are capable of forming biofilm, if all strains form biofilm to a similar degree, or if there is an association between biofilm formation and virulence. Additionally, this assay could systematically be used to test various growth conditions, which may induce or inhibit biofilm formation, as well as the effectiveness of a broad range of antibiotics or antibiofilm compounds.

Given the lack of knowledge concerning which *H. pylori* genes are involved in biofilm formation, utilization of a transposon mutant library followed by screening for

biofilm deficient strains could also aid in the identification of genes that contribute to biofilm formation in *H. pylori*. Given that we know the Δ*arsS* strain forms an enhanced biofilm, perhaps this strain background could be used for the transposon mutagenesis. Each mutant strain could then be compared to Δ*arsS* to identify strains showing reduced biofilm formation. Of note, the ArsRS regulon is predicted to include anywhere from 100-200 genes. To date we have only identified one gene within this regulon, *homB,* which is related to biofilm formation. Transposon mutagenesis could likewise be conducted in wild-type G27 to see if other mutations lead to biofilm phenotypes: enhanced or decreased. Once again, having a more high-throughput system would make screening of these transposon libraries much more efficient. Clearly, implementation of high throughput screening techniques could accelerate progress towards a greater understanding of biofilm development, as well as facilitate the ultimate goal of identification of better therapeutics.

## UNRAVELING *H. PYLORI* TRANSCRIPTIONAL NETWORKS

Early transcriptional studies identified overlap between the Fur, NikR, and ArsRS regulons; however, only a handful of studies, including work from our group, have addressed the idea of coordinate regulation. Notably, coordination between NikR and ArsR was shown to be required for optimal *ureA* transcription (37). Additionally, a recent publication detailed an intricate study of the interplay between NikR, Fur, and ArsR in the regulation of the ArsRS operon (170). In chapter 2, we examined the growth kinetics of a triple mutant strain in which *fur*, *nikR*, and *arsS* were all deleted. Additionally, all

possible single and double mutant combination were examined. The construction of these isogenic strains offers a starting point for future global transcriptional studies to compare gene expression between wild-type, single, double and even triple deletion strains. Furthermore, we note that the studies conducted herein only utilized rich media. Thus, future evaluation of the stress response (e.g. metal chelation, pH shock, oxidative shock, etc) in these strains could help to reveal previously unrecognized regulatory hierarchies.

In addition to comparing the combined effect of mutation of the different regulators, we also directly compared phenotypes demonstrated by the Δ*arsS* and ArsR-D52N mutant strains. The ArsRS system is comprised of the ArsS sensor kinase and the ArsR response regulator. Early investigations of ArsRS revealed that, while the sensor kinase is dispensable for *in vitro* growth, the *arsR* regulatory gene is essential (24; 134; 176). It was subsequently determined that even though ArsR is essential, mutations in the ArsR phosphate-accepting residue (Aspartate at position 52) are tolerated *in vitro* (176). Therefore, mutation of residue 52 of ArsR or deletion of *arsS* should both nullify the pH responsive arm of this two-component system. It is worth noting that while several studies have sought to delineate the ArsRS regulon, few studies have compared the *arsS* deletion strain directly to the ArsR-D52N mutation. Furthermore, investigations of the ArsRS-dependent transcriptome have largely relied on comparison between a Δ*arsS* strain and wild-type *H. pylori* following exposure to low pH (62; 120; 130; 163; 208). Interestingly, our study revealed that while these two mutations are comparable, strains bearing the individual mutations do show phenotypic differences in biofilm formation, biofilm architecture, adherence and aggregation. Similarly, a recent study identified differences in trafficking of the urease proteins to the membrane in ArsR-D52N and

*ΔarsS* strains (130). Additionally, the response to pH shock differed between ArsR-D52N and *ΔarsS*; supplementation with urea led to a rapid increase in cytoplasmic pH of the wild-type and ArsR-D52N strains, whereas little to no change was observed in strains lacking ArsS (130). In light of these observed differences, it would be interesting to determine if the ArsR-D52N strain shows a similar growth defect during animal infection as has been observed with the *ΔarsS* strain (158). Given the observed phenotypic differences between the various ArsRS mutant strains presented in the second and third chapter of this dissertation, there is clearly a need for future work to define the roles of ArsR, ArsS, and ArsRS in *H. pylori* biology. For example, Chromatin immunoprecipitation using purified ArsR or ArsR-D52N followed by high-throughput sequencing (ChIP-seq) could help to identify unique targets of ArsR as compared to ArsR~P. Additionally, RNA-seq could be utilized to quickly identify differences in acid responsive regulons between strains lacking ArsS as compared to the non-phosphorylateable ArsR. RNA-Seq and ChIP-seq could also provide insight into the identification of the essential ArsR targets, an area of study that has been largely ignored.

## OUTER MEMBRANE HETEROGENEITY

The panmitic structure of the *H. pylori* genome is well recognized, and this variability among strains is reflected in the composition of the cell envelope. One of the main sources of this variation arises from difference in the OMP profile. Previous evaluation of OMP expression in clinical isolates revealed that the frequencies of expression of the various OMPs range from expression in 100% of strains to expression

in only 35% of strains (150). In addition to interstrain diversity, passage of strains *in vitro* or through an animal model has been repeatedly shown to result in variable OMP presentation (74; 214). This variation mimics data from long-term non-human primate colonization studies and natural human infections; strains isolated from a single animal or individual over time show variation in OMP expression (101; 119; 187). The diversity in OMP profiles can arise by deletion and/or genomic rearrangement of specific OMPs and also from phase variation (149; 173). Moreover, as demonstrated in chapter 3, some OMPs are also under the influence of transcriptional regulatory proteins.

Given their location at the interface between the bacterium and the host or environment, the array of *H. pylori* OMPs is of particular interest in the *H. pylori* field. As such, numerous studies have sought to assign biological roles for a handful of OMPs. Other studies have sought to identify any associations between various OMPs and the clinical outcome of infection. Several *H. pylori* OMPs have been identified as adhesins, and in some cases, the cognate host receptor has been described (reviewed in (156)). Despite these findings, there is still a huge gap in our knowledge regarding how and under what conditions certain OMPs are required. For instance, while some studies have demonstrated that *hopQ* is necessary for adherence and optimal CagA translocation into host cells, other findings suggest that *hopQ* acted as an anti-adhesin (26; 103; 112; 122). Similarly, initial *homB* characterization showed that the effect of *homB* on adherence varied between tested strains (152). Additionally, when interstrain adherence profiles were compared, differences were also observed (152). Together, these findings suggest that perhaps the collection of OMPs expressed by a particular isolate, and the expression of OMPs at any given time, may greatly influence the ability of the strains to adhere. This

is further exacerbated by the potential redundancy in function of the numerous OMP adhesins. While it would be a significant undertaking, examination of the presence and expression of different combinations of OMPs in an isogenic strain background could be one way to tease out the various interactions and potential redundancies between the numerous OMPs. For example, we are currently constructing various isogenic mutant strains in the *H. pylori* G27 background in order to examine differences between HomA and HomB. Once the strains are constructed, we will have a collection of *H. pylori* G27 strains that carry a single copies of *homA* or *homB* at either locus A or locus B, strains that carry two copies of *homA* or *homB*, and strains that carry a copy of *homA* and *homB* (see Figure 1 as a reference). Thus, we will be able to systematically investigate the role of each of the loci and proteins in *H. pylori* biology. Although HomC and HomD are also members of the Hom family, they are universally found in all strains tested thus far. Because of this, we have chosen to focus our initial studies on HomA and HomB. However, future studies that focus on the different alleles that we have described for *homC* would also be of interest (Appendix B and (48)). We note that a similar strategy could also be used to evaluate the presence of different combinations of *bab* alleles or the *sab* alleles. Due to the heterogeneity seen across *H. pylori* strains, isogenic strain construction clearly presents the best way to tease out interactions between the various OMPs as well as potential redundant functions.

Given that surface adherence and cell-cell interactions are important processes in the initiation of biofilm formation, differences in OMP profiles may affect the biofilm-forming ability of various strains. NapA and CagE are both cell-envelope associated proteins that are predicted to contribute to biofilm formation (45; 216). Moreover, in the

current study, we demonstrated a role for HomB in biofilm formation by G27 (chapter 3). It is clear that given the large level of *H. pylori* heterogeneity, the applicability of our and others findings will need to be examined across additional strains. For example, HomB is not present in all *H. pylori* isolates. Given that *homB* is over 90% homologous to *homA*, does HomA play a role in biofilm formation in the absence of HomB? To date, no strains lacking both *homA*/*homB* have been identified. Thus perhaps, utilization of the isogenic mutant strains described above could test the hypothesis that HomA and HomB provide analogous functions in biofilm formation.

We have focused this discussion primarily on the role of OMPs as bacterial adhesins that contribute to biofilm formation; however, LPS is also a variable outer membrane component that has been associated with both adherence and biofilm formation (14; 66; 115). Published data show that lipid A modifications promote biofilm formation. However, the O-antigen of *H. pylori* is also variable and this variability has yet to be investigated in biofilm associated *H. pylori* (76). The O-antigen of *H. pylori* is distinctive due to the O-side chains that mimic type 1 and/or type 2 blood group antigens (14; 115). Slip-strand mispairing in the enzymes responsible for construction of the O-antigen side chains leads to phase variation in LPS and increased variability(14; 115). Indeed, variations in LPS were detected in response to pH shock (pH 5) (135); thus LPS modifications may occur in response to stress. Taken together, these data emphasize the need to not only investigate the role of OMPs in biofilm formation, but the need to also evaluate the role of LPS and other outer membrane modifications as well.

## CONCLUSION

Infection with *H. pylori* can have truly devastating outcomes. This bacterium is currently the leading cause of infection-associated cancers (165). Despite the decreasing trend in number of gastric cancer cases reported worldwide, gastric cancer still ranks as the fifth most common malignancy in the world, and is the third leading cause of cancer related death world-wide (165). What makes these statistics so striking is the fact that *H. pylori* is a treatable infection. Unfortunately, the current treatment failure rate is approximately 20%, and antibiotic resistance is on the rise (222). Herein, we present insight into the alternative biofilm life-style of *H. pylori*. Given that biofilms are inherently more resistant to antibiotics, biofilm development by *H. pylori* may contribute to the 20% treatment failure. The discoveries presented in this thesis may help to lay the groundwork for more in-depth studies of the mechanisms that underpin biofilm formation. Indeed, a greater understanding of these mechanisms and the role of the biofilm in colonization by *H. pylori* may help to identify bacterial factors that could be used as targets for novel treatment and prevention strategies.

# REFERENCES


1. 1994. Schistosomes, liver flukes and *Helicobacter pylori*. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7-14 June 1994. *IARC Monogr Eval Carcinog Risks Hum* 61:1-241

2. Abdullah SM, Hussein NR, Salih AM, Merza MA, Goreal AA, et al. 2012. Infection with *Helicobacter pylori* strains carrying *babA2* and *cagA* is associated with an increased risk of peptic ulcer disease development in Iraq. *Arab J Gastroenterol* 13:166-9

3. Aberg A, Gideonsson P, Vallstrom A, Olofsson A, Ohman C, et al. 2014. A repetitive DNA element regulates expression of the *Helicobacter pylori* sialic acid binding adhesin by a rheostat-like mechanism. *PLoS Pathog* 10:e1004234

4. Alfizah H, Ramelah M, Rizal AM, Anwar AS, Isa MR. 2012. Association of Malaysian *Helicobacter pylori* virulence polymorphisms with severity of gastritis and patients' ethnicity. *Helicobacter* 17:340-9

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

6. Alm RA, Ling LS, Moir DT, King BL, Brown ED, et al. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 397:176-80

7. Alm RA, Trust TJ. 1999. Analysis of the genetic diversity of *Helicobacter pylori*: the tale of two genomes. *J Mol Med (Berl)* 77:834-46

8. Anand PS, Kamath KP, Anil S. 2014. Role of dental plaque, saliva and periodontal disease in *Helicobacter pylori* infection. *World J Gastroenterol* 20:5639-53

9. Andersen LP. 2007. Colonization and infection by *Helicobacter pylori* in humans. *Helicobacter* 12 Suppl 2:12-5

10. Andersen LP, Rasmussen L. 2009. *Helicobacter pylori*-coccoid forms and biofilm formation. *FEMS Immunol Med Microbiol* 56:112-5

11. Andersen LP, Wadstrom T. 2001. Basic Bacteriology and Culture. In *Helicobacter pylori: Physiology and Genetics*, ed. HLT Mobley, GL Mendz, SL Hazell. Washington (DC). Number of.

12. Anderson JK, Huang JY, Wreden C, Sweeney EG, Goers J, et al. 2015. Chemorepulsion from the Quorum Signal Autoinducer-2 Promotes *Helicobacter pylori* Biofilm Dispersal. *MBio* 6:e00379

13. Angst EC. 1923. The fouling of ships bottoms by bacteria. In *bureau construction and repair*, ed. USN Department

14. Appelmelk BJ, Vandenbroucke-Grauls C. 2001. Lipopolysaccharide Lewis Antigens. In *Helicobacter pylori: Physiology and Genetics*, ed. HLT Mobley, GL Mendz, SL Hazell. Washington (DC). Number of.

15. Armitano RI, Matteo MJ, Goldman C, Wonaga A, Viola LA, et al. 2013. *Helicobacter pylori* heterogeneity in patients with gastritis and peptic ulcer disease. *Infect Genet Evol* 16:377-85

16. Atapoor S, Safarpoor Dehkordi F, Rahimi E. 2014. Detection of *Helicobacter pylori* in Various Types of Vegetables and Salads. *Jundishapur J Microbiol* 7:e10013

17. Attaran B, Falsafi T, Moghaddam AN. 2016. Study of biofilm formation in C57Bl/6J mice by clinical isolates of *Helicobacter pylori*. *Saudi J Gastroenterol* 22:161-8

18. Azevedo NF, Guimaraes N, Figueiredo C, Keevil CW, Vieira MJ. 2007. A new model for the transmission of *Helicobacter pylori*: role of environmental reservoirs as gene pools to increase strain diversity. *Crit Rev Microbiol* 33:157-69

19. Azevedo NF, Pacheco AP, Keevil CW, Vieira MJ. 2006. Adhesion of water stressed *Helicobacter pylori* to abiotic surfaces. *J Appl Microbiol* 101:718-24

20. Azevedo NF, Vieira MJ, Keevil CW. 2003. Establishment of a continuous model system to study *Helicobacter pylori* survival in potable water biofilms. *Water Sci Technol* 47:155-60

21. Baltrus DA, Amieva MR, Covacci A, Lowe TM, Merrell DS, et al. 2009. The complete genome sequence of *Helicobacter pylori* strain G27. *J Bacteriol* 191:447-8

22. Bauer B, Meyer TF. 2011. The Human Gastric Pathogen *Helicobacter pylori* and Its Association with Gastric Cancer and Ulcer Disease. *Ulcers* 2011

23. Becker KW, Skaar EP. 2014. Metal limitation and toxicity at the interface between host and pathogen. *FEMS Microbiol Rev* 38:1235-49

24. Beier D, Frank R. 2000. Molecular characterization of two-component systems of *Helicobacter pylori*. *J Bacteriol* 182:2068-76

25. Beier D, Gross R. 2006. Regulation of bacterial virulence by two-component systems. *Curr Opin Microbiol* 9:143-52

26. Belogolova E, Bauer B, Pompaiah M, Asakura H, Brinkman V, et al. 2013. *Helicobacter pylori* outer membrane protein HopQ identified as a novel T4SS-associated virulence factor. *Cell Microbiol* 15:1896-912

27. Bessa LJ, Grande R, Di Iorio D, Di Giulio M, Di Campli E, Cellini L. 2013. *Helicobacter pylori* free-living and biofilm modes of growth: behavior in response to different culture media. *APMIS* 121:549-60

28. Bijlsma JJ, Waidner B, Vliet AH, Hughes NJ, Hag S, et al. 2002. The *Helicobacter pylori* homologue of the ferric uptake regulator is involved in acid resistance. *Infect Immun* 70:606-11

29. 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

30. Breurec S, Guillard B, Hem S, Brisse S, Dieye FB, et al. 2011. Evolutionary history of *Helicobacter pylori* sequences reflect past human migrations in Southeast Asia. *PLoS One* 6:e22058

31. Bridge DR, Merrell DS. 2013. Polymorphism in the *Helicobacter pylori* CagA and VacA toxins and disease. *Gut Microbes* 4:101-17

32. Bury-Mone S, Thiberge JM, Contreras M, Maitournam A, Labigne A, De Reuse H. 2004. Responsiveness to acidity via metal ion regulators mediates virulence in the gastric pathogen *Helicobacter pylori*. *Mol Microbiol* 53:623-38

33. Cammarota G, Branca G, Ardito F, Sanguinetti M, Ianiro G, et al. 2010. Biofilm demolition and antibiotic treatment to eradicate resistant *Helicobacter pylori*: a clinical trial. *Clin Gastroenterol Hepatol* 8:817-20 e3

34. Cammarota G, Sanguinetti M, Gallo A, Posteraro B. 2012. Review article: biofilm formation by *Helicobacter pylori* as a target for eradication of resistant infection. *Aliment Pharmacol Ther* 36:222-30

35. Carpenter BM, Gancz H, Gonzalez-Nieves RP, West AL, Whitmire JM, et al. 2009. A single nucleotide change affects fur-dependent regulation of *sodB* in *H. pylori*. *PLoS One* 4:e5369

36. Carpenter BM, McDaniel TK, Whitmire JM, Gancz H, Guidotti S, et al. 2007. Expanding the *Helicobacter pylori* genetic toolbox: modification of an endogenous plasmid for use as a transcriptional reporter and complementation vector. *Appl Environ Microbiol* 73:7506-14

37. Carpenter BM, West AL, Gancz H, Servetas SL, Pich OQ, et al. 2015. Crosstalk between the HpArsRS two-component system and HpNikR is necessary for maximal activation of urease transcription. *Front Microbiol* 6:558

38. Carpenter BM, Whitmire JM, Merrell DS. 2009. This is not your mother's repressor: the complex role of *fur* in pathogenesis. *Infect Immun* 77:2590-601

39. Carron MA, Tran VR, Sugawa C, Coticchia JM. 2006. Identification of *Helicobacter pylori* biofilms in human gastric mucosa. *J Gastrointest Surg* 10:712-7

40. Cellini L, Del Vecchio A, Di Candia M, Di Campli E, Favaro M, Donelli G. 2004. Detection of free and plankton-associated *Helicobacter pylori* in seawater. *J Appl Microbiol* 97:285-92

41. Cellini L, Grande R, Di Campli E, Traini T, Di Giulio M, et al. 2008. Dynamic colonization of *Helicobacter pylori* in human gastric mucosa. *Scand J Gastroenterol* 43:178-85

42. Characklis WG. 1973. Attahced microbial growth -II. Frictional resistance due to microbial slimes. *Water Research* 7:1249-58

43. Choi YW, Park SA, Lee HW, Lee NG. 2009. Alteration of growth-phase-dependent protein regulation by a *fur* mutation in *Helicobacter pylori*. *FEMS Microbiol Lett* 294:102-10

44. Clyne M, Ocroinin T, Suerbaum S, Josenhans C, Drumm B. 2000. Adherence of isogenic flagellum-negative mutants of *Helicobacter pylori* and Helicobacter mustelae to human and ferret gastric epithelial cells. *Infect Immun* 68:4335-9

45. Cole SP, Harwood J, Lee R, She R, Guiney DG. 2004. Characterization of monospecies biofilm formation by *Helicobacter pylori*. *J Bacteriol* 186:3124-32

46. Contreras M, Thiberge JM, Mandrand-Berthelot MA, Labigne A. 2003. Characterization of the roles of NikR, a nickel-responsive pleiotropic autoregulator of *Helicobacter pylori*. *Mol Microbiol* 49:947-63

47. Copass M, Grandi G, Rappuoli R. 1997. Introduction of unmarked mutations in the *Helicobacter pylori vacA* gene with a sucrose sensitivity marker. *Infect Immun* 65:1949-52

48. Cordeiro RM, A.; Breurec, S.; Megraud, F.; Oleastro, M. Diversity and phylogeny of the *Helicobacter pylori* outer membrane protein encoding gene *homC*. *Proc. CHRO*, *Vancouver, Canada*, *2011*

49. Correa P, Piazuelo MB. 2012. Evolutionary History of the *Helicobacter pylori* Genome: Implications for Gastric Carcinogenesis. *Gut Liver* 6:21-8

50. Correa P, Piazuelo MB. 2012. The gastric precancerous cascade. *J Dig Dis* 13:2-9

51. Costerton JW, Geesey GG, Cheng KJ. 1978. How bacteria stick. *Sci Am* 238:86-95

52. Coticchia JM, Sugawa C, Tran VR, Gurrola J, Kowalski E, Carron MA. 2006. Presence and density of *Helicobacter pylori* biofilms in human gastric mucosa in patients with peptic ulcer disease. *J Gastrointest Surg* 10:883-9

53. Covacci A, Censini S, Bugnoli M, Petracca R, Burroni D, et al. 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

54. Covacci A, Telford JL, Del Giudice G, Parsonnet J, Rappuoli R. 1999. *Helicobacter pylori* virulence and genetic geography. *Science* 284:1328-33

55. Cover TL. 2016. *Helicobacter pylori* Diversity and Gastric Cancer Risk. *MBio* 7:e01869-15

56. Danielli A, Amore G, Scarlato V. 2010. Built shallow to maintain homeostasis and persistent infection: insight into the transcriptional regulatory network of the gastric human pathogen *Helicobacter pylori*. *PLoS Pathog* 6:e1000938

57. Danielli A, Roncarati D, Delany I, Chiarini V, Rappuoli R, Scarlato V. 2006. *in vivo* dissection of the *Helicobacter pylori* Fur regulatory circuit by genome-wide location analysis. *J Bacteriol* 188:4654-62

58. Danielli A, Scarlato V. 2010. Regulatory circuits in *Helicobacter pylori* : network motifs and regulators involved in metal-dependent responses. *FEMS Microbiol Rev* 34:738-52

59. de Jonge R, Durrani Z, Rijpkema SG, Kuipers EJ, van Vliet AH, Kusters JG. 2004. Role of the *Helicobacter pylori* outer-membrane proteins AlpA and AlpB in colonization of the guinea pig stomach. *J Med Microbiol* 53:375-9

60. de Vries N, van Vliet AHM, Kusters JG. 2001. Gene Regulation. In *Helicobacter pylori: Physiology and Genetics*, ed. HLT Mobley, GL Mendz, SL Hazell. Washington (DC). Number of.

61. Delany I, Ieva R, Soragni A, Hilleringmann M, Rappuoli R, Scarlato V. 2005. In vitro analysis of protein-operator interactions of the NikR and fur metal-

responsive regulators of coregulated genes in *Helicobacter pylori*. *J Bacteriol* 187:7703-15

62. Dietz P, Gerlach G, Beier D. 2002. Identification of target genes regulated by the two-component system HP166-HP165 of *Helicobacter pylori*. *J Bacteriol* 184:350-62

63. Dixon MF. 2001. Pathology of Gastritis and Peptic Ulceration. In *Helicobacter pylori: Physiology and Genetics*, ed. HLT Mobley, GL Mendz, SL Hazell. Washington (DC). Number of.

64. Doig P, Trust TJ. 1994. Identification of surface-exposed outer membrane antigens of *Helicobacter pylori*. *Infect Immun* 62:4526-33

65. Dong QJ, Wang Q, Xin YN, Li N, Xuan SY. 2009. Comparative genomics of *Helicobacter pylori*. *World J Gastroenterol* 15:3984-91

66. Edwards NJ, Monteiro MA, Faller G, Walsh EJ, Moran AP, et al. 2000. Lewis X structures in the O antigen side-chain promote adhesion of *Helicobacter pylori* to the gastric epithelium. *Mol Microbiol* 35:1530-9

67. Ernst FD, Bereswill S, Waidner B, Stoof J, Mader U, et al. 2005. Transcriptional profiling of *Helicobacter pylori* Fur- and iron-regulated gene expression. *Microbiology* 151:533-46

68. Ernst FD, Kuipers EJ, Heijens A, Sarwari R, Stoof J, et al. 2005. The nickel-responsive regulator NikR controls activation and repression of gene transcription in *Helicobacter pylori*. *Infect Immun* 73:7252-8

69. Ernst PB, Gold BD. 2000. The disease spectrum of *Helicobacter pylori*: the immunopathogenesis of gastroduodenal ulcer and gastric cancer. *Annu Rev Microbiol* 54:615-40

70. Evans DJ, Jr., Evans DG. 2000. *Helicobacter pylori* adhesins: review and perspectives. *Helicobacter* 5:183-95

71. Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, et al. 2003. Traces of human migrations in *Helicobacter pylori* populations. *Science* 299:1582-5

72. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269:496-512

73. Foster JW. 2004. *Escherichia coli* acid resistance: tales of an amateur acidophile. *Nat Rev Microbiol* 2:898-907

74. Franco AT, Friedman DB, Nagy TA, Romero-Gallo J, Krishna U, et al. 2009. Delineation of a carcinogenic *Helicobacter pylori* proteome. *Mol Cell Proteomics* 8:1947-58

75. Freedberg A.S. BLE. 1940. The presence of spirochetes in human gastric mucosa. *American Journal of Digestive Disease* 7:443-5

76. Gaddy JA, Radin JN, Cullen TW, Chazin WJ, Skaar EP, et al. 2015. *Helicobacter pylori* Resists the Antimicrobial Activity of Calprotectin via Lipid A Modification and Associated Biofilm Formation. *MBio* 6:e01349-15

77. Gaddy JA, Tomaras AP, Actis LA. 2009. The *Acinetobacter baumannii* 19606 OmpA protein plays a role in biofilm formation on abiotic surfaces and in the interaction of this pathogen with eukaryotic cells. *Infect Immun* 77:3150-60

78. Gancz H, Censini S, Merrell DS. 2006. Iron and pH homeostasis intersect at the level of Fur regulation in the gastric pathogen *Helicobacter pylori*. *Infect Immun* 74:602-14

79. Gancz H, Merrell DS. 2011. The *Helicobacter pylori* Ferric Uptake Regulator (Fur) is essential for growth under sodium chloride stress. *J Microbiol* 49:294-8

80. Garcia A, Salas-Jara MJ, Herrera C, Gonzalez C. 2014. Biofilm and *Helicobacter pylori*: from environment to human host. *World J Gastroenterol* 20:5632-8

81. Giao MS, Azevedo NF, Wilks SA, Vieira MJ, Keevil CW. 2008. Persistence of *Helicobacter pylori* in heterotrophic drinking-water biofilms. *Appl Environ Microbiol* 74:5898-904

82. Giao MS, Azevedo NF, Wilks SA, Vieira MJ, Keevil CW. 2011. Interaction of *Legionella pneumoph*ila and *Helicobacter pylori* with bacterial species isolated from drinking water biofilms. *BMC Microbiol* 11:57

83. Gilbreath JJ, West AL, Pich OQ, Carpenter BM, Michel S, Merrell DS. 2012. Fur activates expression of the 2-oxoglutarate oxidoreductase genes (*oorDABC*) in *Helicobacter pylori*. *J Bacteriol* 194:6490-7

84. Gisbert JP. 2008. "Rescue" regimens after *Helicobacter pylori* treatment failure. *World J Gastroenterol* 14:5385-402

85. Goh KL, Chan WK, Shiota S, Yamaoka Y. 2011. Epidemiology of *Helicobacter pylori* infection and public health implications. *Helicobacter* 16 Suppl 1:1-9

86. Goodwin AC, Weinberger DM, Ford CB, Nelson JC, Snider JD, et al. 2008. Expression of the *Helicobacter pylori* adhesin SabA is controlled via phase variation and the ArsRS signal transduction system. *Microbiology* 154:2231-40

87. Goodwin C, Collins M, Blincow E. 1986. The absence of thermoplasmaquinones in Campylobacter pyloridis, and its temperature and pH growth range. *Microbios letters* 32:137-40

88. Gotoh Y, Eguchi Y, Watanabe T, Okamoto S, Doi A, Utsumi R. 2010. Two-component signal transduction as potential drug targets in pathogenic bacteria. *Curr Opin Microbiol* 13:232-9

89. Graham DY. 1998. Antibiotic resistance in *Helicobacter pylori*: implications for therapy. *Gastroenterology* 115:1272-7

90. Grande R, Di Campli E, Di Bartolomeo S, Verginelli F, Di Giulio M, et al. 2012. *Helicobacter pylori* biofilm: a protective environment for bacterial recombination. *J Appl Microbiol* 113:669-76

91. Grande R, Di Giulio M, Di Campli E, Di Bartolomeo S, Cellini L. 2010. A model of *Helicobacter pylori* persistence in a case of gastric cancer. *New Microbiol* 33:343-9

92. Grande R, Di Marcantonio MC, Robuffo I, Pompilio A, Celia C, et al. 2015. *Helicobacter pylori* ATCC 43629/NCTC 11639 Outer Membrane Vesicles (OMVs) from Biofilm and Planktonic Phase Associated with Extracellular DNA (eDNA). *Front Microbiol* 6:1369

93. Gressmann H, Linz B, Ghai R, Pleissner KP, Schlapbach R, et al. 2005. Gain and loss of multiple genes during the evolution of *Helicobacter pylori*. *PLoS Genet* 1:e43

94. Haley KP, Gaddy JA. 2015. Metalloregulation of *Helicobacter pylori* physiology and pathogenesis. *Front Microbiol* 6:911

95. Hall-Stoodley L, Costerton JW, Stoodley P. 2004. Bacterial biofilms: from the natural environment to infectious diseases. *Nat Rev Microbiol* 2:95-108

96. Harvey VC, Acio CR, Bredehoft AK, Zhu L, Hallinger DR, et al. 2014. Repetitive sequence variations in the promoter region of the adhesin-encoding gene *sabA* of *Helicobacter pylori* affect transcription. *J Bacteriol* 196:3421-9

97. Hennig EE, Allen JM, Cover TL. 2006. Multiple chromosomal loci for the *babA* gene in *Helicobacter pylori*. *Infect Immun* 74:3046-51

98. Henrici AT. 1933. Studies of Freshwater Bacteria: I. A Direct Microscopic Technique. *J Bacteriol* 25:277-87

99. Heukelekian H, Heller A. 1940. Relation between Food Concentration and Surface for Bacterial Growth. *J Bacteriol* 40:547-58

100. Hussein NR. 2011. A study of *Helicobacter pylori* outer-membrane proteins (*hom*) A and B in Iraq and Turkey. *J Infect Public Health* 4:135-9

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

102. Jang S, Jones KR, Olsen CH, Joo YM, Yoo YJ, et al. 2010. Epidemiological link between gastric disease and polymorphisms in VacA and CagA. *J Clin Microbiol* 48:559-67

103. Javaheri A, Kruse T, Moonens K, Mejias-Luque R, Debraekeleer A, et al. 2016. *Helicobacter pylori* adhesin HopQ engages in a virulence-enhancing interaction with human CEACAMs. *Nat Microbiol* 2:16189

104. Jefferson KK. 2004. What drives bacteria to produce a biofilm? *FEMS Microbiol Lett* 236:163-73

105. Jones HC, Roth IL, Sanders WM, 3rd. 1969. Electron microscopic study of a slime layer. *J Bacteriol* 99:316-25

106. Jones KR, Whitmire JM, Merrell DS. 2010. A Tale of Two Toxins: *Helicobacter pylori* CagA and VacA Modulate Host Pathways that Impact Disease. *Front Microbiol* 1:115

107. Jung SW, Sugimoto M, Graham DY, Yamaoka Y. 2009. *homB* status of *Helicobacter pylori* as a novel marker to distinguish gastric cancer from duodenal ulcer. *J Clin Microbiol* 47:3241-5

108. Kang J, Jones KR, Jang S, Olsen CH, Yoo YJ, et al. 2012. The geographic origin of *Helicobacter pylori* influences the association of the *homB* gene with gastric cancer. *J Clin Microbiol* 50:1082-5

109. Kawai M, Furuta Y, Yahara K, Tsuru T, Oshima K, et al. 2011. Evolution in an oncogenic bacterial species with extreme genome plasticity: *Helicobacter pylori* East Asian genomes. *BMC Microbiol* 11:104

110. Kim A, Servetas SL, Kang J, Kim J, Jang S, et al. 2015. *Helicobacter pylori bab* Paralog Distribution and Association with *cagA*, *vacA*, and *homA*/*B* Genotypes in American and South Korean Clinical Isolates. *PLoS One* 10:e0137078

111. Koebnik R, Locher KP, Van Gelder P. 2000. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol Microbiol* 37:239-53

112. Koniger V, Holsten L, Harrison U, Busch B, Loell E, et al. 2016. *Helicobacter pylori* exploits human CEACAMs via HopQ for adherence and translocation of CagA. *Nat Microbiol* 2:16188

113. Kusters JG, van Vliet AH, Kuipers EJ. 2006. Pathogenesis of *Helicobacter pylori* infection. *Clin Microbiol Rev* 19:449-90

114. Lee HW, Choe YH, Kim DK, Jung SY, Lee NG. 2004. Proteomic analysis of a ferric uptake regulator mutant of *Helicobacter pylori*: regulation of *Helicobacter pylori* gene expression by ferric uptake regulator and iron. *Proteomics* 4:2014-27

115. Li H, Liao T, Debowski AW, Tang H, Nilsson HO, et al. 2016. Lipopolysaccharide Structure and Biosynthesis in *Helicobacter pylori*. *Helicobacter*. [Epub ahead of print]

116. Linke S, Lenz J, Gemein S, Exner M, Gebel J. 2010. Detection of *Helicobacter pylori* in biofilms by real-time PCR. *Int J Hyg Environ Health* 213:176-82

117. Linz B, Balloux F, Moodley Y, Manica A, Liu H, et al. 2007. An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* 445:915-8

118. Linz B, Windsor HM, McGraw JJ, Hansen LM, Gajewski JP, et al. 2014. A mutation burst during the acute phase of *Helicobacter pylori* infection in humans and rhesus macaques. *Nat Commun* 5:4165

119. Liu H, Fero JB, Mendez M, Carpenter BM, Servetas SL, et al. 2015. Analysis of a single *Helicobacter pylori* strain over a 10-year period in a primate model. *Int J Med Microbiol* 305:392-403

120. Loh JT, Cover TL. 2006. Requirement of histidine kinases HP0165 and HP1364 for acid resistance in *Helicobacter pylori*. *Infect Immun* 74:3052-9

121. Loh JT, Gupta SS, Friedman DB, Krezel AM, Cover TL. 2010. Analysis of protein expression regulated by the *Helicobacter pylori* ArsRS two-component signal transduction system. *J Bacteriol* 192:2034-43

122. Loh JT, Torres VJ, Algood HM, McClain MS, Cover TL. 2008. *Helicobacter pylori* HopQ outer membrane protein attenuates bacterial adherence to gastric epithelial cells. *FEMS Microbiol Lett* 289:53-8

123. Lu H, Wu JY, Beswick EJ, Ohno T, Odenbreit S, et al. 2007. Functional and intracellular signaling differences associated with the *Helicobacter pylori* AlpAB adhesin from Western and East Asian strains. *J Biol Chem* 282:6242-54

124. Lu Y, Redlinger TE, Avitia R, Galindo A, Goodman K. 2002. Isolation and genotyping of *Helicobacter pylori* from untreated municipal wastewater. *Appl Environ Microbiol* 68:1436-9

125. Mack WN, Mack JP, Ackerson AO. 1975. Microbial film development in a trickling filter. *Microb Ecol* 2:215-26

126. Mackay WG, Gribbon LT, Barer MR, Reid DC. 1998. Biofilms in drinking water systems: a possible reservoir for *Helicobacter pylori*. *J Appl Microbiol* 85 Suppl 1:52S-9S

127. Mahdavi J, Sonden B, Hurtig M, Olfat FO, Forsberg L, et al. 2002. *Helicobacter pylori* SabA adhesin in persistent infection and chronic inflammation. *Science* 297:573-8

128. Maixner F, Krause-Kyora B, Turaev D, Herbig A, Hoopmann MR, et al. 2016. The 5300-year-old *Helicobacter pylori* genome of the Iceman. *Science* 351:162-5

129. Marcus EA, Moshfegh AP, Sachs G, Scott DR. 2005. The periplasmic alpha-carbonic anhydrase activity of *Helicobacter pylori* is essential for acid acclimation. *J Bacteriol* 187:729-38

130. Marcus EA, Sachs G, Wen Y, Scott DR. 2016. Phosphorylation-dependent and Phosphorylation-independent Regulation of *Helicobacter pylori* Acid Acclimation by the ArsRS Two-component System. *Helicobacter* 21:69-81

131. Marshall BJ. 2001. One Hundred Years of Discovery and Rediscovery of *Helicobacter pylori* and Its Association with Peptic Ulcer Disease. In *Helicobacter pylori: Physiology and Genetics*, ed. HLT Mobley, GL Mendz, SL Hazell. Washington (DC). Number of.

132. Marshall BJ, Armstrong JA, McGechie DB, Glancy RJ. 1985. Attempt to fulfil Koch's postulates for pyloric Campylobacter. *Med J Aust* 142:436-9

133. Marshall BJ, Warren JR. 1984. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* 1:1311-5

134. McDaniel TK, Dewalt KC, Salama NR, Falkow S. 2001. New approaches for validation of lethal phenotypes and genetic reversion in *Helicobacter pylori*. *Helicobacter* 6:15-23

135. McGowan CC, Necheva A, Thompson SA, Cover TL, Blaser MJ. 1998. Acid-induced expression of an LPS-associated gene in *Helicobacter pylori*. *Mol Microbiol* 30:19-31

136. McNulty CA, Watson DM. 1984. Spiral bacteria of the gastric antrum. *Lancet* 1:1068-9

137. Megraud F, Lehours P, Vale FF. 2016. The history of *Helicobacter pylori*: from phylogeography to paleomicrobiology. *Clin Microbiol Infect*

138. Menard AC, R.; Breurec, S.; Megraud, F.; Oleastro, M. High worldwid conservation of a *Helicobacter pylori* outer membrane protein gene, *homD*. *Proc. CHRO, Vancouver, Canada, 2011*

139. Merrell DS, Goodrich ML, Otto G, Tompkins LS, Falkow S. 2003. pH-regulated gene expression of the gastric pathogen *Helicobacter pylori*. *Infect Immun* 71:3529-39

140. Merrell DS, Thompson LJ, Kim CC, Mitchell H, Tompkins LS, et al. 2003. Growth phase-dependent response of *Helicobacter pylori* to iron starvation. *Infect Immun* 71:6510-25

141. Miles S, Piazuelo MB, Semino-Mora C, Washington MK, Dubois A, et al. 2010. Detailed in vivo analysis of the role of *Helicobacter pylori* Fur in colonization and disease. *Infect Immun* 78:3073-82

142. Mitchell HM. 2001. Epidemiology of Infection. In *Helicobacter pylori: Physiology and Genetics*, ed. HLT Mobley, GL Mendz, SL Hazell. Washington (DC). Number of.

143. Montano V, Didelot X, Foll M, Linz B, Reinhardt R, et al. 2015. Worldwide Population Structure, Long-Term Demography, and Local Adaptation of *Helicobacter pylori*. *Genetics* 200:947-63

144. Moodley Y, Linz B. 2009. *Helicobacter pylori* Sequences Reflect Past Human Migrations. *Genome Dyn* 6:62-74

145. Moodley Y, Linz B, Bond RP, Nieuwoudt M, Soodyall H, et al. 2012. Age of the association between *Helicobacter pylori* and man. *PLoS Pathog* 8:e1002693

146. Moreno Y, Ferrus MA. 2012. Specific detection of cultivable *Helicobacter pylori* cells from wastewater treatment plants. *Helicobacter* 17:327-32

147. Ng CG, Loke MF, Goh KL, Vadivelu J, Ho B. 2017. Biofilm formation enhances *Helicobacter pylori* survivability in vegetables. *Food Microbiology* 62:68-76

148. O'Toole G, Kaplan HB, Kolter R. 2000. Biofilm formation as microbial development. *Annu Rev Microbiol* 54:49-79

149. O'Toole PW, Clyne M. 2001. Cell Envelope. In *Helicobacter pylori: Physiology and Genetics*, ed. HLT Mobley, GL Mendz, SL Hazell. Washington (DC). Number of.

150. Odenbreit S, Swoboda K, Barwig I, Ruhl S, Boren T, et al. 2009. Outer membrane protein expression profile in *Helicobacter pylori* clinical isolates. *Infect Immun* 77:3782-90

151. Odenbreit S, Till M, Hofreuter D, Faller G, Haas R. 1999. Genetic and functional characterization of the *alpAB* gene locus essential for the adhesion of *Helicobacter pylori* to human gastric tissue. *Mol Microbiol* 31:1537-48

152. Oleastro M, Cordeiro R, Ferrand J, Nunes B, Lehours P, et al. 2008. Evaluation of the clinical significance of *homB*, a novel candidate marker of *Helicobacter pylori* strains associated with peptic ulcer disease. *J Infect Dis* 198:1379-87

153. Oleastro M, Cordeiro R, Menard A, Gomes JP. 2010. Allelic diversity among *Helicobacter pylori* outer membrane protein genes *homB* and *homA* generated by recombination. *J Bacteriol* 192:3961-8

154. Oleastro M, Cordeiro R, Menard A, Yamaoka Y, Queiroz D, et al. 2009. Allelic diversity and phylogeny of *homB*, a novel co-virulence marker of *Helicobacter pylori*. *BMC Microbiol* 9:248

155. Oleastro M, Cordeiro R, Yamaoka Y, Queiroz D, Megraud F, et al. 2009. Disease association with two *Helicobacter pylori* duplicate outer membrane protein genes, *homB* and *homA*. *Gut Pathog* 1:12

156. Oleastro M, Menard A. 2013. The Role of *Helicobacter pylori* Outer Membrane Proteins in Adherence and Pathogenesis. *Biology (Basel)* 2:1110-34

157. Oleastro M, Monteiro L, Lehours P, Megraud F, Menard A. 2006. Identification of markers for *Helicobacter pylori* strains isolated from children with peptic ulcer disease by suppressive subtractive hybridization. *Infect Immun* 74:4064-74

158. Panthel K, Dietz P, Haas R, Beier D. 2003. Two-component systems of *Helicobacter pylori* contribute to virulence in a mouse infection model. *Infect Immun* 71:5381-5

159. Park SR, Mackay WG, Reid DC. 2001. *Helicobacter* sp. recovered from drinking water biofilm sampled from a water distribution system. *Water Res* 35:1624-6

160. Parkhill J, Wren BW, Mungall K, Ketley JM, Churcher C, et al. 2000. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* 403:665-8

161. Percival SL, Thomas JG. 2009. Transmission of *Helicobacter pylori* and the role of water and biofilms. *J Water Health* 7:469-77

162. Pflock M, Dietz P, Schar J, Beier D. 2004. Genetic evidence for histidine kinase HP165 being an acid sensor of *Helicobacter pylori*. *FEMS Microbiol Lett* 234:51-61

163. Pflock M, Finsterer N, Joseph B, Mollenkopf H, Meyer TF, Beier D. 2006. Characterization of the ArsRS regulon of *Helicobacter pylori*, involved in acid adaptation. *J Bacteriol* 188:3449-62

164. Pflock M, Kennard S, Finsterer N, Beier D. 2006. Acid-responsive gene regulation in the human pathogen *Helicobacter pylori*. *J Biotechnol* 126:52-60

165. Plummer M, de Martel C, Vignat J, Ferlay J, Bray F, Franceschi S. 2016. Global burden of cancers attributable to infections in 2012: a synthetic analysis. *Lancet Glob Health* 4:e609-16

166. Pride DT, Meinersmann RJ, Blaser MJ. 2001. Allelic Variation within *Helicobacter pylori babA* and *babB*. *Infect Immun* 69:1160-71

167. Qu W, Zhou Y, Shao C, Sun Y, Zhang Q, et al. 2009. *Helicobacter pylori* proteins response to nitric oxide stress. *J Microbiol* 47:486-93

168. Ramsay PT, Carr A. 2011. Gastric acid and digestive physiology. *Surg Clin North Am* 91:977-82

169. Ratledge C, Dover LG. 2000. Iron metabolism in pathogenic bacteria. *Annu Rev Microbiol* 54:881-941

170. Roncarati D, Pelliciari S, Doniselli N, Maggi S, Vannini A, et al. 2016. Metal-responsive promoter DNA compaction by the ferric uptake regulator. *Nat Commun* 7:12593

171. Sachs G, Weeks DL, Wen Y, Marcus EA, Scott DR, Melchers K. 2005. Acid acclimation by *Helicobacter pylori*. *Physiology (Bethesda)* 20:429-38

172. Salama N, Guillemin K, McDaniel TK, Sherlock G, Tompkins L, Falkow S. 2000. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc Natl Acad Sci U S A* 97:14668-73

173. Salaun L, Linz B, Suerbaum S, Saunders NJ. 2004. The diversity within an expanded and redefined repertoire of phase-variable genes in *Helicobacter pylori*. *Microbiology* 150:817-30

174. Saravanan NS, S.T. 2015. Biofilm: A Sequential Analysis. *International Journal of Life Sciences Research* 3:90-8

175. Scarlato V, Delany I, Spohn G, Beier D. 2001. Regulation of transcription in *Helicobacter pylori*: simple systems or complex circuits? *Int J Med Microbiol* 291:107-17

176. Schar J, Sickmann A, Beier D. 2005. Phosphorylation-independent activity of atypical response regulators of *Helicobacter pylori*. *J Bacteriol* 187:3100-9

177. Schooling SR, Beveridge TJ. 2006. Membrane vesicles: an overlooked component of the matrices of biofilms. *J Bacteriol* 188:5945-57

178. Schreiber S, Konradt M, Groll C, Scheid P, Hanauer G, et al. 2004. The spatial orientation of *Helicobacter pylori* in the gastric mucus. *Proc Natl Acad Sci U S A* 101:5024-9

179. Schulz GE. 2002. The structure of bacterial outer membrane proteins. *Biochim Biophys Acta* 1565:308-17

180. Scott DR, Marcus EA, Weeks DL, Sachs G. 2002. Mechanisms of acid resistance due to the urease system of *Helicobacter pylori*. *Gastroenterology* 123:187-95

181. Senkovich OA, Yin J, Ekshyyan V, Conant C, Traylor J, et al. 2011. *Helicobacter pylori* AlpA and AlpB bind host laminin and influence gastric inflammation in gerbils. *Infect Immun* 79:3106-16

182. Servetas SL, Carpenter BM, Haley KP, Gilbreath JJ, Gaddy JA, Merrell DS. 2016. Characterization of Key *Helicobacter pylori* Regulators Identifies a Role for ArsRS in Biofilm Formation. *J Bacteriol* 198:2536-48

183. Shao C, Sun Y, Wang N, Yu H, Zhou Y, et al. 2013. Changes of proteome components of *Helicobacter pylori* biofilms induced by serum starvation. *Mol Med Rep* 8:1761-6

184. Sheu SM, Sheu BS, Chiang WC, Kao CY, Wu HM, et al. 2012. *H. pylori* clinical isolates have diverse *babAB* genotype distributions over different topographic sites of stomach with correlation to clinical disease outcomes. *BMC Microbiol* 12:89

185. Skindersoe ME, Rasmussen L, Andersen LP, Krogfelt KA. 2015. A Novel Assay for Easy and Rapid Quantification of *Helicobacter pylori* Adhesion. *Helicobacter* 20:199-205

186. Smith JL. 2003. The role of gastric acid in preventing foodborne disease and how bacteria overcome acid conditions. *J Food Prot* 66:1292-303

187. Solnick JV, 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

188. Srivastava SY, de Silva AM. 2009. Characterization of *Borrelia burgdorferi* aggregates. *Vector Borne Zoonotic Dis* 9:323-9

189. Stark RM, Gerwig GJ, Pitman RS, Potts LF, Williams NA, et al. 1999. Biofilm formation by *Helicobacter pylori*. *Lett Appl Microbiol* 28:121-6

190. Steer HW, Colin-Jones DG. 1975. Mucosal changes in gastric ulceration and their response to carbenoxolone sodium. *Gut* 16:590-7

191. Stein M, Rappuoli R, Covacci A. 2001. The *cag* Pathogenicity Island. In *Helicobacter pylori: Physiology and Genetics*, ed. HLT Mobley, GL Mendz, SL Hazell. Washington (DC). 345-354.

192. Styer CM, Hansen LM, Cooke CL, Gundersen AM, Choi SS, et al. 2010. Expression of the BabA adhesin during experimental infection with *Helicobacter pylori*. *Infect Immun* 78:1593-600

193. Suerbaum S. 2000. Genetic variability within *Helicobacter pylori*. *Int J Med Microbiol* 290:175-81

194. Suerbaum S, Achtman M. 2001. Population Genetics. In *Helicobacter pylori: Physiology and Genetics*, ed. HLT Mobley, GL Mendz, SL Hazell. Washington (DC). Number of.

195. Svensson SL, Davis LM, MacKichan JK, Allan BJ, Pajaniappan M, et al. 2009. The CprS sensor kinase of the zoonotic pathogen *Campylobacter jejuni* influences biofilm formation and is required for optimal chick colonization. *Mol Microbiol* 71:253-72

196. Svensson SL, Hyunh S, Parker CT, Gaynor EC. 2015. The *Campylobacter jejuni* CprRS two-component regulatory system regulates aspects of the cell envelope. *Mol Microbiol* 96:189-209

197. Svensson SL, Pryjma M, Gaynor EC. 2014. Flagella-mediated adhesion and extracellular DNA release contribute to biofilm formation and stress tolerance of *Campylobacter jejuni*. *PLoS One* 9:e106063

198. Talarico S, Whitefield SE, Fero J, Haas R, Salama NR. 2012. Regulation of *Helicobacter pylori* adherence by gene conversion. *Mol Microbiol* 84:1050-61

199. Talebi Bezmin Abadi A, Rafiei A, Ajami A, Hosseini V, Taghvaei T, et al. 2011. *Helicobacter pylori homB*, but not *cagA*, is associated with gastric cancer in Iran. *J Clin Microbiol* 49:3191-7

200. Talebi Bezmin Abadi A, Taghvaei T, Mohabbati Mobarez A, Vaira G, Vaira D. 2013. High correlation of *babA* 2-positive strains of *Helicobacter pylori* with the presence of gastric cancer. *Intern Emerg Med* 8:497-501

201. Thompson LJ, Merrell DS, Neilan BA, Mitchell H, Lee A, Falkow S. 2003. Gene expression profiling of *Helicobacter pylori* reveals a growth-phase-dependent switch in virulence gene expression. *Infect Immun* 71:2643-55

202. Thung I, Aramin H, Vavinskaya V, Gupta S, Park JY, et al. 2016. Review article: the global emergence of *Helicobacter pylori* antibiotic resistance. *Aliment Pharmacol Ther* 43:514-33

203. Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, et al. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388:539-47

204. van Vliet AH, Ernst FD, Kusters JG. 2004. NikR-mediated regulation of *Helicobacter pylori* acid adaptation. *Trends Microbiol* 12:489-94

205. van Vliet AH, Kuipers EJ, Stoof J, Poppelaars SW, Kusters JG. 2004. Acid-responsive gene induction of ammonia-producing enzymes in *Helicobacter pylori* is mediated via a metal-responsive repressor cascade. *Infect Immun* 72:766-73

206. Vianna JS, Ramis IB, Ramos DF, A VONG, Silva PE. 2016. Drug Resistance in *Helicobacter pylori*. *Arq Gastroenterol* 53:215-23

207. Weinberg ED. 1975. Nutritional immunity. Host's attempt to withold iron from microbial invaders. *JAMA* 231:39-41

208. Wen Y, Feng J, Scott DR, Marcus EA, Sachs G. 2006. Involvement of the HP0165-HP0166 two-component system in expression of some acidic-pH-upregulated genes of *Helicobacter pylori*. *J Bacteriol* 188:1750-61

209. Wen Y, Feng J, Scott DR, Marcus EA, Sachs G. 2007. The HP0165-HP0166 two-component system (ArsRS) regulates acid-induced expression of HP1186 alpha-carbonic anhydrase in *Helicobacter pylori* by activating the pH-dependent promoter. *J Bacteriol* 189:2426-34

210. Wen Y, Marcus EA, Matrubutham U, Gleeson MA, Scott DR, Sachs G. 2003. Acid-adaptive genes of *Helicobacter pylori*. *Infect Immun* 71:5921-39

211. Williams JC, McInnis KA, Testerman TL. 2008. Adherence of *Helicobacter pylori* to abiotic surfaces is influenced by serum. *Appl Environ Microbiol* 74:1255-8

212. Wimley WC. 2003. The versatile beta-barrel membrane protein. *Curr Opin Struct Biol* 13:404-11

213. Yahaghi E, Khamesipour F, Mashayekhi F, Safarpoor Dehkordi F, Sakhaei MH, et al. 2014. *Helicobacter pylori* in vegetables and salads: genotyping and antimicrobial resistance properties. *Biomed Res Int* 2014:757941

214. Yamaoka Y, Kita M, Kodama T, Imamura S, Ohno T, et al. 2002. *Helicobacter pylori* infection in mice: Role of outer membrane proteins in colonization and inflammation. *Gastroenterology* 123:1992-2004

215. Yamaoka Y, Ojo O, Fujimoto S, Odenbreit S, Haas R, et al. 2006. *Helicobacter pylori* outer membrane proteins and gastroduodenal disease. *Gut* 55:775-81

216. Yang FL, Hassanbhai AM, Chen HY, Huang ZY, Lin TL, et al. 2011. Proteomannans in biofilm of *Helicobacter pylori* ATCC 43504. *Helicobacter* 16:89-98

217. Yonezawa H, Osaki T, Hanawa T, Kurata S, Ochiai K, Kamiya S. 2013. Impact of *Helicobacter pylori* biofilm formation on clarithromycin susceptibility and generation of resistance mutations. *PLoS One* 8:e73301

218. Yonezawa H, Osaki T, Kurata S, Fukuda M, Kawakami H, et al. 2009. Outer membrane vesicles of *Helicobacter pylori* TK1402 are involved in biofilm formation. *BMC Microbiol* 9:197

219. Yonezawa H, Osaki T, Kurata S, Zaman C, Hanawa T, Kamiya S. 2010. Assessment of *in vitro* biofilm formation by *Helicobacter pylori*. *J Gastroenterol Hepatol* 25 Suppl 1:S90-4

220. Yonezawa H, Osaki T, Woo T, Kurata S, Zaman C, et al. 2011. Analysis of outer membrane vesicle protein involved in biofilm formation of *Helicobacter pylori*. *Anaerobe* 17:388-90

221. Zaric S, Bojic B, Jankovic L, Dapcevic B, Popovic B, et al. 2009. Periodontal therapy improves gastric *Helicobacter pylori* eradication. *J Dent Res* 88:946-50

222. Zhang M. 2015. High antibiotic resistance rate: A difficult issue for *Helicobacter pylori* eradication treatment. *World J Gastroenterol* 21:13432-7

223. Zhang XS, Tegtmeyer N, Traube L, Jindal S, Perez-Perez G, et al. 2015. A specific A/T polymorphism in Western tyrosine phosphorylation B-motifs

regulates *Helicobacter pylori* CagA epithelial cell interactions. *PLoS Pathog* 11:e1004621

224. Zobell CE, Allen EC. 1935. The Significance of Marine Bacteria in the Fouling of Submerged Surfaces. *J Bacteriol* 29:239-51

# Appendix A: *Helicobacter pylori bab* Paralog Distribution and Association with *cagA*, *vacA*, and *homA/B* Genotypes in American and South Korean Clinical Isolates

Published as: **Aeryun Kim¶, Stephanie L. Servetas¶, Jieun Kang, Jinmoon Kim, Sungil Jang, Ho Jin Cha, Wan Jin Lee, June Kim, Judith Romero-Gallo, Richard M. Peek Jr., D. Scott Merrell, and Jeong-Heon Cha.** 2015. *Helicobacter pylori bab* paralog distribution and association with *cagA, vacA,* and *homA/B* genotypes in American and South Korean clinical isolates. *PLos ONE*. **10**(8): e0137078.

¶These authors contributed equally to this work

The work presented in this chapter is the sole work of S.L. Servetas with the following exceptions: A. Kim, J. Kang, J. Kim, S. Jang, H. J. Cha, W. J. Lee, J. Kim J. H. Cha conducted sequencing analysis and genotyping; J. Romero-Gallo, R. M. Peek Jr. provided AH clinical isolates.

## ABSTRACT

*Helicobacter pylori* genetic variation is a crucial component of colonization and persistence within the inhospitable niche of the gastric mucosa. As such, numerous *H. pylori* genes have been shown to vary in terms of presence and genomic location within this pathogen. Among the variable factors, the Bab family of outer membrane proteins (OMPs) has been shown to differ within subsets of strains. To better understand genetic variation among the *bab* genes and to determine whether this variation differed among

isolates obtained from different geographic locations, we characterized the distribution of the Bab family members in 80 American *H. pylori* clinical isolates (AH) and 80 South Korean *H. pylori* clinical isolates (KH). Overall, we identified 23 different *bab* genotypes (19 in AH and 11 in KH), but only 5 occurred in greater than 5 isolates. Regardless of strain origin, a strain in which locus A and locus B were both occupied by a *bab* gene was the most common (85%); locus C was only occupied in those isolates that carried *bab* paralog at locus A and B. While the *babA*/*babB*/- genotype predominated in the KH (78.8%), no single genotype could account for greater than 40% in the AH collection. In addition to basic genotyping, we also identified associations between *bab* genotype and well known virulence factors, *cagA* and *vacA*. Specifically, significant associations between *babA* at locus A and the *cagA* EPIYA-ABD motif ($P<0.0001$) and the *vacA* s1/i1/m1 allele ($P<0.0001$) were identified. Log-linear modeling further revealed a three-way association between *bab* carried at locus A, *vacA*, and number of OMPs from the HOM family ($P<0.002$). *En masse* this study provides a detailed characterization of the *bab* genotypes from two distinct populations. Our analysis suggests greater variability in the AH, perhaps due to adaptation to a more diverse host population. Furthermore, when considering the presence or absence of both the *bab* and *homA/B* paralogs at their given loci and the *vacA* genotype, an association was observed. Our results highlight the multifactorial nature of *H. pylori* mediated disease and the importance of considering how the specific combinations of *H. pylori* virulence genes and their multiple interactions with the host will collectively impact disease progression.

## INTRODUCTION

*Helicobacter pylori* (*H. pylori*) is a successful pathogen, colonizing the gastric mucosa of over 50% of the world's population (2; 43). This pathogen has gained notoriety for its ability to colonize the inhospitable niche of the stomach and to cause gastric diseases (2; 43). *H. pylori* cause persistent, potentially lifelong infections; however, only about 20% of individuals infected will develop symptomatic infection. Although clinical manifestations occur only in a subset of infected individuals, these can be severe and include peptic ulcers and gastric cancer. Rates of *H. pylori* infection and gastric cancer mortality vary geographically. East Asian countries such as China, Japan, and South Korea have high rates of *H. pylori* infection, and these populations also have some of the highest rates of gastric cancer (17; 41; 45; 47).

Environmental, host, and bacterial factors are known to influence *H. pylori*-associated disease outcome and also vary geographically (10). Interestingly, *H. pylori* is a bacterial species that shows exceptionally high rates of genetic variability and intra-species diversity. Indeed, these polymorphisms have been used to determine the geographic origin of *H. pylori* isolates (26; 30). Furthermore, it is highly probable that these genetic differences influence virulence. Previous studies have shown that severe disease outcomes such as gastric cancer and ulcers are associated with polymorphism in *H. pylori* virulence factors such as the cytotoxin-associated gene A (*cagA*) and vacuolating cytotoxin (*vacA*) (17; 20; 21; 47; 48). In addition, disease outcome has been associated with several outer membrane proteins (OMPs). Specifically, OMPs such as AlpA and B (adherence-associated lipoprotein A and B), BabA (blood group antigen binding adhesin), HomB (*Helicobacter* outer membrane B), HopZ (*Helicobacter* outer membrane protein Z), OipA (outer membrane inflammatory protein A) and SabA (sialic

acid binding adhesion) are all associated with variable disease outcomes (14; 25; 28; 44; 49; 50). In comparison to the polymorphisms within *cagA* and *vacA*, genetic variability within the OMP families is based upon presence and absence of different closely related paralogs. For example, the *bab*-family of genes is made up of three paralogs *babA*, *babB* and *babC*, which can be located at three different *H. pylori* chromosomal loci referred to as locus A, B and C (7; 18; 38). Recently, we presented detailed epidemiological studies of *cagA*, *vacA,* and *homB* from a collection of 260 *H. pylori* clinical isolates from South Korea (20-22; 25). Our studies showed a significant association between infection with *H. pylori* strains carrying the EPIYA-ABD *cagA* genotype and the development of gastric cancer (22). Moreover, the majority of *H. pylori* isolates from the South Korean population encoded the most virulent toxins, CagA EPIYA-ABD motif and VacA s1/i1/m1 allele (20). While, we found no association between the presence of *homB* and the progression to severe gastric disease (25), other groups working with isolates from different geographic regions have shown that the presence of *homB* was associated with development of peptic ulcer disease in children and young adults (33; 36)and with gastric cancer development and the presence of *cagA* (23). The variations in these results suggest that the impact of the *homB* gene on disease is geographically dependent (25).

Our studies and others have revealed that certain polymorphisms are more prevalent within specific populations (3; 6; 20-22; 24; 25; 51). Thus, it is possible that the combination of virulence factor polymorphisms present within a population may influence the rate of severe disease outcome. Therefore, the polymorphism of each virulence factor must be evaluated within various populations to help understand the

relationship between these genotypes and their associations with gastric diseases including cancer.

The *bab* gene family has been shown to have a positive association with gastric cancer and duodenal ulcers (1; 16; 40; 47). BabA, one member of this family, was the first identified adhesin in *H. pylori* and mediates the binding of bacteria to the fucosylated blood group O antigen Lewis b ($Le^b$), which is highly expressed in gastrointestinal epithelial cells (11; 19). Similar to several *H. pylori* OMPs, BabA has two closely related paralogous, BabB (HopT) and BabC (HopU); however, the functions of BabB and BabC are unknown. These OMPs, particularly BabA and BabB, have nearly identical N- and C-terminal domains but a divergent middle region, suggesting that the middle variable region of BabA is most likely responsible for the specific adhesin function (8; 19; 37; 38). Interestingly, it has been shown that BabB expression is modulated by changes in the number of cytosine-thymidine (CT) dinucleotide repeats in the 5′ region of the *babB* gene resulting in slipped-strand mispairing (5; 42; 46). Changes in the number of CT repeats results in a frame shift ; 'IN'-frame *bab* genes are translated into a full functional Bab or 'OUT'-of-frame *bab* genes into a premature partial Bab (5; 42; 46). Furthermore, *bab* expression can be modulated by gene conversion between the loci, most likely by intragenomic homologous recombination (35; 37). This gene conversion results in chimeric *bab* genes where the CT repeats of the *babB* 5′ region become associated with *babA* or *babC* subjecting these genes to phase variation. These recombination events may play an important role in regulation of adherence. For example, *babBA* chimera results in the regulation of *babA* by changes in CT repeats that lead to changes in the binding to $Le^b$ on gastric epithelial cells (9).

To gain further insight into genetic diversity of *H. pylori* isolates within the South Korean population, we analyzed the *bab* gene family. In addition, we expanded our analysis to include clinical isolates obtained from an American hospital to allow for strain comparisons between different geographic regions.

## MATERIALS AND METHODS

### Bacterial Strains and Culture Conditions

A total of 160 *H. pylori* clinical isolates were used in our analysis. The 80 South Korean *H. pylori* clinical isolates (KH) were a subset of a collection of South Korean strains previously characterized for distribution of *cagA* EPIYA polymorphism, *vacA* s/i/m polymorphism, and *homA/B* paralog (20-22; 25), while the other 80 isolates were obtained from patients at Vanderbilt University Medical Center, Nashville TN, USA. Written informed consent was received from each patient, and the protocol was approved by the Vanderbilt University and the Nashville Department of Veterans Affairs Institutional Review Board (IRB# 5190). Gastric biopsies were harvested from individuals at the Veterans Affairs Medical Center in Nashville undergoing upper endoscopy, and used for bacterial culture. Isolates from biopsies were confirmed to be *H. pylori* by positive urease, catalase, and oxidase tests, and typical appearance on Gram stain. These strains are referred to as American *H. pylori* clinical isolates (AH). For 80 AH, polymorphisms in *cagA* and *vacA*, and *homB* paralog were determined using the strategies previously developed for characterization of KH (20-22; 25). The 160 KH and AH are described in detail in Table 6 and Table 7. All *H. pylori* clinical isolates were

**Table 6 (S1-AH). *cagA*, *vacA* and *homA/B* genotyping of 80 American *H. pylori* clinical isolates**

| Strain Name | Strain Designation | Disease | Race | Sex | Age | CagA EPIYA | VacA s/i/m | Hom locusA/B |
|---|---|---|---|---|---|---|---|---|
| **B99** | 868 | GU | B | M | 52 | ABCC | 1/1/1 | *homB/homA* |
| **B101** | 869 | DU | W | M | 33 | ABC | 1/1/1 | *homA/-* |
| **B105** | 870 | DU | B | M | 44 | ABC | 1/1/1 | *homB/homA* |
| **B107** | 871 | DU | W | M | 60 | ABCC | 2/2/2 | *homA/-* |
| **B108** | 872 | G | W | M | 56 | ABC | 2/2/2 | *homA/-* |
| **B128** | 873 | GU | W | M | 48 | ABC | 1/2/2 | *homB/homB* |
| **B129** | 874 | DU | W | M | 50 | ABC | 1/1/1 | *homA/homA* |
| **B134** | 875 | GU | B | M | 58 | ABC | 1/1/1 | *homA/homA* |
| **B138** | 876 | CA | W | M | 48 | ABC | 1/1/1 | *homB/homB* |
| **B140** | 877 | E | W | M | 31 | ABC | 1/1/1 | *homA/-* |
| **B148** | 879 | DU | W | M | 59 | ABCC | 1/1/1 | *homB/homB* |
| **B151** | 880 | G | W | M | 84 | ABC | 1/1/1 | *homB/homB* |
| **B173** | 881 | E | B | M | 64 | ABCC | 2/2/2 | *homA/-* |
| **B177** | 882 | E | W | M | 72 | ABC | 1/1/1 | *homA/homA* |
| **B179** | 883 | G | B | M | 52 | ABC | 1/1/1 | *homB/homB* |
| **B181** | 884 | BE | W | M | 40 | ABCC | 2/2/2 | *homA/-* |
| **B186** | 885 | GU | B | M | 44 | ABC | 1/1/1 | *homB/homA* |
| **B211** | 886 | G | W | M | 49 | ABCC | 2/2/2 | *homA/-* |
| **B213** | 887 | E | B | M | 66 | ABC | 1/1/1 | *homB/homB* |
| **B218** | 888 | G | W | M | 80 | ABC | 1/2/2 | *homA/homA* |
| **B221** | 889 | E | W | F | 50 | ABCC | 2/2/2 | *homB/-* |
| **B235** | 891 | G | B | M | 55 | ABCC | 1/1/1 | *homB/homB* |
| **B240** | 892 | DU | B | M | 48 | ABCC | 2/2/2 | *homA/-* |
| **B289** | 893 | G | B | M | 50 | ABC | 1/1/2 | *homB/homB* |
| **J58** | 895 | G | W | M | 52 | ABCC | 2/2/2 | *homB/-* |
| **J68** | 896 | DU | W | M | 33 | ABCC | 2/2/2 | *homA/-* |
| **J99** | 897 | DU | W | M | 48 | ABC | 1/1/1 | *homB/homB* |
| **J104** | 898 | DU | W | M | 65 | ABC | 1/1/1 | *homA/homA* |
| **J117** | 899 | DU | W | M | 59 | ABCC | 1/2/2 | *homB/-* |
| **J133** | 900 | DU | B | M | 61 | ABC | 1/1/2 | *homA/homA* |
| **J154** | 901 | E | W | M | 61 | ABCC | 2/2/2 | *homA/-* |
| **J166** | 902 | DU | W | M | 56 | ABC | 1/1/1 | $homA^{+/-}homA$ |
| **J178** | 903 | G | B | M | 61 | ABC | 1/1/1 | *homB/homA* |
| **J226** | 905 | DU | W | M | 31 | ABCC | 1/1/2 | *homB/homA* |
| **J238** | 906 | G | W | M | 56 | ABC | 1/1/1 | *ihom/ihom* |
| **J243** | 907 | GU | B | M | 73 | ABC | 1/1/1 | *homB/homB* |
| **J262** | 908 | E | W | M | 63 | ABCC | 2/2/2 | *homB/-* |
| **J300** | 909 | CA | W | M | 76 | ABCC | 1/1/1 | *homB/homA* |
| **J44** | 1303 | DU | W | M | 64 | ABC | 1/1/2 | *homB/-* |
| **J63** | 1304 | G | W | M | 52 | AB* | 2/2/2 | *homB/-* |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **J75** | 1305 | G | W | M | 61 | ABCC* | 2/2/2 | *homB/-* |
| **J150** | 1306 | G | W | M | 52 | ABCC | 2/2/2 | *homA/-* |
| **J170** | 1307 | GU | W | M | 56 | ABC | 1/1/1 | *homB/homB* |
| **J182** | 1308 | DU | W | M | 60 | ABC | 1/1/2 | *homA/homA* |
| **J195** | 1309 | G | W | M | 44 | ABCC | 2/2/2 | *ihom/ihom* |
| **J254** | 1312 | DU | W | M | 45 | ABC | 1/1/1 | *homB/-* |
| **J258** | 1313 | DU | W | M | 65 | ABC | 1/1/1 | *homA/homA* |
| **J291** | 1314 | G | B | M | 27 | ABC | 1/1/1 | *homB/homA* |
| **J314** | 1315 | BE | W | M | 46 | ABC | 1/1/1 | *homA/homA* |
| **B96** | 1316 | G | B | M | 31 | ABC | 1/1/1 | *homA/homA* |
| **B116A** | 1317 | G | W | M | 83 | ABC | 2/1/2 | *homA/homA* |
| **B125A** | 1319 | GU | W | M | 74 | ABC | 1/1/2 | *homB/-* |
| **B127A** | 1320 | GU | W | M | 68 | ABC* | 2/2/2 | *homA/-* |
| **B130A** | 1321 | GU | W | M | 62 | AB | 1/1/1 | *homA/homA* |
| **B136A** | 1322 | E | W | M | 66 | (A/B)C | 1/1/2 | *homA/-* |
| **B137A** | 1323 | G | W | M | 78 | ABC | 2/2/2 | *homA/-* |
| **B147** | 1326 | E | W | M | 58 | AB | 1/2/2 | *homA/-* |
| **B158A** | 1327 | BE | W | M | 55 | ABC | 1/2/2 | *homA/ihom* |
| **B166A** | 1328 | G | W | M | 61 | B* | 2/2/2 | *homA/-* |
| **B175A** | 1329 | E | W | M | 65 | ABC | 2/2/2 | *homB/-* |
| **B185A** | 1330 | DU | W | M | 65 | ABC | 1/1/1 | *homB/homB* |
| **B201A** | 1331 | E | W | M | 67 | (A/B)C* | 2/2/2 | *homA/-* |
| **B223A** | 1336 | GU | W | M | 75 | ABC | 1/1/1 | *homB/homB* |
| **B227C** | 1337 | E | W | M | 45 | ABC | 1/2/2 | *homA/ihom* |
| **B234A** | 1338 | G | W | M | 69 | BC* | 2/2/2 | *homB/-* |
| **B251A** | 1340 | G | W | M | 66 | ABC | 1/1/1 | *homB/homA* |
| **B253A** | 1341 | G | W | M | 71 | ABC | 1/2/2 | *homA/-* |
| **B255A** | 1343 | DU | W | M | 66 | ABCC | 1/1/2 | *homB/homA* |
| **B258A** | 1345 | BE | W | M | 52 | ABC | 1/1/2 | *homA/homB* |
| **B263A** | 1346 | G | W | M | 50 | ABCC | 2/2/2 | *homA/-* |
| **B266A** | 1348 | G | W | M | 59 | ABC | 1/1/1 | *homB/homB* |
| **B267A** | 1349 | DU | W | M | 76 | ABCC | 1/1/1 | *homB/homB* |
| **B268A** | 1350 | G | W | M | 74 | ABCC | 1/1/1 | *homB/homB* |
| **B269A** | 1351 | DU | W | M | 52 | ABCC* | 2/2/2 | *homA/-* |
| **B279A** | 1352 | DU | W | M | 51 | ABCC | 2/2/2 | *homA/-* |
| **B292A** | 1354 | DU | W | M | 52 | ABC | 1/1/2 | *homB/homB* |
| **B293C** | 1355 | DU | W | M | 54 | ABC | 1/1/1 | *homA/homB* |
| **B294A** | 1356 | E | B | M | 56 | ABC | 1/1/1 | *homA/ihom* |
| **B300A** | 1358 | DU | W | M | 56 | ABCC* | 2/2/2 | *homA/-* |
| **B301 A** | 1359 | G | W | M | 68 | ABCC | 2/2/2 | *homA/-* |

**Abbreviations of disease names are following; G: Gastritis, E : Esophogitis, GU: Gastric ulcer, DU: Duodenal Ulcer, BE: Barrett's Esophagus, CA: Gastric Cancer**

**B indicates Black, W; White**

***homA*$^{+/-}$ or *homB*$^{+/-}$ indicates that the amplicaton of the respective locus was unsuccesful. For example, *homB*$^{+/-}$ /*homB* indicates that homB gene is amplicated in locusB, but we don't know that *homB* exists in locusA.**

*** : 10 isolates were failed to get the sequence of CagA EPIYA motif region, so those genotypes were decided by comparing PCR results of other isolates or their own unclear pcr bands.**

****-B motif's proline is replaced with a serine, ESIYA, and is therefore, classified as "other "**

*****-ABABD second -B motif's proline is replaced with leucine, ELIYA, and is therefore, classified as "other"**

**N/A, not applicable**

**Table 7 (S1-KH) *cagA*, *vacA* and *homA/B* genotyping of 80 Korean *H. pylori* clinical isolates**

| Strain Name | Disease | Sex | Age | CagA EPIYA | VacA s/i/m | Hom locusA/B |
|---|---|---|---|---|---|---|
| **k3** | **CA** | **F** | **65** | ABD | 1/1/1 | - / *homB* |
| **k6** | **CA** | **F** | **45** | ABD | 1/1/1 | - / *homA* |
| **k10** | **CA** | **M** | **52** | ABD | 1/1/1 | - / *homB* |
| **k16** | **CA** | **M** | **48** | ABD | 1/1/1 | - / *homB* |
| **k17** | **CA** | **F** | **56** | ABD | 1/1/1 | - / *homB* |
| **k19** | **CA** | **F** | **86** | ABD | 1/1/1 | - / *homB* |
| **k21** | **CA** | **M** | **44** | ABD | 1/1/1 | - / *homB* |
| **k23** | **DU** | **M** | **47** | ABD | 1/1/1 | - / *homB* |
| **k24** | **DU** | **M** | **38** | ABD | 1/1/1 | * |
| **k25** | **DU** | **M** | **44** | ABD | 1/1/1 | - / *homB* |
| **k26** | **DU** | **M** | **20** | ABD | 1/1/1 | - / *homB* |
| **k27** | **DU** | **M** | **47** | ABD | 1/1/1 | - / *homA* |
| **k28** | **DU** | **M** | **28** | ABD | 1/1/1 | - / *homB* |
| **k34** | **DU** | **M** | **43** | ABD | 1/1/1 | - / *homB* |
| **k35** | **DU** | **F** | **56** | ABD | 1/1/2 | - / *homB* |
| **k36** | **DU** | **M** | **46** | ABD | 1/1/1 | - / *homB* |
| **k37** | **DU** | **M** | **61** | ABD | 1/2/2 | - / *homB* |
| **k41** | **DU** | **M** | **55** | ABD | 1/1/1 | - / *homA* |
| **k42** | **DU** | **F** | **48** | ABD | 1/1/2 | - / *homB* |
| **k43** | **DU** | **M** | **70** | ABD | 1/1/1 | - / *homB* |
| **k44** | **DU** | **M** | **42** | ABD | 1/1/1 | - / *homB* |
| **k45** | **DU** | **M** | **22** | ABD | 1/1/1 | - / *homB* |
| **k46** | **DU** | **F** | **61** | ABD | 1/1/1 | - / *homB* |
| **k47** | **DU** | **F** | **72** | ABD | 1/2/2 | *homA/-* |
| **k48** | **DU** | **F** | **41** | ABD | 1/1/1 | - / *homB* |
| **k57** | **G** | **F** | **63** | ABD | 1/1/1 | *homA / homB* |
| **k60** | **G** | **M** | **53** | ABC | 1/1/1 | - / *homB* |
| **k64** | **G** | **F** | **61** | ABCC | 1/1/1 | - / *homB* |
| **k74** | **G** | **F** | **52** | ABD | 1/1/1 | - / *homA* |
| **k77** | **G** | **F** | **37** | ABD | 1/1/1 | - / *homB* |
| **k78** | **G** | **M** | **36** | AABD | 1/1/1 | *homA/homA* |
| **k80** | **CA** | **F** | **61** | ABD | 1/1/1 | - / *homB* |
| **k82** | **G** | **M** | **39** | ABD | 1/1/1+ | - / *homB* |
| **k85** | **G** | **F** | **28** | BD | 1/1/1 | - / *homA* |
| **k93** | **DU** | **F** | **37** | ABC | 1/1/1 | - / *homB* |
| **k104** | **CA** | **M** | **46** | ABD | 1/1/1 | - / *homB* |
| **k107** | **DU** | **M** | **26** | ABD | 1/1/1 | - / *homB* |
| **k109** | **G** | **M** | **40** | ABD | 1/1/1 | - / *homB* |
| **k111** | **DU** | **F** | **36** | ABD | 1/1/1 | - / *homA* |
| **k112** | **G** | **M** | **57** | ABD | 1/1/1 | - / *homB* |

| | | | | | | |
|---|---|---|---|---|---|---|
| **k113** | **G** | **M** | **29** | ABD | 1/1/1 | *- / homB* |
| **k115** | **G** | **F** | **82** | ABC | 1/1/2 | *- / homB* |
| **k117** | **G** | **F** | **21** | ABD | 1/1/1 | *- / homA* |
| **k123** | **G** | **M** | **76** | ABD | 1/1/1 | *- / homB* |
| **k131** | **G** | **F** | **61** | ABD | 1/1/1 | *- / homB* |
| **k146** | **G** | **M** | **40** | ABD** | 1/1/1 | *- / homB* |
| **k154** | **G** | **F** | **55** | ABCCCC | 1/1/2 | *- / homB* |
| **k162** | **G** | **F** | **63** | ABD | 1/1/1 | *- / homB* |
| **k165** | **G** | **M** | **28** | ABD | 1/1/1 | *- / homB* |
| **k172** | **G** | **F** | **31** | ABCC | 1/1/1 | *- / homB* |
| **k175** | **G** | **F** | **41** | ABD | 1/1/2 | *- / homA* |
| **k178** | **G** | **F** | **40** | ABD | 1/1/1 | *- / homB* |
| **k182** | **DU** | **N/A** | **N/A** | ABD | 1/1/1 | *- / homB* |
| **k183** | **G** | **M** | **40** | ABD | 1/1/1 | *- / homB* |
| **k185** | **G** | **F** | **52** | ABD | 1/1/1 | *- / homA* |
| **k193** | **G** | **F** | **50** | ABD | 1/1/1 | *- / homB* |
| **k196** | **G** | **F** | **50** | ABD | 1/1/1 | *- / homB* |
| **k197** | **G** | **F** | **45** | ABD | 1/1/1 | *homB+/- / homB* |
| **k208** | **G** | **F** | **56** | ABD | 1/1/1 | *- / homA* |
| **k209** | **G** | **F** | **24** | ABD | 1/1/1 | *- / homB* |
| **k218** | **G** | **F** | **62** | ABD | 1/1/1 | *- / homB* |
| **k220** | **DU** | **N/A** | **N/A** | ABD | 1/1/1 | *- / homA* |
| **k223** | **G** | **F** | **25** | ABD | 1/1/1 | *- / homB* |
| **k235** | **G** | **F** | **50** | ABD | 1/1/1 | *- / homA* |
| **k238** | **DU** | **M** | **55** | ABD | 1/1/1 | *- / homB* |
| **k241** | **G** | **M** | **41** | ABD | 1/1/1 | *-/homB* |
| **k248** | **G** | **M** | **58** | ABD | 1/1/1 | *- / homB* |
| **k255** | **CA** | **F** | **74** | ABD | 1/1/1 | *- / homA* |
| **k258** | **CA** | **M** | **68** | ABD | 1/1/1 | *- / homB* |
| **k259** | **CA** | **M** | **44** | ABD | 1/1/2 | *- / homB* |
| **k260** | **CA** | **M** | **58** | ABD | 1/1/1 | *- / homA* |
| **k261** | **CA** | **F** | **48** | ABD | 1/1/1 | *- / homA* |
| **k262** | **G** | **F** | **56** | ABC | 1/1/1 | *homA/homA* |
| **k263** | **G** | **M** | **59** | ABABD*** | 1/1/1 | *- / homB* |
| **k264** | **DU** | **M** | **32** | ABD | 1/1/1 | *- / homB* |
| **k265** | **DU** | **M** | **42** | ABD | 1/1/1 | *- / homB* |
| **k266** | **G** | **F** | **34** | ABD | 1/1/1 | *- / homB* |
| **k7** | **CA** | **M** | **56** | ABD | 1/1/1 | *- / homB* |
| **k8** | **CA** | **M** | **56** | ABD | 1/1/1 | *- / homB* |
| **k22** | **GU** | **M** | **42** | ABD | 1/1/1 | *- / homB* |

**Abbreviations of disease names are following; G: Gastritis, E : Esophogitis, GU: Gastric ulcer, DU: Duodenal Ulcer, BE: Barrett's Esophagus, CA: Gastric Cancer**

**B indicates Black, W; White**

***homA*$^{+/-}$ or *homB*$^{+/-}$ indicates that the amplicaton of the respective locus was unsuccesful. For example, *homB*$^{+/-}$ /*homB* indicates that *homB* gene is amplicated in locusB, but we don't know that *homB* exists in locusA.**

*** : 10 isolates were failed to get the sequence of CagA EPIYA motif region, so those genotypes were decided by comparing PCR results of other isolates or their own unclear pcr bands.**

****-B motif's proline is replaced with a serine, ESIYA, and is therefore, classified as "other "**

*****-ABABD second -B motif's proline is replaced with leucine, ELIYA, and is therefore, classified as "other"**

**N/A, not applicable**

cultured as previously described [22]. Briefly, bacterial isolates preserved at -80°C, were grown and expanded on antibiotic-supplemented horse blood agar plates under microaerophilic conditions created in GasPak jars with Campypak generators (Mitsubishi Gas Chemical Company, Inc, Tokyo, Japan).

### *bab* Genes Genotyping

Chromosomal DNA was extracted from *H. pylori* strains G27, J99, and 26695 and from the 160 clinical isolates using the G-spin total DNA extraction kit (Intron Biotechnology, inc., Seoul, South Korea). Primers used for genotyping and sequencing *bab* genes in this study were based on the genome sequences of *H. pylori* strains G27, J99 and 26695 (GenBank accession numbers CP001173, AE001439, and CP003904, respectively) (Table 8). PCRs using a locus-specific forward primer and *bab*-specific reverse primer (babAr1, babBr1 or babCr1) were carried out to investigate the presence or absence of *babA*, *babB* or *babC*, respectively (Figure 25) [20, 42]. If all three PCRs per locus were negative, the locus was considered empty. Sequencing of amplicons was used to identify *bab* chimeras and whether the *bab* genes were 'IN'-frame or 'OUT'-of-frame according to the number of CT repeats. In addition, to assess the entire *babC* gene (approximately 2200 bp), PCR amplicons of locusAf/babCr1 and babCf1/locusAr were sequenced using primers babCf1, babCf2, babCf3, babCr1, babCr2, locusAf, and locusAr and the resulting overlapping DNA fragments were aligned. Sanger dideoxy sequencing was performed at Cosmo Genetech Co., Ltd. (Seoul, South Korea) and Geno Tech Corp (Seoul, South Korea), and the resulting DNA sequences were analyzed using Vector NTI

**Table 8. Primer sequences for analysis of *H. pylori bab* genes**

| Primer name | Primer sequence (5’ →3’ ) | Reference |
|---|---|---|
| locusAf | TTTTGAGCCGGTGGATATATTAG | HypDF1 (13)[d] |
| locusBf | CTTTAATCCCCTACATTGTGGA | S18F1 (13)[d] |
| locusCf[a] | ACCCTAGTGGGCATGTGGTA | Hp1-AS (18)[d] |
| locusAr[b] | GGAAATGCGCACACGAGGG | this study |
| babAr1[b] | TTTGCCGTCTATGGTTTGG | BabAR1 (13)[d] |
| babAr2[b] | GAAAGGATGTGTTTTTTCATG | this study |
| babBr1[b] | TCGCTTGTTTTAAAAGCTCTTGA | BabBR1 (13)[d] |
| babBr2[b] | CTGCCAGGACCACAAGCAGT | this study |
| babCf1[b] | GCGGCAAACATCATGCAAGTC | this study |
| babCr1[b] | GACTTGCATGATGTTTGCCGC | this study |
| babCf2[b] | CGGGTTTGCTCAAAGAAAAAAYC[c] | this study |
| babCr2[b] | GRTTTTTTCTTTGAGCAAACCCG[c] | this study |
| babCf3[b] | AAAGAATAACCCCTATAGCC | this study |

[a] One nucleotide in the primer was modified from the indicated primer in reference
[b] Primer was used for sequencing
[c] Underlined Y indicates C or T, and underlined R indicates G or A
[d] Primer name was used in the indicated reference

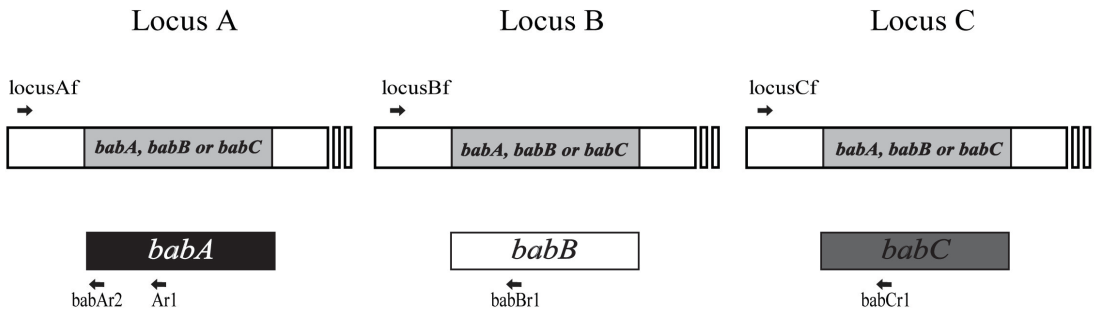



**Figure 25. Outline of *bab* genotyping by PCR.**

(Top) Schematic representation of the three loci where the *bab* genes are generally detected: Locus A, B and C. The annealing positions (arrows) and names of each locus-specific forward are shown. (Bottom) Annealing positions (arrows) and names of *bab*-specific reverse primer are indicated with their respective *bab* gene: *babA* depicted by the black box, *babB* depicted by the white box, *babC* depicted by the grey box. Primers are listed in Table 1, and a full explanation of the genotyping scheme can be found in the Materials and Methods.

version 9.1 (Invitrogen, Carlsbad, CA,USA) and Sequencher 5.1 (Gene Codes Corp., Ann Arbor, MI, USA).

**Phylogenetic Analysis**

Phylogenetic trees were constructed using Phylip v3.695 (Phylogeny Inference Package) (15). First, an input file was generated based on gene variation at seven distinct loci: *cagA* (EPIYA type), *vacA* (s/i/m type), *hom* loci A/B (*hom* type) and *bab* loci A/B/C (*bab* type). Each locus was assigned a discrete numerical character corresponding to its type. Consequently, each clinical isolate was distinguished by a seven-number code representing its genotype. For example, strain G27 has the following genotype: *cagA* EPIYA (ABCC), *vacA* (s1/i1/m1), *hom* locus A (*homB*), *hom* locus B (empty), *bab* locus A (*babC*), *bab* locus B (*babB*), and *bab* locus C (*babA*). This genotype corresponds to the number 2020321 in the input file. All resulting codes were utilized to generate the strain input files used in subsequent analyses (Table 9).

We began constructing the trees via the SEQBOOT program to generate multiple data sets. With this program we used a combination of bootstrapping and character permutation to generate 100 data sets. These data sets were then used as the input file in the PARS program, a general parsimony program that carries out the Wagner parsimony method and allows for up to 8 discrete character states and the use of '?' for states that are unknown. PARS was run using the default settings with the following exceptions: species order was randomized using a random number seed of 5, with 10 replicates and the trees were constructed from the 100 data sets generated in SEQBOOT. PARS searches for the best fit tree, saving the 100 best options. As a final step, we used the

**Table 9 (S3). Phylogenetic coding scheme**

| KEY | | |
|---|---|---|
| Position | Gene | Code |
| 1 | *cagA* EPIYA motif | AB=0, ABC=1, ABCC=2, ABCCC=3, ABCCCC=4, ABD=5, other=6 |
| 2 | *vacA* variant | s1/i1/m1=0, si/i1/m2=1, s1/i2/m2=2, s2/i2/m2=3, s2/i1/m2=4 |
| 3 | *hom* locus A | empty=0, *homA*=1, *homB*=2, *ihom*=3, indeterminant genotype=? |
| 4 | *hom* locus B | empty=0, *homA*=1, *homB*=2, *ihom*=3, indeterminant genotype=? |
| 5 | *bab* locus A | empty=0, *babA*=1, *babB*=2, *babC*=3, *babAB*=4, *babBA*=5, *babBC*=6 |
| 6 | *bab* locus B | empty=0, *babA*=1, *babB*=2, *babC*=3, *babAB*=4, babBA=5, babBC=7 |
| 7 | *bab* locus C | empty=0, *babA*=1, *babB*=2, *babC*=3, *babAB*=4, *babBA*=5, *babBC*=8 |

| AH | | |
|---|---|---|
| **Strain Name** | **Strain Designation** | **Genotype code** |
| **G27** | 1 | 2020321 |
| **B99** | 868 | 2021120 |
| **B101** | 869 | 1010222 |
| **B105** | 870 | 1021420 |
| **B107** | 871 | 2310440 |
| **B108** | 872 | 1310422 |
| **B128** | 873 | 1222120 |
| **B129** | 874 | 1011420 |
| **B134** | 875 | 1011420 |
| **B138** | 876 | 1022420 |
| **B140** | 877 | 1010221 |
| **B148** | 879 | 2022420 |
| **B151** | 880 | 1022120 |
| **B173** | 881 | 2310400 |
| **B177** | 882 | 1011220 |
| **B179** | 883 | 1022120 |

| KH | |
|---|---|
| **Strain Name** | **Genotype code** |
| **k3** | 5002120 |
| **k6** | 5001100 |
| **k10** | 5002120 |
| **k16** | 5002120 |
| **k17** | 5002321 |
| **k19** | 5002120 |
| **k21** | 5002120 |
| **k23** | 5002120 |
| **k24** | 50??120 |
| **k25** | 5002120 |
| **k26** | 5002120 |
| **k27** | 5001120 |
| **k28** | 5002120 |
| **k34** | 5002120 |
| **k35** | 5102120 |
| **k36** | 5002120 |

| | | | | |
|---|---|---|---|---|
| **B181** | 884 | 2310420 | **k37** | 5202120 |
| **B186** | 885 | 1021120 | **k41** | 5001120 |
| **B211** | 886 | 2310420 | **k42** | 5102124 |
| **B213** | 887 | 1022120 | **k43** | 5002140 |
| **B218** | 888 | 1211120 | **k44** | 5002120 |
| **B221** | 889 | 2320220 | **k45** | 5002120 |
| **B235** | 891 | 2022120 | **k46** | 5002120 |
| **B240** | 892 | 2310440 | **k47** | 5210140 |
| **B289** | 893 | 1122420 | **k48** | 5002120 |
| **J58** | 895 | 2320320 | **k57** | 5012120 |
| **J68** | 896 | 2310622 | **k60** | 1002120 |
| **J99** | 897 | 1022120 | **k64** | 2002120 |
| **J104** | 898 | 1011120 | **k74** | 5001120 |
| **J117** | 899 | 2220530 | **k77** | 5002120 |
| **J133** | 900 | 1111140 | **k78** | 6011120 |
| **J154** | 901 | 2310440 | **k80** | 5002120 |
| **J166** | 902 | 10?1500 | **k82** | 5002120 |
| **J178** | 903 | 1021120 | **k85** | 6001140 |
| **J226** | 905 | 2121420 | **k93** | 1002120 |
| **J238** | 906 | 1033100 | **k104** | 5002120 |
| **J243** | 907 | 1022120 | **k107** | 5002120 |
| **J262** | 908 | 2320100 | **k109** | 5002120 |
| **J300** | 909 | 2021120 | **k111** | 5001120 |
| **J44** | 1303 | 1120320 | **k112** | 5002120 |
| **J63** | 1304 | 0320320 | **k113** | 5002120 |
| **J75** | 1305 | 2320320 | **k115** | 1102120 |
| **J150** | 1306 | 2310020 | **k117** | 5001120 |
| **J170** | 1307 | 1022120 | **k123** | 5002120 |
| **J182** | 1308 | 1111424 | **k131** | 5002120 |
| **J195** | 1309 | 2333110 | **k146** | 5002110 |
| **J254** | 1312 | 1020124 | **k154** | 4102120 |
| **J258** | 1313 | 1011120 | **k162** | 5002120 |
| **J291** | 1314 | 1021120 | **k165** | 5002100 |
| **J314** | 1315 | 1011120 | **k172** | 2002120 |
| **B96** | 1316 | 1011420 | **k175** | 5101100 |
| **B116A** | 1317 | 1411120 | **k178** | 5002120 |
| **B125A** | 1319 | 1120122 | **k182** | 5002120 |
| **B127A** | 1320 | 1310220 | **k183** | 5002120 |
| **B130A** | 1321 | 0011124 | **k185** | 5001120 |
| **B136A** | 1322 | 6110120 | **k193** | 5002120 |
| **B137A** | 1323 | 1310420 | **k196** | 5002120 |
| **B147** | 1326 | 0210320 | **k197** | 50?2420 |
| **B158A** | 1327 | 1213120 | **k208** | 5001120 |
| **B166A** | 1328 | 6310420 | **k209** | 5002120 |

| | | |
|---|---|---|
| **B175A** | 1329 | 1320320 |
| **B185A** | 1330 | 1022120 |
| **B201A** | 1331 | 6310320 |
| **B223A** | 1336 | 1022120 |
| **B227C** | 1337 | 1213120 |
| **B234A** | 1338 | 6320320 |
| **B251A** | 1340 | 1021120 |
| **B253A** | 1341 | 1210320 |
| **B255A** | 1343 | 2121120 |
| **B258A** | 1345 | 1112120 |
| **B263A** | 1346 | 2310400 |
| **B266A** | 1348 | 1022220 |
| **B267A** | 1349 | 2022120 |
| **B268A** | 1350 | 2022120 |
| **B269A** | 1351 | 2310320 |
| **B279A** | 1352 | 2310320 |
| **B292A** | 1354 | 1122120 |
| **B293C** | 1355 | 1012120 |
| **B294A** | 1356 | 1013120 |
| **B300A** | 1358 | 2310420 |
| **B301 A** | 1359 | 2310320 |

| | |
|---|---|
| **k218** | 5002120 |
| **k220** | 5001120 |
| **k223** | 5002120 |
| **k235** | 5001120 |
| **k238** | 5002120 |
| **k241** | 5002320 |
| **k248** | 5002120 |
| **k255** | 5001121 |
| **k258** | 5002120 |
| **k259** | 5102120 |
| **k260** | 5001120 |
| **k261** | 5001520 |
| **k262** | 1011220 |
| **k263** | 6002120 |
| **k264** | 5002120 |
| **k265** | 5002120 |
| **k266** | 5002120 |
| **k7** | 5002100 |
| **k8** | 5002100 |
| **k22** | 5002000 |

PARS output trees in the CONSENSE program to generate a majority-rules consensus tree. In doing this, the branch lengths of our final tree represent the number of times a given branching event occurred in the 100 trees generate by PARS. FigTree v1.4.2 is a tree editing program (39).

### Statistical Analysis

Statistical analysis was conducted as previously described (21). Briefly, two-way associations between *bab* genotype, *hom* genotype, *cagA* EPIYA polymorphisms, *vacA* s/i/m polymorphisms, and disease state were analyzed using the Fisher's exact test. Log linear modeling was used to assess higher order associations that were significant at a 5% level. Using categorical values to represent our data sets, we fit a saturated model. We then applied a backward selection algorithm that eliminates the least significant associations at each step and then reforms the model to look for associations. Data analysis was conducted using SPSS version 22 software (SPSS Inc., Chicago, IL, USA) or SAS version 9.3 software (SAS Institute Inc., Cary, NC, USA).

### Nucleotide Sequence Accession Numbers

The sequences for *bab* genotypes in American and South Korean strains were deposited in the NCBI GenBank database with accession numbers KP339308 to KP339491.

## RESULTS

**Sample Population**

Complete demographic data was available for all 80 AH (Tables 6). For the American population, the mean patient age was 57 years, with an age range of 27-84 years. 20.0% (16 patients) of the population was black, and 80.0% (64 patients) was white. Finally, 79 of the patients were male; there was one white female patient, age 50. Of these clinical isolates, 7.5% were from patients with cancer/pre-malignant lesions (2.5% with gastric carcinoma and 5.0% with Barrett's Esophagus), 43.8% were from patients with peptic ulcer disease (31.3% with duodenal ulcers and 12.5% with gastric ulcers), 32.5% were from patients with gastritis, and 16.3% were from patients with esophagitis. Within the South Korean population, age and gender were missing for 2 individuals. Of the remaining 78, the mean age was 48 years, ranging from 20-86 years of age. The South Korean population had a more even distribution between males and females, with 48.7% (38 patients) female and 51.3% (40 patients) male. Within the South Korean female patient group, the mean age was 51, with an age range of 21-86 years. Within the South Korean male patient group, the mean age was 46, with an age range of 20-76 years. Of the 80 KH samples analyzed, 18.8% were from patients with gastric cancer, 33.8% were from patients with peptic ulcer disease (32.5% with duodenal ulcers and 1.3% with gastric ulcers), and 47.5% were from patients with gastritis.

***bab* Genes Genotyping**

Complete *bab* genotyping data can be found in Tables 10 and 11, and the distribution of identified *bab* genes at locus A, B and C is shown in Figure 26. Genotypic

**Table 10 (S2-AH). *bab* genotyping of 80 American *H. pylori* clinical isolates**

| Strain Name | Strain Designation | *bab* locusA/B/C | Status of *bab* locusA/B/C |
|---|---|---|---|
| **B99** | 868 | *babA/babB/-* | - / ON[11] / - |
| **B101** | 869 | *babB/babB/babB* | OFF[9] / OFF[10] / ON[8] |
| **B105** | 870 | *babAB/babB/-* | - / ON[14] / - |
| **B107** | 871 | *babAB/babAB/-* | - / - / - |
| **B108** | 872 | *babAB/babB/babB* | - / OFF[10] / ON[8] |
| **B128** | 873 | *babA/babB/-* | - / OFF[10] / - |
| **B129** | 874 | *babAB/babB/-* | - / ON[8] / - |
| **B134** | 875 | *babAB/babB/-* | - / ON[14] / - |
| **B138** | 876 | *babAB/babB/-* | - / OFF[13] / - |
| **B140** | 877 | *babB/babB/babA* | ON[11] / OFF[9] / - |
| **B148** | 879 | *babAB/babB/-* | - / OFF[10] / - |
| **B151** | 880 | *babA/babB/-* | - / ON[8] / - |
| **B173** | 881 | *babAB/-/-* | - / - / - |
| **B177** | 882 | *babB/babB/-* | ON[8] / OFF[6] / - |
| **B179** | 883 | *babA/babB/-* | - / OFF[9] / - |
| **B181** | 884 | *babAB/babB/-* | - / OFF[9] / - |
| **B186** | 885 | *babA/babB/-* | - / ON[8] / - |
| **B211** | 886 | *babAB/babB/-* | - / OFF[7] / - |
| **B213** | 887 | *babA/babB/-* | - / OFF[10] / - |
| **B218** | 888 | *babA/babB/-* | - / OFF[7] / - |
| **B221** | 889 | *babB/babB/-* | OFF[10] / OFF[9] / - |
| **B235** | 891 | *babA/babB/-* | - / ON[8] / - |
| **B240** | 892 | *babAB/babAB/-* | - / - / - |
| **B289** | 893 | *babAB/babB/-* | - / ON[8] / - |
| **J58** | 895 | *babC/babB/-* | - / ON[11] / - |
| **J68** | 896 | *babBC/babB/babB* | ON[8]/ OFF[9] / OFF[10] |
| **J99** | 897 | *babA/babB/-* | - / ON[8] / - |
| **J104** | 898 | *babA/babB/-* | - / OFF[9] / - |
| **J117** | 899 | *babBA/babC/-* | ON[8] / - / - |
| **J133** | 900 | *babA/babAB/-* | - / - / - |
| **J154** | 901 | *babAB/babAB/-* | - / - / - |
| **J166** | 902 | *babBA/-/-* | ON[8] / - / - |
| **J178** | 903 | *babA/babB/-* | - / OFF[10] / - |
| **J226** | 905 | *babAB/babB/-* | - / OFF[10] / - |
| **J238** | 906 | *babA/-/-* | - / - / - |
| **J243** | 907 | *babA/babB/-* | - / ON[11] / - |
| **J262** | 908 | *babA/-/-* | - / - / - |
| **J300** | 909 | *babA/babB/-* | - / ON[11] / - |
| **J44** | 1303 | *babC/babB/-* | - |
| **J63** | 1304 | *babC/babB/-* | - / OFF[10] / - |
| **J75** | 1305 | *babC/babB/-* | - / OFF[9] / - |

| | | | |
|---|---|---|---|
| **J150** | 1306 | *-/babB/-* | - / ON[8] / - |
| **J170** | 1307 | *babA/babB/-* | - / ON[11] / - |
| **J182** | 1308 | *babAB/babB/babAB* | - / OFF[9] / - |
| **J195** | 1309 | *babA/babA/-* | - / - / - |
| **J254** | 1312 | *babA/babB/babAB* | - / OFF[10] / - |
| **J258** | 1313 | *babA/babB/-* | - / ON[11] / - |
| **J291** | 1314 | *babA/babB/-* | - / ON[8] / - |
| **J314** | 1315 | *babA/babB/-* | - / ON[11] / - |
| **B96** | 1316 | *babAB/babB/-* | - / OFF[6] / - |
| **B116A** | 1317 | *babA/babB/-* | - / ON[8] / - |
| **B125A** | 1319 | *babA/babB/babB* | - / ON[8] / ON[8] |
| **B127A** | 1320 | *babB/babB/-* | OFF[7] / OFF[7] / - |
| **B130A** | 1321 | *babA/babB/babAB* | - / OFF[6] / - |
| **B136A** | 1322 | *babA/babB/-* | - / ON[11] / - |
| **B137A** | 1323 | *babAB/babB/-* | - / OFF[12] / - |
| **B147** | 1326 | *babC/babB/-* | - / OFF[6] / - |
| **B158A** | 1327 | *babA/babB/-* | - / ON[8] / - |
| **B166A** | 1328 | *babAB/babB/-* | - / OFF[7] / - |
| **B175A** | 1329 | *babC/babB/-* | - / OFF[9] / - |
| **B185A** | 1330 | *babA/babB/-* | - / ON[8] / - |
| **B201A** | 1331 | *babC/babB/-* | - / OFF[9] / - |
| **B223A** | 1336 | *babA/babB/-* | - / OFF[9] / - |
| **B227C** | 1337 | *babA/babB/-* | - / ON[8] / - |
| **B234A** | 1338 | *babC/babB/-* | - / ON[11] / - |
| **B251A** | 1340 | *babA/babB/-* | - / ON[8] / - |
| **B253A** | 1341 | *babC/babB/-* | - / OFF[12] / - |
| **B255A** | 1343 | *babA/babB/-* | - / OFF[10] / - |
| **B258A** | 1345 | *babA/babB/-* | - / ON[8] / - |
| **B263A** | 1346 | *babAB/-/-* | - / - / - |
| **B266A** | 1348 | *babB/babB/-* | OFF[10] / OFF[9] / - |
| **B267A** | 1349 | *babA/babB/-* | - / OFF[7] / - |
| **B268A** | 1350 | *babA/babB/-* | - / ON[11] / - |
| **B269A** | 1351 | *babC/babB/-* | - / ON[8] / - |
| **B279A** | 1352 | *babC/babB/-* | - / OFF[9] / - |
| **B292A** | 1354 | *babA/babB/-* | - / ON[8] / - |
| **B293C** | 1355 | *babA/babB/-* | - / OFF[9] / - |
| **B294A** | 1356 | *babA/babB/-* | - / OFF[7] / - |
| **B300A** | 1358 | *babAB/babB/-* | - / OFF[9] / - |
| **B301 A** | 1359 | *babC/babB/-* | - / OFF[7] / - |

**[ ] indicates CT dinucleotide repeats number.**

**'-' indicates an empty Locus in column 2 and lack of CT repeats in column 3**

**Table 11 (S2-KH). *bab* genotyping of 80 South Korean *H. pylori* clinical isolates**

| Strain Name | *bab* locusA/B/C | Status of *bab* locusA/B/C |
|---|---|---|
| k3 | *babA/babB/-* | - / ON[8] / - |
| k6 | *babA/-/-* | - / - / - |
| k10 | *babA/babB/-* | - / OFF[12] / - |
| k16 | *babA/babB/-* | - / ON[8] / - |
| k17 | *babC/babB/babA* | - / OFF[10] / - |
| k19 | *babA/babB/-* | - / ON[11] / - |
| k21 | *babA/babB/-* | - / ON[8] / - |
| k23 | *babA/babB/-* | - / OFF[10] / - |
| k24 | *babA/babB/-* | - / OFF[9] / - |
| k25 | *babA/babB/-* | - / ON[8] / - |
| k26 | *babA/babB/-* | - / ON[5] / - |
| k27 | *babA/babB/-* | - / OFF[10] / - |
| k28 | *babA/babB/-* | - / OFF[7] / - |
| k34 | *babA/babB/-* | - / ON[5] / - |
| k35 | *babA/babB/-* | - / OFF[10] / - |
| k36 | *babA/babB/-* | - / ON[8] / - |
| k37 | *babA/babB/-* | - / ON[8] / - |
| k41 | *babA/babB/-* | - / ON[8] / - |
| k42 | *babA/babB/babAB* | - / OFF[7] / - |
| k43 | *babA/babAB/-* | *-/-/-* |
| k44 | *babA/babB/-* | - / OFF[9] / - |
| k45 | *babA/babB/-* | - / OFF[7] / - |
| k46 | *babA/babB/-* | - / ON[8] / - |
| k47 | *babA/babAB/-* | - / - / - |
| k48 | *babA/babB/-* | - / ON[8] / - |
| k57 | *babA/babB/-* | - / ON[8] / - |
| k60 | *babA/babB/-* | - / OFF[7] / - |
| k64 | *babA/babB/-* | - / ON[8] / - |
| k74 | *babA/babB/-* | - / ON[8] / - |
| k77 | *babA/babB/-* | - / OFF[7] / - |
| k78 | *babA/babB/-* | - / ON[8] / - |
| k80 | *babA/babB/-* | - / ON[8] / - |
| k82 | *babA/babB/-* | - / ON[8] / - |
| k85 | *babA/babAB/-* | - / - / - |
| k93 | *babA/babB/-* | - / ON[8] / - |
| k104 | *babA/babB/-* | - / OFF[7] / - |
| k107 | *babA/babB/-* | - / ON[8] / - |
| k109 | *babA/babB/-* | - / OFF[9] / - |
| k111 | *babA/babB/-* | - / ON[11] / - |
| k112 | *babA/babB/-* | - / ON[11] / - |
| k113 | *babA/babB/-* | - / ON[8] / - |

| | | |
|---|---|---|
| k115 | *babA/babB/-* | - / ON[8] / - |
| k117 | *babA/babB/-* | - / OFF[9] / - |
| k123 | *babA/babB/-* | - / OFF[9] / - |
| k131 | *babA/babB/-* | - / ON[8] / - |
| k146 | *babA/babA/-* | - / - / - |
| k154 | *babA/babB/-* | - / OFF[10] / - |
| k162 | *babA/babB/-* | - / ON[8] / - |
| k165 | *babA/-/-* | - / - / - |
| k172 | *babA/babB/-* | - / ON[5] / - |
| k175 | *babA/-/-* | - / - / - |
| k178 | *babA/babB/-* | - / ON[11] / - |
| k182 | *babA/babB/-* | - / OFF[TT+6] / - |
| k183 | *babA/babB/-* | - / ON[8] / - |
| k185 | *babA/babB/-* | - / OFF[10] / - |
| k193 | *babA/babB/-* | - / OFF[10] / - |
| k196 | *babA/babB/-* | - / ON[8] / - |
| k197 | *babAB/babB/-* | - / OFF[9] / - |
| k208 | *babA/babB/-* | - / ON[8] / - |
| k209 | *babA/babB/-* | - / OFF[9] / - |
| k218 | *babA/babB/-* | - / OFF[TT+6] / - |
| k220 | *babA/babB/-* | - / ON[8] / - |
| k223 | *babA/babB/-* | - / ON[11] / - |
| k235 | *babA/babB/-* | - / ON[8] / - |
| k238 | *babA/babB/-* | - / ON[11] / - |
| k241 | *babC/babB/babA* | - / OFF[TT+6] / - |
| k248 | *babA/babB/-* | - / OFF[10] / - |
| k255 | *babA/babB/babA* | - / ON[11] / - |
| k258 | *babA/babB/-* | - / ON[8] / - |
| k259 | *babA/babB/-* | - / ON[8] / - |
| k260 | *babA/babB/-* | - / ON[8] / - |
| k261 | *babBA/babB/-* | ON[8] / ON[8] / - |
| k262 | *babB/babB/-* | OFF[10] / OFF[10] / - |
| k263 | *babA/babB/-* | - / ON[8] / - |
| k264 | *babA/babB/-* | - / ON[5] / - |
| k265 | *babA/babB/-* | - / OFF[TT+8] / - |
| k266 | *babA/babB/-* | - / OFF[10] / - |
| k7 | *babA/-/-* | - / - / - |
| k8 | *babA/-/-* | - / - / - |
| k22 | *-/-/-* | - / - / - |

**[ ] indicates CT dinucleotide repeats number.**

**'-' indicates an empty Locus in column 2 and lack of CT repeats in column 3**

**Figure 26. Distribution of the *bab* genotypes in AH and KH (n=160).**

Genotypes of one, two, three, or none *bab* genes are grouped into I, II, III, and IV, respectively. These classifications were made on the basis of number of *bab* genes present at locus A, B, and C. A black, white, or grey box indicated *babA*, *babB*, and *babC*, respectively. '-' indicates the absence of *bab* gene at the designated loci. 'AH%' and 'KH%' indicate the frequency of each individual genotype within the specified population (n=80) and 'Total%' shows the percent of the respective genotype in the overall population (n=160). Bracketed numbers indicate the number of isolates possessing each respective genotype out of the population examined. '*' indicates genotypes containing chimeric form(s).

**Figure 26. Distribution of the *bab* genotypes in AH and KH (n=160).**

| Genotype | | Locus A | Locus B | Locus C | AH % | KH % | Total % |
|---|---|---|---|---|---|---|---|
| Single copy | | A | – | – | 2.5 (2/80) | 6.3 (5/80) | 6.9 (11/160) |
| | | – | B | – | 1.3 (1/80) | 0.0 (0/80) | |
| | Chimeric form | A B | – | – | 2.5 (2/80) | 0.0 (0/80) | |
| | | B A | – | – | 1.3 (1/80) | 0.0 (0/80) | |
| Double copy | | A | B | – | 38.8 (31/80) | 78.8 (63/80) | 85.0 (136/160) |
| | | C | B | – | 15.0 (12/80) | 0.0 (0/80) | |
| | | B | B | – | 5.0 (4/80) | 1.3 (1/80) | |
| | | A | A | – | 1.3 (1/80) | 1.3 (1/80) | |
| | Chimeric form | A B | B | – | 16.3 (13/80) | 1.3 (1/80) | |
| | | A B | A B | – | 3.8 (3/80) | 0.0 (0/80) | |
| | | A | A B | – | 1.3 (1/80) | 3.8 (3/80) | |
| | | B A | C | – | 1.3 (1/80) | 0.0 (0/80) | |
| | | B A | B | – | 0.0 (0/80) | 1.3 (1/80) | |
| Triple copy | | A | B | B | 1.3 (1/80) | 0.0 (0/80) | 7.5 (12/160) |
| | | B | B | A | 1.3 (1/80) | 0.0 (0/80) | |
| | | B | B | B | 1.3 (1/80) | 0.0 (0/80) | |
| | | A | B | A | 0.0 (0/80) | 1.3 (1/80) | |
| | | C | B | A | 0.0 (0/80) | 2.5 (2/80) | |
| | | A | B | A B | 2.5 (2/80) | 1.3 (1/80) | |
| | Chimeric form | A B | B | B | 1.3 (1/80) | 0.0 (0/80) | |
| | | A B | B | A B | 1.3 (1/80) | 0.0 (0/80) | |
| | | B C | B | B | 1.3 (1/80) | 0.0 (0/80) | |
| None | | – | – | – | 0.0 (0/80) | 1.3 (1/80) | 0.6 (1/160) |
| | | | | | 100 | 100 | 100 |

analysis revealed that in addition to *babA*, *babB* and *babC*, three types of chimeric *bab* genes were also present: *babAB*, *babBA* and *babBC*. A total of 23 different *bab* genotypes, including a single strain where *bab* was absent from the three known loci (-/-/-), were identified within the 160 isolates chosen for this study. While 6.9% of strains (11/160) carried a single *bab* gene, the majority of the strains (85.0%, 136/160) carried two *bab* genes, one at locus A and one at locus B (Figure 26). Of note, locus C was only occupied if both other loci also carried a *bab* gene (Figure 26). In the KH, 11 of the 23 genotypes were identified, and isolates carrying two *bab* genes were most common, accounting for 87.5% of the isolates (Table 11). The genotype *babA*/*babB*/-, position 1, 2, and 3 correspond to locus A, B and C, respectively, occurred in the 78.8% of the isolates (Figure 26). In comparison, there were many more *bab* genotypes within the AH. In fact, 19 of the 23 genotypes were identified. Twelve genotypes were unique to the AH while 4 were unique to the KH. Similar to the KH, isolates carrying two *bab* genes were most common among AH, accounting for 82.5% of the isolates; however, while the most common genotype of *babA*/*babB*/- occurred in only 38.8%, genotypes of *babAB*/*babB*/- (16.3%) and *babC*/*babB*/- (15.0%) showed a higher frequency in this population (Figure 26). Whereas locus A of KH was occupied by *babA* in greater than 90% of isolates (74/80), most of the diversity within the *bab* genotypes for the AH was driven by locus A, which was significantly more variable ($P<0.0001$) than seen in KH. Indeed, *babA* occurred at locus A only 47.5% of the time (Figure 27). The next most frequent occupants of locus A were chimeric *bab* genes (28.8%), followed by *babC* (15.0%) and *babB* (7.5%).

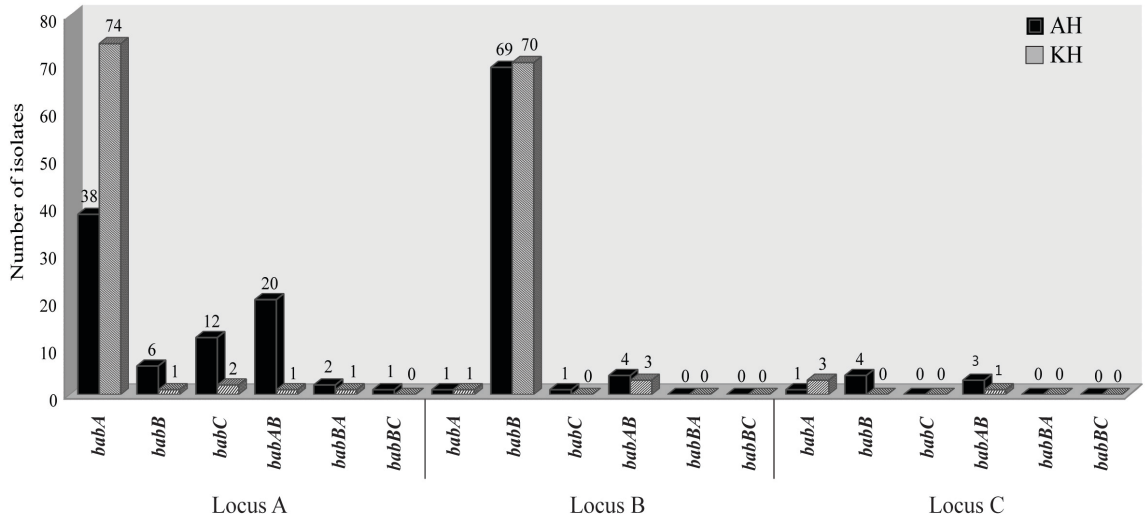



**Figure 27 (S1). Frequency of *bab* gene by locus**

Comparison of the frequency of the presence of *bab*-family genes in loci A, B and C between 80 AH isolates and 80 KH isolates

In total, we identified 3 chimeric *bab* genes, *babAB*, *babBA* and *babBC* that were the result of gene conversion. This gene conversion likely occurs in the 5′ region (~60 to 240 bp from the initiation codon) and in the 3′ region (~1200 to 2100 bp) but this later event is hard to detect due to the homology in this region. It can be speculated that the characteristic of the chimera would follow that of the second *bab* gene, which would constitute the majority of the chimeric gene. *bab* chimeric forms were relatively common in the AH group, occurring at one or more loci in 26/80 (32.5%) of isolates versus 6/80 (7.5%) in the KH group (Figure 26, Table 10, Table 11). The most common chimeric gene was *babAB,* which accounted for 88.9% of chimeric genes; 90.0% of chimeras in AH (27/30) and 83.3% in KH (5/6) (Figure 26, Table 10, Table 11). Within the *babB* coding sequence there are variable CT repeats approximately 56 bp downstream from the initiation codon that are responsible for phase variation. In the *babAB* chimera, this region is displaced with the corresponding part of *babA* [41], resulting in loss of phase variable *babB*. Although it appeared much less frequently, babBA and babBC chimeras were also present (Figure 26, Figure 27). In these instances, CT repeats from the 5′ region of babB were introduced to babA and babC, respectively, resulting in phase variable babA and babC.

Eighty-two *bab* genes from the AH group and 72 *bab* genes from the KH group contained a run of CT repeats (5-14 CT repeats) (Table 12). In all cases, the CT repeats were associated with the 5' region of *babB* genes; therefore, *babBC* and *babBA* also carry CT repeats. CT repeats were found at all three loci within the AH, with the majority (84.2%) occurring at locus B (Table 12). Similarly, in the KH group a majority of the *bab* genes containing CT repeats were found at locus B (97.2%) (Table 12). Interestingly,

**Table 12. 'IN'-frame or 'OUT'-of-frame status deduced by the number of CT dinucleotide repeats in the *babB* coding region.**

| Group | Locus | Frame Status | Pattern of CT repeats | Sequence of CT-repeats region | Number Identified* |
|---|---|---|---|---|---|
| AH (82) | A (9) | IN (5) | 8 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCTCGTTTTTG | 4 |
| | | | 11 | ATGAAAAAAACCCTCCTACTCTCTCTCTCTCTCTCTCTCGTTTTTG | 1 |
| | | OUT (4) | 10 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCTCTCGTTTTTG | 2 |
| | | | 9 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCTCGTTTTTG | 1 |
| | | | 7 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCGTTTTTG | 1 |
| | B (69) | IN (29) | 8 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCTCGTTTTTG | 17 |
| | | | 11 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCTCTCTCGTTTTTG | 10 |
| | | | 14 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCTCTCTCTCTCTCGTTTTTG | 2 |
| | | OUT (40) | 9 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCTCGTTTTTG | 15 |
| | | | 10 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCTCTCGTTTTTG | 10 |
| | | | 7 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCGTTTTTG | 8 |
| | | | 6 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCGTTTTTG | 4 |
| | | | 12 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCTCTCTCTCGTTTTTG | 2 |
| | | | 13 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCTCTCTCTCTCGTTTTTG | 1 |
| | C (4) | IN (3) | 8 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCTCGTTTTTG | 3 |
| | | OUT (1) | 10 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCTCTCGTTTTTG | 1 |
| KH (72) | A (2) | IN (1) | 8 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCTCGTTTTTG | 1 |
| | | OUT (1) | 10 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCTCTCGTTTTTG | 1 |
| | B (70) | IN (42) | 8 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCTCGTTTTTG | 31 |
| | | | 11 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCTCTCTCGCTTTTG | 7 |
| | | | 5 | ATGAAAAAAACCCTTTTACTCTCTCTCTCGTTTTTG | 4 |
| | | OUT (28) | 10 | ATGAAAAAAATCCTTTTACTCTCTCTCTCTCTCTCTCGTTTTTG | 10 |
| | | | 9 | ATGAAAAAACCCCTTTTTCTCTCTCTCTCTCTCTCGTTTTTG | 7 |
| | | | 7 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCGTTTTTG | 6 |
| | | | TT+6 | ATGAAAAAAACCCTTTTATTCTCTCTCTCTCTCGTTTTTG | 3 |
| | | | TT+8 | ATGAAAAAAACCCTTTTATTCTCTCTCTCTCTCTCGTTTTTG | 1 |
| | | | 12 | ATGAAAAAAACCCTTTTACTCTCTCTCTCTCTCTCTCTCTCGTTTTTA | 1 |

* This number corresponds to the number of times the specific CT-repeat pattern was identified at the given locus.

( ) The bracketed number indicates the number of times a *bab* gene carrying CT repeats was identified.

when comparing locus B in the AH group to the KH, significantly more *babB* genes were 'IN'-frame in the KH group than in the AH (*P*=0.0341, data not shown). Although not significant, the trend was similar for all *babB* genes regardless of loci, where more *babB* genes were 'OUT'-of-frame (54.9%) than 'IN'-frame (45.1%) in the AH group (Figure 28). The opposite trend was observed in the KH group; *babB* genes tended to be 'IN'-frame (59.7%) as compared to 'OUT'-of-frame (40.3%) (Figure 28).

Since very little is known about *babC*, we took this opportunity to sequence 15 *babC* genes from our collection and compare the deduced BabCs. As shown in Figure 29. we defined a BabC consensus sequence by comparing 14 BabCs and one BabBC from this study along with the BabC from G27. According to the *babC* sequences, we identified only one chimeric *babBC* gene (AH-J68) and no *babAC* gene. Furthermore, we found that the N-terminus of BabC was most similar to that of BabA, and the C-terminus of BabC was homologous to that of both BabA and BabB (Figure 30). The amino acid sequence of BabC contained two variable regions, VR1 spanning from amino acid 212 to 246 and VR2 spanning from amino acid 409 to 417. The BabC sequences, including VR1 and VR2, shared over 70% identities (Figure 29 and Figure 30).

To further understand the *bab* genotypes in these populations, we next sought to determine if carrying a specific *bab* gene at one locus was associated with the presence of *bab* genes at the other loci. Given the difference in *bab* genotypes between the two populations, we first examined each population separately. A significant association was found in the KH group between the genotype at locus A and locus C (*P*=0.0027) (Table 13). Conversely, in the AH group a significant association was found between locus A and locus B (*P*=0.0028). When KH and AH were combined, there was also a significant

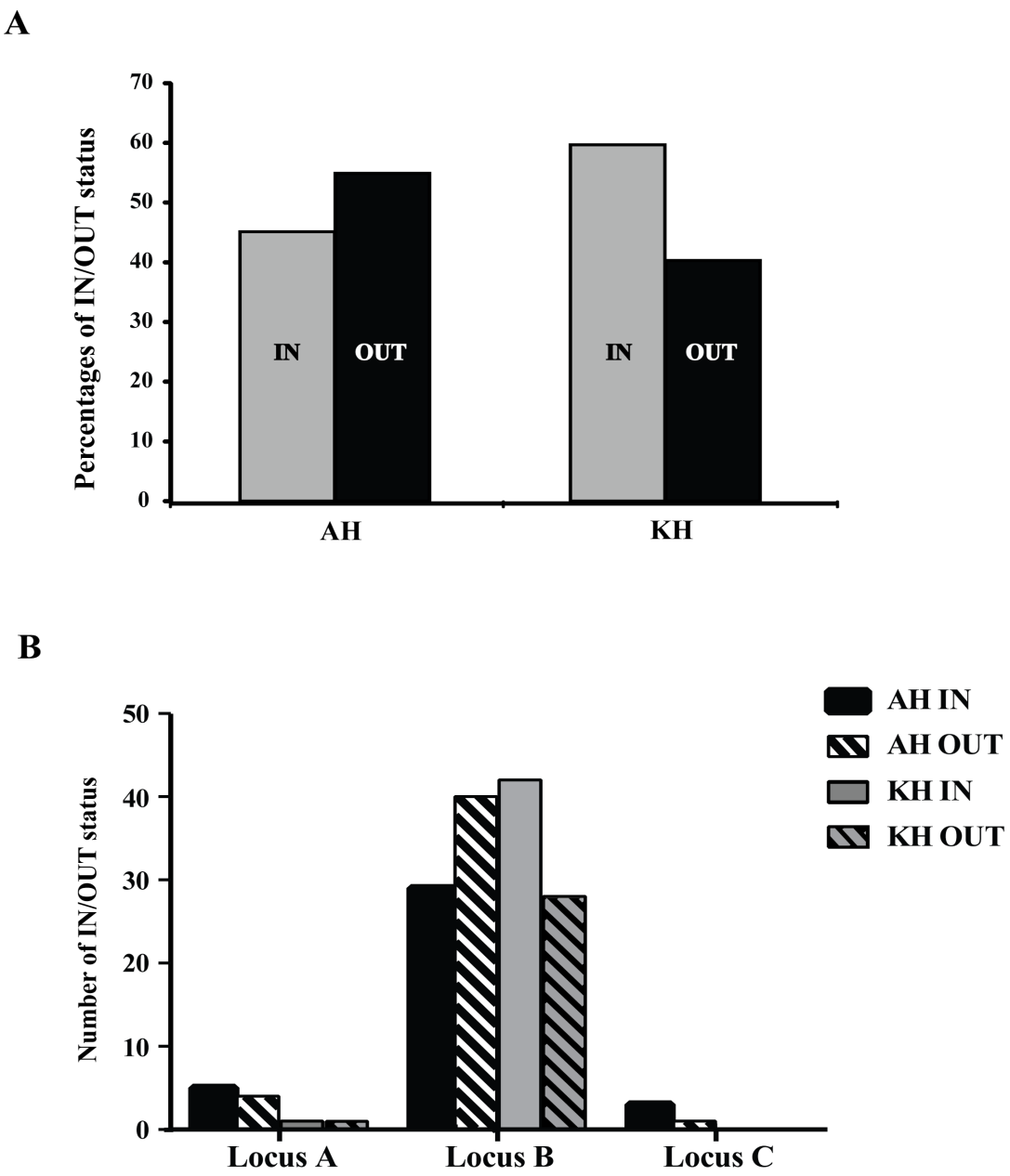



**Figure 28 (S2). Status of *babB* reading frame.**

**A**. Percentage of ' IN' and ' OUT' babB gene reading frame in each population. **B.** Number of ' IN' and ' OUT' babB gene reading frame at each locus.

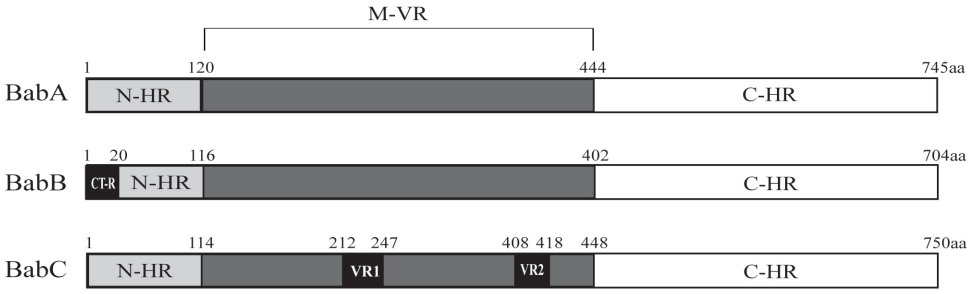



**Figure 29. Schematic comparison of BabA, BabB, and BabC.**

The representative structures of BabA and BabB are based on the J99 amino acid sequence, whereas the structure of BabC is based on the consensus sequence defined in this study. The N-HR indicates N-terminal homology region, M-VR indicates middle variable region, which is characterized by sequence difference among the three Bab proteins. Note that the M-VR is conserved for each of the Bab proteins. The C-HR indicates C-terminal homologous region that shows >90% identity. The CT-R present at the N-terminus of BabB refers to the CT repeat region. VR-1 and VR-2 in BabC indicate regions of variability among the BabC amino acid sequences analyzed in this study (n= 15)

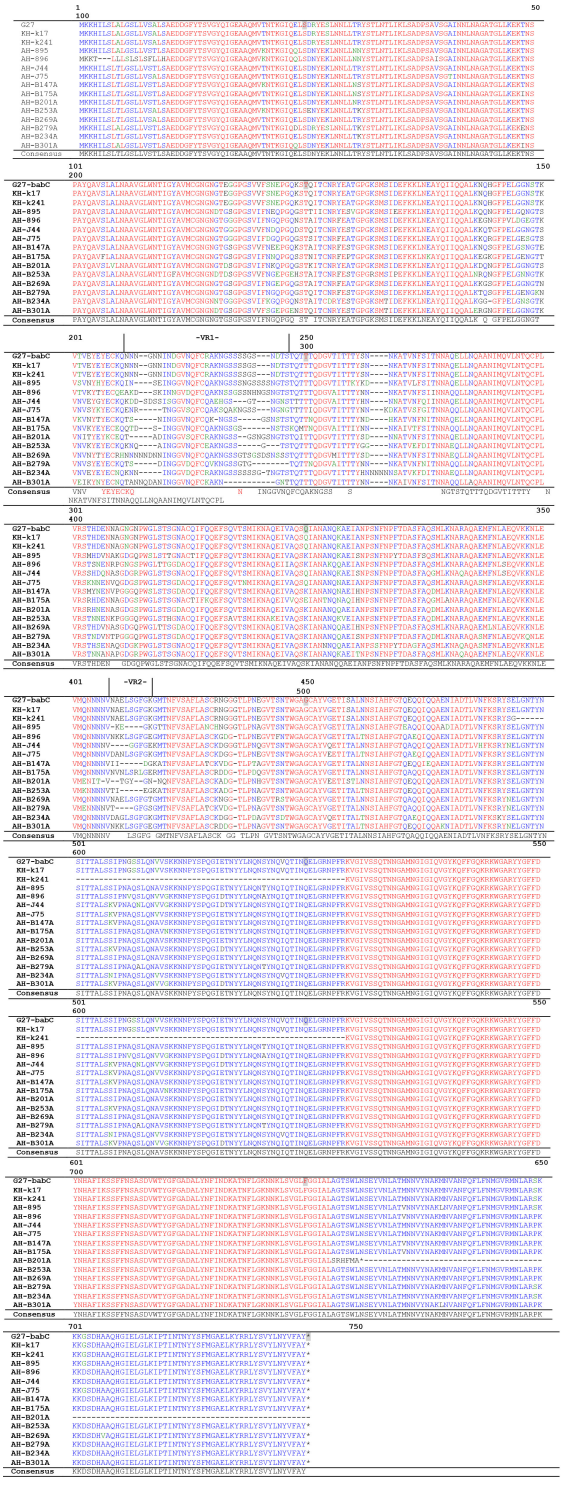


**Figure 30 (S3). Alignment of the amino acid sequences of 13 BabCs and 1 BabBC (AH-J68).**

Variable regions 1 and 2 are indicated, VR1 and VR2, respectively, above the BabC amino acid sequence.

**Table 13. Significant two-way comparisons of *bab* genotype and other factors in both populations[a]**

| *bab* genotype comparison | *P* value[b] for distribution within | | |
|---|---|---|---|
| | KH | AH | Combined[c] |
| *babA,B,*C or empty at Locus A vs *babA,B,C* or empty at Locus B | 0.1046 | **0.0028** | **0.0131** |
| *babA,B,*C or empty at Locus A vs Full or empty *bab* at Locus C | **0.0027** | 0.4111 | 0.4500 |
| *babA,B,*C or empty at Locus A vs one or two *hom* loci occupied | **0.0167** | **<0.0001** | **0.0029** |
| *babB* 'IN' or Other at Locus B vs one or two *hom* loci occupied | **0.0210** | 1.0000 | 0.6070 |
| Full or empty *bab* at Locus C vs one or two *hom* loci occupied | 1.0000 | **0.0390** | 0.1750 |
| *babA,B,*C or empty at Locus A vs *homA or homB* | 0.6390 | **<0.0001** | **<0.0001** |
| *babA,B,C* or empty at Locus B vs *homA* or *homB* | 0.7340 | **0.0160** | **0.0260** |
| *babA,B,*C or empty at Locus A vs v*acA* s1 or *vacA* s2 | 1.0000 | **<0.0001** | **<0.0001** |
| *babA,B,*C or empty at Locus A vs v*acA* i1 or *vacA* i2 | 0.9999 | **<0.0001** | **<0.0001** |
| *babA,B,*C or empty at Locus A vs v*acA* m1 or *vacA* m2 | 0.9999 | **<0.0001** | **<0.0001** |
| *babA,B,*C or empty at Locus A vs *cagA* EPIYA-ABD or Other | 0.5010 | N/A* | **<0.0001** |
| *babA,B,C* or empty at Locus B vs Cancer/Gastric ulcer or Duodenal ulcer/Gastritis | **0.0101** | 0.828[#] | 0.483[#] |
| *babB* or other at Locus B vs Cancer or Gastric ulcer or Duodenal ulcer or Gastritis | **0.0012** | 0.806[#] | 0.447[#] |

[a] For simplicity, table 3 only contains associations for which a statistically significant association was found in at least one grouping; however, an exhaustive analysis was conducted on numerous other permutations of the data.

[b] Statistically significant *P* values are in boldface type.

[c] All isolates (n=160) analyzed as a single group

*No ABD in the AH

[#]Also includes a category for Esophogitis/Barrett's Esophagus

association between locus A and locus B (*P*=0.0131) (Table 13). Of note, in the AH/KH-combined analysis *babC* was more likely to occur at locus A when *babB* at locus B was 'OUT'-of-frame (78.6%, 11/14). Furthermore, if BabB was being expressed from locus A, then *babB* at locus B was more likely to be 'OUT'-of-frame (60.0%, 15/25). Yet interestingly, if *babA* occupied locus A, then *babB* at locus B was more likely to be 'IN'-frame (58.9%, 66/112).

**Associations Between *bab* and *hom* Genes**

Given the associations among *bab* loci, we next assessed whether *bab* genotype was associated with other outer membrane proteins. We have previously looked at the association between *homB* and disease in the South Korean population (25). *hom* paralog (*homA* and *homB*) from the Hom protein family, which shares over 90% identity, are known to occupy two possible loci (also called locus A and B) (4). Therefore, we chose to evaluate any association between the *bab* genes, and *homA* and *homB*. Once again, given the differences in the two populations, each was analyzed individually as well as combined. There were no significant associations when comparing strains for presence of *homA* versus *homB*. Given that *homA* and *homB* are homologous and share two conserved loci, locus A and locus B, we also considered *hom* gene number in which we looked at whether or not both *hom* locus A and locus B were occupied regardless of if it was *homA* or *homB*. We found a significant association with the *bab* paralog present at locus A and number of *hom* loci occupied. This association was seen in both the AH and KH groups (*P*<0.0001 and *P*=0.0167, respectively) (Table 13). In the AH group, presence of a *babB* at locus A was more often associated with a single *hom* gene (61.5%, Table 6

and Table 10). Conversely, in the KH *babA* at locus A occurred most often with a single *hom* (97.3%, S1B and S2B Tables). When the groups were combined, the association between number of *hom* loci occupied and the *bab* gene at locus A remained significant (*P*=0.0029) (Table 13). One factor that stands out when the groups were combined was that although *babC* was rare (n=14), it occurred only in isolates with a single *hom* gene. Also of note in the AH group, in strains where locus C was occupied there was a significant difference between carrying a single *hom* gene versus two *hom* genes (*P*=0.039). Although the finding wasn't significant in the KH group, the same trend was seen; locus C was only occupied in strains with a single *hom* gene (Table 13 and data not shown).

**Associations Between *bab*, *cagA*, and *vacA***

It has been well established that particular *cagA* and *vacA* polymorphisms are associated with more severe disease outcomes (3; 12; 21; 22). Due to their association with disease outcome, we next assessed whether particular *bab* genotypes were found in association with known *cagA* or *vacA* polymorphisms. There were no significant associations between *cagA* and *bab* when the groups were looked at separately; however, when considering the clinical isolates as a whole, there was a significant association between the *bab* gene at locus A and the *cagA* EPIYA polymorphism (*P*<0.0001) (Table 13). The *cagA* EPIYA-ABD polymorphism occurred more frequently in strains that carry a *babA* at locus A (92.9%, 65/70), whereas strains carrying a *babB* or *babC* at locus A were most likely to have a non EPIYA-ABD (93.1%, 27/29 and 85.7%, 12/14, respectively) (S1 and S2 Tables). Similar to *cagA*, *vacA* associations were only apparent

when KH and AH were combined. When we analyzed all 160 isolates, there were significant associations between *bab* at locus A and the s, i, and m regions of *vacA* ($P<0.0001$ for each) (Table 13). Further analysis of the associations between the *vacA* polymorphic regions and *bab* at locus A uncovered a three-way association between *bab* at locus A, the *vacA* s, i, m regions, and occupation of a single or both *hom* loci (Table 14). If we first consider strains carrying *vacA* s1 polymorphisms, these strains were most likely to harbor a *babA* at locus A. Interestingly, *babB* occurs almost equally with the s1 and s2 *vacA* polymorphisms; however, 100.0% of strains containing *babB* at locus A and s2 carried a single *hom* gene compared to strains carrying s1, which were more likely to carry two *hom* genes. The association between s2/*babB*/single *hom* and s1/*babB*/two *hom* genes were both significant ($P=0.00062$ and $P=0.01846$, respectively, data not shown). The i and m regions of *vacA* clustered together in that i1 compared to i2 and m1 compared to m2 resulted in the same distributions. *babA* at locus A was more frequently associated with i1 and m1 as compared to *babC,* which was much more commonly associated with i2 and m2. Once again, *babB* appeared to be found in relatively equal frequencies with i1m1 and i2m2; however, when we analyzed the *vacA* genotype of strains carrying *babB* at locus A, the strains with only an i2m2 *vacA* type were more likely to have a single *hom* gene (87.5%, 14/16, $P=0.00744$) and strains with an i1m1 *vacA* classification were more likely to carry two *hom* genes (76.9%, 10/13, $P= 0.00166$) (data not shown).

**Associations Between *bab* Genotype and Disease**

**Table 14. Three-way comparisons of *bab* genotype and other factors in both populations.**

| *bab* genotype comparison | *P* value[a] for distribution within | | |
|---|---|---|---|
| | KH | AH | Combined[c] |
| *babA, B* or *C* at Locus A vs *vacA* i1/i2 vs one or two *hom* loci occupied | N/A | 0.097 | **0.001** |
| *babA, B* or *C* at Locus A vs *vacA* s1/s2 vs one or two *hom* loci occupied | N/A | 0.173 | **0.002** |
| *babA, B* or *C* at Locus A vs *vacA* m1/m2 vs one or two *hom* loci occupied | N/A | 0.882 | **0.001** |
| *babA, B* or *C* at Locus A vs *vacA* s1i1m1/other vs one or two *hom* loci occupied | 1.000 | N/A[#] | N/A[#] |
| *babA, B* or *C* at Locus A vs *cagA* (AB &Other[$]/ABCs/ABD*)* vs *vacA* m1/m2 | N/A | 1.000 | 0.158 |
| *babA, B* or *C* at Locus A vs *cagA* (AB &Other[$]/ABCs/ABD) vs *vacA* s1/s2 | N/A | 0.963 | 0.881 |
| *babA, B* or *C* at Locus A vs *cagA* (AB &Other[$]/ABCs/ABD) vs *vacA* i1/i2 | N/A | 0.916 | 0.158 |
| *babA, B* or *C* at Locus A vs *cagA* (Other/ABD) vs *vacA* s1i1m1/other | 1.000 | N/A[#] | N/A[#] |
| *babA, B* or *C* at Locus A vs *cagA* (AB &Other[$]/ABCs/ABD) vs one or two *hom* loci occupied | 0.976 | 1.000 | [8)]**0.044** |
| *babA, B* or *C* at Locus B vs one or two *hom* loci occupied vs *vacA* s1/s2 | N/A* | 0.823 | 0.061 |

[a] Statistically significant *P* values are in boldface type.
[b] All isolates (n=160) analyzed as a single group.
[#]These comparison were only done in KH; for AH and combined we looked at i, s, and m regions of *vacA* separately, which wasn't feasible with KH since was overwhelmingly s1/i1/m1.
*All KH strains are *vacA* s1.
[$]AB&Other refers to any *cagA* EPIYA motif that is not $ABC_{(1-4)}$, or ABD.

Given the association between the virulence factors, *cagA* and *vacA* and the *bab* genes, particularly at locus A, we asked if there was an association between *bab* genotype and disease. The only significant association between *bab* and disease was seen in the KH group. In this group, there was a significant association between the *bab* genes present at locus B and disease outcome (*P*=0.0101) (Table 13). Specifically, this association appeared to be driven by the fact that isolates that lacked a *bab* gene at locus B were more likely (66.7%) to come from individuals suffering from more severe disease outcome (gastric cancer and gastric ulcers) than from individuals suffering from gastritis and duodenal ulcers (33.3%). It was also interesting to note, although the finding was not significant (*P*=0.176), that individuals with cancer (gastric carcinoma) or pre-malignant lesions (Barrett's Esophagus) were more likely to have a *babB* gene that was 'IN'-frame' (82.4%, 14/17) compared to 'OUT'-of-frame' (17.6%, 3/17) (data not shown).

**Phylogenetic Analysis of AH and KH**

To complement our statistical analysis, we also conducted a phylogenetic analysis based on gene variation at seven loci: *cagA* (EPIYA type), *vacA* (s/i/m type), *hom* locus A and B (*hom* type), and *bab* locus A, B, and C (*bab* type). As shown in Figure 31A, using these parameters led to a distinct geographical separation between the KH and AH. Only, 6 KH (7.5%) cluster with the AH: K47, K78, K262, K57, K24, and K197. Interestingly, while most of these strains have the predominant KH *cagA* ABD EPIYA motif and *vacA* s1/i1/m1 type, their *bab* and *hom* genotypes more closely resembled the AH. Given this observation, we next constructed a separate consensus tree using only the *hom* and *bab* genes (Figure 31B). Remarkably, this tree was very similar to the tree

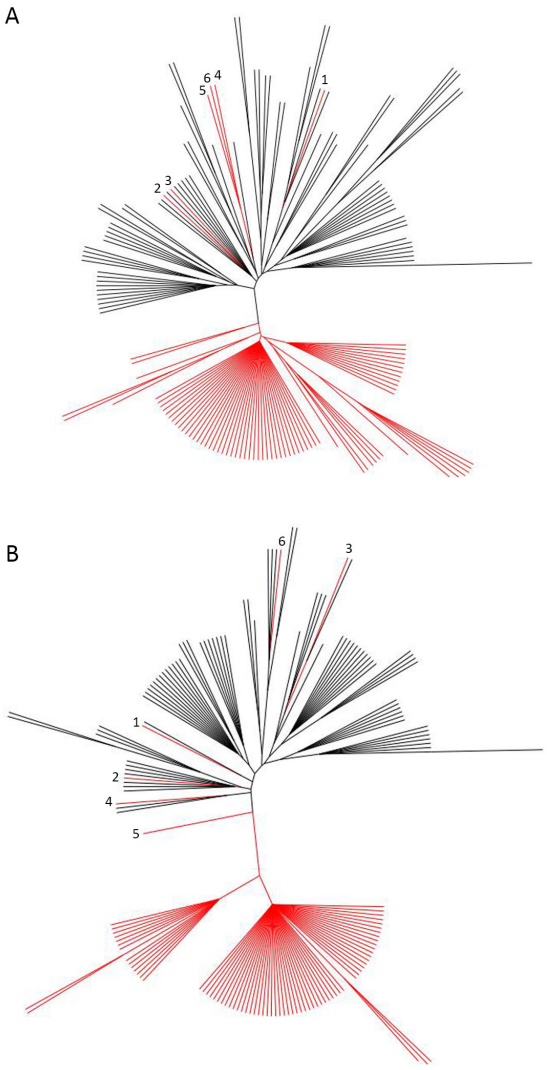



**Figure 31. Phylogenetic analysis of AH and KH.**

**A**. An unrooted consensus tree based on *cagA* EPIYA type, *vacA* s/i/m type, *homA/B* genotype, and *bab* genotype. **B.** An unrooted consensus tree based on *homA/B* and *bab* genotypes only. Trees were created in PHYLIP with Wagner parsimony followed by the majority-rules consensus method. KH branches/nodes are red and AH branches/nodes are black. The six KH grouped with AH are identified by numbers 1-6 in both trees and correspond to the following strains:(1) K47, (2) K78, (3) K262, (4) K57, (5) K24, and (6) K197.

generated using all 7 loci; once again, the majority of the KH and AH segregated, and the same 6 KH grouped more closely with the AH. Closer inspection of the OMP genotypes revealed that the *hom* genotypes of the 6 KH that group among the AH, deviate from the single *hom* at locus B genotype that predominates in the KH.

## DISCUSSION

The *bab* loci have been investigated in detail in several human populations as well as in long term animal colonization studies (7; 27; 29; 32; 42); however, in this study we not only asked what the *bab* genotypes were in South Korean and American populations, but also which *bab* genes were phase variable, which of the phase variable *bab* genes were 'IN'-frame or 'OUT'-of-frame, and how the *bab* genotypes were associated with other genetic markers. Furthermore, we sought to compare these findings between two distinct populations. The KH came from a population with high levels of *H. pylori* infection and a large burden of stomach cancer, while the AH came from a population known to have lower levels of infection and stomach cancer. Although higher rates of *H. pylori* infection in the KH group can partially explain the increased rates of stomach cancer, it is also possible that genetic differences between the *H. pylori* circulating in the South Korean compared to the strains circulating in the United States may also play a role. For our study we began by genotyping the three *bab* loci in 80 clinical isolates obtained from South Korea and 80 clinical isolates from the United States. Overall, we identified 23 different *bab* genotypes. Of these 23, 19 were found in the AH and 11 in the

KH group (Figure 26). Interestingly, although there are a thousand potential *bab* genotypes, our work and previous studies have shown that not all combinations are observed (7; 13; 29; 31). For example, consistent with previous studies (7; 13; 31), we found that carrying two *bab* genes at locus A and locus B was the most common genotype. In the KH group, close to 80% of isolates showed *babA*/*babB*/- genotype. Conversely, even though the AH group did not have one genotype that was found in greater than 40% of the population, the majority of isolates still carried two *bab* genes: *babA*/*babB*/-, *babAB*/*babB*/-, or *bab*C/*babB*/-. Our data suggest that while there are a large number of possible combinations of the *bab* genes at different loci, certain genotypes predominate and the others are rare. Although we identified 23 genotypes, only 5 genotypes were found in at least 5 of the isolates, and 4 genotypes were found only within the KH group while 12 genotypes were exclusive to the AH group. Also of note, in both the KH and AH collections, locus A was most frequently occupied by *babA*, and locus B by *babB*. Furthermore, locus C was only occupied when locus A and B carried a *bab* gene (Figure 22 and Figure 28). Taken together, these data show that isolates carrying two *bab* genes are most common regardless of geographic origin of the strain. Based upon these observations, one might speculate that in an ancestral strain locus A was occupied by *babA* and locus B by *babB*, and as *H. pylori* adapted, gene conversion events have resulted in varying *bab* genotypes. Since the *bab* genes encode OMPs, and since BabA is known to act as an adhesin, it is likely that differences in the *bab* genotype affect how the bacteria interacts with the host. Therefore, perhaps the genotypes that appear to occur in only 1 or 2 individuals reflect adaptation of *H. pylori* to

an individual host, or across gastric regions within a host, and also within an individual over time (13).

In addition to *bab* locus genotyping, we also examined which *bab* genes were subject to phase variation. As previously shown, CT repeats associated with the 5′ region of *babB* can result in slipped-strand mispairing leading to a premature stop in translation (5; 42; 46). In the KH group, *babB* at locus B was significantly more likely to be 'IN' frame' compared to the AH which may reflect a difference in the host populations since *H. pylori* likely uses phase variation as a response to host changes. Probably through gene conversion, CT repeats were also identified in association with *babA* and *babC* at locus A. These data suggest that gene conversion occurred in these strains and resulted in phase variable *babA* and *babC*, imparting *H. pylori* with additional means of regulation over these genes. Interestingly, the most common gene conversion of *babAB* among 3 types of chimeric *bab* genes deprived *babB* of the regulation of phase variation.

Evidence from this study and others suggest that varying the *bab* genotype is a mechanism of *H. pylori* adaptation (9; 11). To try and gain a better understanding of these adaptations, we asked if there were associations among the *bab* loci. In both populations, *babB* was more likely to be expressed from locus B when *babA* occupied locus A. Conversely, if *babB* was expressed from locus A then the *babB* at locus B was more likely to be 'OUT' of frame. Through these analyses we also noted that the *babC* gene was far more likely to occur in strains that also carry a *babB*. Furthermore, in strains containing both *babB* and *babC*, *babB* was more likely to be 'OUT' of frame. Taken together, these data suggest that *H. pylori* actively adjusts its repertoire of *bab* genes, not

only through gene replacement, but also through the formation of chimeric proteins and phase variation supporting the finding of other authors (13; 37). Although the function of *babB* and *babC* are unknown, it could be speculated that the Bab OMPs share overlapping functions, perhaps acting as adhesins similar to BabA. The differences between *babA*, *babB*, and *babC* may affect their function. For instance, if they are functioning as adhesins, these differences may result from different binding targets. Alternatively, the difference between these three genes may also make them antigenically distinct and allow them to escape the immune system.

To this end, we expanded our analysis to include other OMPs from the Hom family. There were no statistically significant differences when we compared strains carrying *homA* to those carrying *homB*; however, given that *homA* and *homB* are greater than 90% identical we decided to ask whether or not we saw an difference in *vacA*, *cagA* and *bab* genotypes between strains carrying a single *hom* gene at locus A or locus B versus strains in which both loci were occupied, regardless of whether it was *homA* or *homB*. We analyzed each population individually as well as combined, and observed an association between the *bab* gene carried at locus A and number of *hom*A and/or B genes. In both population, *babA* occurred most frequently with a single *homA* or *homB* gene, *babB* occurred relatively equally whether a single or both *hom* loci were occupied, and *babC* was only found in strains with a single *homA* or *homB* gene. If one considers the environment in which *H. pylori* persist, having a variable repertoire of OMPs is likely paramount to the success of the organism. While *H. pylori* has greater than 50 OMPs, to our knowledge this is the first evidence showing an association between the Hom and Bab families (4; 46; 50)Since only the function of *babA* is known, it is difficult to

speculate how these proteins may interact functionally (11); however, one can surmise that *H. pylori* finds the right combination of OMPs as a means to balance between the adherent population, which is close to nutrients, and a free population, which is less susceptible to immune clearance. Through mechanisms such as phase variation and gene conversion *H. pylori* is able to fine tune its OMP repertoire in response to changes in the environment.

In addition to investigating the relationship between the *bab* genes and other OMPs, we were also interested in how *bab* genotypes related to known virulence factors such as *cagA* and *vacA* in KH (12; 20-22) and AH (Table 6). In case of *cagA* genotypes, interestingly the AH group contained none of the EPIYA-D motif while KH included 7 strains possessing the EPIYA-C motifs. Importantly, we discovered several associations. A previous study found that strains with *babA* were more likely to have a *cagA* gene than strains that lacked *babA* (18). While all of the strains in our collection contained a *cagA* gene (22), we did find that there was a positive association between having *babA* at locus A and having a *cagA* EPIYA-ABD motif. Strains that carried *babB* or *babC* at locus A were significantly more likely to have a non EPIYA-ABD motif. Carrying a *babA* at locus A was also significantly more likely to occur with *vacA* type s1/i1/m1. Our group has previously shown that *cagA* EPIYA-ABD and *vacA* s1/i1/m1 are associated with more severe disease in the KH (20). Our current data suggest that carriage of *babA* at locus A may also be associated with these more virulent genotypes. Furthermore, we observed a three-way association between *bab* genotype at locus A, occupation of one or both *hom* loci and the *vacA* s1/i1/m1 genotype. Based on these three-way associations, we observed some interesting patterns. For instance, strains harboring a single *homA* or

*homB* gene were more likely to be associated with *vacA* s1, i1 and m1 if there was a *babA* at locus A; however, if there was a *babB* at locus A, then strains with a single *homA* or *homB* gene were more likely to be found with a strain possessing s2/i2/m2, supporting the idea that *babA* may be associated with more virulent genotypes. However, if both *hom* loci are occupied and *babB* is at locus A, then this strain was more likely to be associated with s1, i1, and m1. These findings perhaps demonstrate that it is a specific combination of virulence factors that combine to result in more severe disease outcomes. Although we did not see higher order associations with disease, it could be that our sample size was not adequate to obtain statistical significance over the number of variables that were included.

Taken *en masse*, the *H. pylori* isolates obtained from the American population showed a larger degree of diversity in *bab* genotypes compared to the South Korean population. This may reflect adaptions to a more diverse human population. Furthermore, difference seen between the South Korean and American isolates due to *babB* phase variation may also be due to differences in the populations. Given that difference in host genetics are known to exist, perhaps *babB* is beneficial to *H. pylori* in the South Korean population but not in the American population. Even though the populations were distinct, we identified associations that were seen in both populations, as well as associations that became more apparent when we combined the populations and increased the sample size. Based on our analyses it appears that *babA* may be associated with more virulent genotypes of *cagA* and *vacA*. We also identified an association between OMPs from the Hom family of proteins with the *bab* genes, suggesting a potential coordination between OMP families. Interestingly, our phylogenetic analysis

suggests that the using only the *hom* and *bab* genotypes segregated strains as well as the full genotyping scheme. Specifically the *hom* genotype appears to influence whether the strains cluster with AH or KH. Previous studies have demonstrated that strains with the same *cagA* EPIYA type cluster geographically (12); however, to our knowledge this is the first report in which the number and type of OMP genes have been shown to segregate strains geographically. Although *homA* and *homB* sequence variation has been shown to cluster strains (34), the prior work focused on strains with a single *homA* or *homB* gene, whereas our study focused on which *hom* loci were occupied and not sequence differences. Thus, our observations suggest that perhaps in addition to geographic variation of *homA* and *homB* sequence, there is also a geographic difference in location of the *hom* genes, as well as number of loci occupied (1 vs. 2). Furthermore, these data suggest that OMPs may be a valuable tool to help develop a comprehensive classification system to identify *H. pylori* populations and sub-populations. In all, this study has expanded the profile of *bab* possible genotypes and characterized a new population. Furthermore, it highlights several associations that should be investigated next at a functional level.


## ACKNOWLEDGMENT

We thank Dr. In-Sik Chung for the collection of Korean *H. pylori* clinical strain and miss him deeply.

## REFERENCES


1. Abdullah SM, Hussein NR, Salih AM, Merza MA, Goreal AA, et al. 2012. Infection with *Helicobacter pylori* strains carrying *babA2* and *cagA* is associated with an increased risk of peptic ulcer disease development in Iraq. *Arab J Gastroenterol* 13:166-9

2. Ahuja V, Sharma MP. 2002. High recurrence rate of *Helicobacter pylori* infection in developing countries. *Gastroenterology* 123:653-4

3. Alfizah H, Ramelah M, Rizal AM, Anwar AS, Isa MR. 2012. Association of Malaysian *Helicobacter pylori* virulence polymorphisms with severity of gastritis and patients' ethnicity. *Helicobacter* 17:340-9

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

5. Alm RA, Ling LS, Moir DT, King BL, Brown ED, et al. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 397:176-80

6. Ando T, Peek RM, Pride D, Levine SM, Takata T, et al. 2002. Polymorphisms of *Helicobacter pylori* HP0638 reflect geographic origin and correlate with *cagA* status. *J Clin Microbiol* 40:239-46

7. Armitano RI, Matteo MJ, Goldman C, Wonaga A, Viola LA, et al. 2013. *Helicobacter pylori* heterogeneity in patients with gastritis and peptic ulcer disease. *Infect Genet Evol* 16:377-85

8. Aspholm-Hurtig M, Dailide G, Lahmann M, Kalia A, Ilver D, et al. 2004. Functional adaptation of BabA, the *H. pylori* ABO blood group antigen binding adhesin. *Science* 305:519-22

9. Backstrom A, Lundberg C, Kersulyte D, Berg DE, Boren T, Arnqvist A. 2004. Metastability of *Helicobacter pylori bab* adhesin genes and dynamics in Lewis b antigen binding. *Proc Natl Acad Sci U S A* 101:16923-8

10. Beswick EJ, Suarez G, Reyes VE. 2006. *H. pylori* and host interactions that influence pathogenesis. *World J Gastroenterol* 12:5599-605

11. Boren T, Falk P, Roth KA, Larson G, Normark S. 1993. Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens. *Science* 262:1892-5

12. Bridge DR, Merrell DS. 2013. Polymorphism in the *Helicobacter pylori* CagA and VacA toxins and disease. *Gut Microbes* 4:101-17

13. Colbeck JC, Hansen LM, Fong JM, Solnick JV. 2006. Genotypic profile of the outer membrane proteins BabA and BabB in clinical isolates of *Helicobacter pylori*. *Infect Immun* 74:4375-8

14. de Jonge R, Durrani Z, Rijpkema SG, Kuipers EJ, van Vliet AH, Kusters JG. 2004. Role of the *Helicobacter pylori* outer-membrane proteins AlpA and AlpB in colonization of the guinea pig stomach. *J Med Microbiol* 53:375-9

15. Felsenstein J. 2009. PHYLIP (Phylogeny Inference Package). Department of Genome Sciences: University of Wasghington, Seattle

16. Gerhard M, Lehn N, Neumayer N, Boren T, Rad R, et al. 1999. Clinical relevance of the *Helicobacter pylori* gene for blood-group antigen-binding adhesin. *Proc Natl Acad Sci U S A* 96:12778-83

17. Gwack J, Shin A, Kim CS, Ko KP, Kim Y, et al. 2006. CagA-producing *Helicobacter pylori* and increased risk of gastric cancer: a nested case-control study in Korea. *Br J Cancer* 95:639-41

18. Hennig EE, Allen JM, Cover TL. 2006. Multiple chromosomal loci for the *babA* gene in *Helicobacter pylori*. *Infect Immun* 74:3046-51

19. Ilver D, Arnqvist A, Ogren J, Frick IM, Kersulyte D, et al. 1998. *Helicobacter pylori* adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science* 279:373-7

20. Jang S, Jones KR, Olsen CH, Joo YM, Yoo YJ, et al. 2010. Epidemiological link between gastric disease and polymorphisms in VacA and CagA. *J Clin Microbiol* 48:559-67

21. Jones KR, Jang S, Chang JY, Kim J, Chung IS, et al. 2011. Polymorphisms in the intermediate region of VacA impact *Helicobacter pylori*-induced disease development. *J Clin Microbiol* 49:101-10

22. Jones KR, Joo YM, Jang S, Yoo YJ, Lee HS, et al. 2009. Polymorphism in the CagA EPIYA motif impacts development of gastric cancer. *J Clin Microbiol* 47:959-68

23. Jung SW, Sugimoto M, Graham DY, Yamaoka Y. 2009. *homB* status of *Helicobacter pylori* as a novel marker to distinguish gastric cancer from duodenal ulcer. *J Clin Microbiol* 47:3241-5

24. Kalaf EA, Al-Khafaji ZM, Yassen NY, Al-Abbudi FA, Sadwen SN. 2013. Study of the cytoxin-associated gene a (CagA gene) in *Helicobacter pylori* using gastric biopsies of Iraqi patients. *Saudi J Gastroenterol* 19:69-74

25. Kang J, Jones KR, Jang S, Olsen CH, Yoo YJ, et al. 2012. The geographic origin of *Helicobacter pylori* influences the association of the *homB* gene with gastric cancer. *J Clin Microbiol* 50:1082-5

26. Kawai M, Furuta Y, Yahara K, Tsuru T, Oshima K, et al. 2011. Evolution in an oncogenic bacterial species with extreme genome plasticity: *Helicobacter pylori* East Asian genomes. *BMC Microbiol* 11:104

27. Liu H, Fero JB, Mendez M, Carpenter BM, Servetas SL, et al. 2015. Analysis of a single *Helicobacter pylori* strain over a 10-year period in a primate model. *Int J Med Microbiol* 305:392-403

28. Mahdavi J, Sonden B, Hurtig M, Olfat FO, Forsberg L, et al. 2002. *Helicobacter pylori* SabA adhesin in persistent infection and chronic inflammation. *Science* 297:573-8

29. Matteo MJ, Armitano RI, Romeo M, Wonaga A, Olmos M, Catalano M. 2011. *Helicobacter pylori bab* genes during chronic colonization. *Int J Mol Epidemiol Genet* 2:286-91

30. Nagasawa S, Motani-Saitoh H, Inoue H, Iwase H. 2013. Geographic diversity of *Helicobacter pylori* in cadavers: forensic estimation of geographical origin. *Forensic Sci Int* 229:7-12

31. Nell S, Kennemann L, Schwarz S, Josenhans C, Suerbaum S. 2014. Dynamics of Lewis b binding and sequence variation of the *babA* adhesin gene during chronic *Helicobacter pylori* infection in humans. *MBio* 5

32. Ohno T, Vallstrom A, Rugge M, Ota H, Graham DY, et al. 2011. Effects of blood group antigen-binding adhesin expression during *Helicobacter pylori* infection of Mongolian gerbils. *J Infect Dis* 203:726-35

33. Oleastro M, Cordeiro R, Ferrand J, Nunes B, Lehours P, et al. 2008. Evaluation of the clinical significance of *homB*, a novel candidate marker of *Helicobacter pylori* strains associated with peptic ulcer disease. *J Infect Dis* 198:1379-87

34. Oleastro M, Cordeiro R, Menard A, Yamaoka Y, Queiroz D, et al. 2009. Allelic diversity and phylogeny of *homB*, a novel co-virulence marker of *Helicobacter pylori*. *BMC Microbiol* 9:248

35. Oleastro M, Menard A. 2013. The Role of *Helicobacter pylori* Outer Membrane Proteins in Adherence and Pathogenesis. *Biology (Basel)* 2:1110-34

36. Oleastro M, Monteiro L, Lehours P, Megraud F, Menard A. 2006. Identification of markers for *Helicobacter pylori* strains isolated from children with peptic ulcer disease by suppressive subtractive hybridization. *Infect Immun* 74:4064-74

37. Pride DT, Blaser MJ. 2002. Concerted evolution between duplicated genetic elements in *Helicobacter pylori*. *J Mol Biol* 316:629-42

38. Pride DT, Meinersmann RJ, Blaser MJ. 2001. Allelic Variation within *Helicobacter pylori babA* and *babB*. *Infect Immun* 69:1160-71

39. Rambaut A. 2014. Available: http://tree.bio.ed.ac.uk/software/figtree.

40. Sheu SM, Sheu BS, Chiang WC, Kao CY, Wu HM, et al. 2012. *H. pylori* clinical isolates have diverse *babAB* genotype distributions over different topographic sites of stomach with correlation to clinical disease outcomes. *BMC Microbiol* 12:89

41. Shin A, Shin HR, Kang D, Park SK, Kim CS, Yoo KY. 2005. A nested case-control study of the association of *Helicobacter pylori* infection with gastric adenocarcinoma in Korea. *Br J Cancer* 92:1273-5

42. Solnick JV, 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

43. Suerbaum S, Michetti P. 2002. *Helicobacter pylori* infection. *N Engl J Med* 347:1175-86

44. Talebi Bezmin Abadi A, Taghvaei T, Mohabbati Mobarez A, Vaira G, Vaira D. 2013. High correlation of *babA* 2-positive strains of *Helicobacter pylori* with the presence of gastric cancer. *Intern Emerg Med* 8:497-501

45. Tokudome S, Ando R, Ghadimi R, Tanaka T, Hattori N, et al. 2007. Are there any real *Helicobacter pylori* infection-negative gastric cancers in Asia? *Asian Pac J Cancer Prev* 8:462-3

46. Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, et al. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388:539-47

47. Torres LE, Melian K, Moreno A, Alonso J, Sabatier CA, et al. 2009. Prevalence of *vacA*, *cagA* and *babA2* genes in Cuban *Helicobacter pylori* isolates. *World J Gastroenterol* 15:204-10

48. Wada A, Yamasaki E, Hirayama T. 2004. *Helicobacter pylori* vacuolating cytotoxin, VacA, is responsible for gastric ulceration. *J Biochem* 136:741-6

49. Yamaoka Y, Kita M, Kodama T, Imamura S, Ohno T, et al. 2002. *Helicobacter pylori* infection in mice: Role of outer membrane proteins in colonization and inflammation. *Gastroenterology* 123:1992-2004

50. Yamaoka Y, Ojo O, Fujimoto S, Odenbreit S, Haas R, et al. 2006. *Helicobacter pylori* outer membrane proteins and gastroduodenal disease. *Gut* 55:775-81

51. Zhang XS, Tegtmeyer N, Traube L, Jindal S, Perez-Perez G, et al. 2015. A specific A/T polymorphism in Western tyrosine phosphorylation B-motifs regulates *Helicobacter pylori* CagA epithelial cell interactions. *PLoS Pathog* 11:e1004621

# Appendix B: *Helicobacter pylori* Outer Membrane Protein, HomC, Shows Geographic Dependent Polymorphism that is Influenced by Bab Family

Published as: **Aeryun Kim†, Stephanie L. Servetas†, Jieun Kang, Jinmoon Kim, Sungil Jang, Yun Hui Choi, Hanfu Su, Yeong-Eui Jeon, Youngmin A. Hong, Yun-Jung Yoo, D. Scott Merrell, and Jeong-Heon Cha.** 2016. *Helicobacter pylori* outer membrane protein, HomC, shows geographic dependent polymorphism that is influenced by Bab family. *Journal of Microbiology.* **54**(12). in press

†These authors contributed equally to this work

The work presented in this chapter is the sole work of S.L. Servetas with the following exceptions: A. Kim, J. Kang, J. Kim, S. Jang, Y. H. Choi, H. Su, Y.E. Jeon, Y. A. Hong, Y.J. Yoo, J. H. Cha conducted sequencing analysis and genotyping.

## ABSTRACT

The array of outer membrane proteins (OMPs) found in *Helicobacter pylori* provides a crucial component for persistent colonization within the gastric niche. Not only does *H. pylori* harbor a wide number of OMPs, but these OMPs often vary across strains; this likely contributes to immune evasion, adaptation during long term colonization, and potentially differential disease progression. Previous work from our group described OMP differences among the Bab family (*babA*, *babB*, and *babC*) and

Hom family (*homA* and *homB)* from 80 American *H. pylori* clinical isolates (AH) and 80 South Korean *H. pylori* clinical isolates (KH). In the current study, we expanded our investigation to include the less well characterized Hom family member, HomC. Overall, we identified and genotyped three *homC* variants: $homC^S$, $homC^L$ and $homC^M$, in both populations. Similar to other polymorphic genes, the KH group showed less overall diversity, with 97.5% of strains harboring $homC^L$. In contrast, a more heterogeneous profile was observed in strains derived from an American population; we found nearly equal distribution of $homC^S$ and $homC^L$. Further analysis of the AH group identified associations between *homC* polymorphism and *bab* genotype; in AH strains, there was a significant association between $homC^L$ and carriage of *babA* at locus A. Since *babA* is an important virulence factor for the development of severe gastric disease, these data may suggest that *homC* polymorphism plays a role in *H. pylori* pathogenesis.

## INTRODUCTION

The Gram-negative bacterium *Helicobacter pylori* (*H. pylori*) establishes a persistent gastric infection in humans (2; 33). which results in a number of different clinical outcomes. Disease presentation ranges from mild disease, such as gastritis, to more severe outcomes, such as peptic ulcers (duodenal and gastric ulcers) and gastric cancers (5; 14). Due to the association between *H. pylori* and development of gastric cancers, the World Health Organization (WHO) classified *H. pylori* as class I carcinogen in 1994 (1)

The high levels of genetic diversity and corresponding geographic variations observed among *H. pylori* isolates are two intriguing characteristics of the bacterium. This is especially true when considered in the context of the variable disease outcomes that occur following infection with this bacterium (17; 30; 36). Identification and characterization of the genetic polymorphisms seen in *H. pylori* virulence factors, as well as a greater understanding of the prevalence and distribution of these polymorphisms within geographically distinct populations appears paramount to understanding pathogenesis and development of human gastric disease. Indeed, the association of disease severity with various *H. pylori* virulence factors, such as *cagA*, *vacA,* and several genes encoding outer-membrane proteins (OMPs), has been shown to depend on the population being studied (13; 15-18; 21; 31).

Focusing specifically on OMPs reveals that many of the more commonly studied OMPs are polymorphic and are associated with variable disease outcomes (21; 24; 29). Included among this list are AlpA and B (adherence-associated lipoprotein A and B), BabA (blood group antigen binding adhesin), HomB (Helicobacter outer membrane B), HopZ (Helicobacter outer membrane protein Z), OipA (outer membrane inflammatory protein A) and SabA (sialic acid binding adhesion) (6; 12; 19; 20; 25; 29; 34). In general, the genetic polymorphisms described for the various OMP families include the presence and/or absence of closely related paralogs at one or multiple loci. For example, the *bab* gene family is comprised of three *bab* paralogs: *babA*, *babB,* and *babC* (3). Any of these three genes can be located at three possible chromosomal loci that are referred to as locus A, B and C (21). Previous studies have reported a positive association between the *bab* gene family and development of duodenal ulcers and gastric cancer (11; 21; 34). Among

the *bab* paralogs, BabA is one of the best studied *H. pylori* OMPs and has been shown to function as an *H. pylori* adhesion (10). BabA mediates binding to fucosylated Lewis b present on gastrointestinal epithelial cells (8). Interestingly, the functions of the closely related BabB (HopT) and BabC (HopU) paralogs are unknown (21). At the sequence level, BabA, BabB, and BabC have nearly identical N- and C-terminal domains, but they contain a divergent middle region (21)( this suggests that the middle variable region of BabA is most likely responsible for the specific adhesin function (8-10; 21). Not only do *H. pylori* isolates carry a variety of *bab* combinations, the expression of BabB and BabB chimeric proteins (BabBA, BabBC) can be modulated by changes in the number of cytosine-thymidine (CT) dinucleotide repeats found in the 5' region of the *babB* gene; these changes occur as a result of slipped-strand mispairing (4; 32; 35). Changes in the number of CT repeats result in *babB* variants being 'IN'-frame or 'OUT'-of frame; 'IN'-frame *bab* genes are translated into a full-length functional Bab protein (4; 32; 35). Alternatively, 'OUT'-of frame *bab* genes contain a premature stop codon that leads to production of a nonfunctional truncated Bab protein (4; 32; 35). The above mentioned chimeric *bab* genes are formed by intragenomic homologous recombination between *babB* and *babA* or *babC* (21; 28). This recombination event results in introduction of the CT repeats from *babB* into the 5' region of *babA* or *babC*, which leads to the subsequent modulation of expression via slipped-strand mispairing (21; 28). This well-studied variation in the Bab family is just one example of the highly dynamic nature of OMPs in *H. pylori*.

Another OMP family that has become of interest is the Hom family. However, in comparison to the Bab family, relatively little is known about this family of proteins that

is comprised of four paralogs: HomA, B, C, and D (3; 26). Two of the more well characterized *hom* genes, *homA* and *homB,* are 90% identical at the nucleotide level and share two conserved loci (3). In contrast, very few studies have considered *homC* or *homD,* which are divergent from *homA* and *homB* and occupy two distinct loci. As mentioned above, if present within a strain, *homA* and *homB* can be carried at two possible conserved loci: locus A and locus B (19; 27). Previous studies revealed that the presence of *homB* is significantly associated with development of peptic ulcer disease in Portuguese children and young adults (23; 27) as well as with *cagA* presence and development of gastric cancer in U.S. and Colombian patients (18). However, work from our group found that the impact of the *homB* allele on disease is geographically dependent (19). Most recently, we conducted an analysis of the *bab* gene family in a total of 160 *H. pylori* clinical isolates (21). As part of that analysis, we investigated the relationship between *homA* and *homB,* and the *bab* family (21). Herein, we sought to further expand our understanding of the relationships among the various OMPs. Specifically, we present the first published characterization of *homC* polymorphism as well as an analysis of the distribution of the various forms of *homC* across 160 clinical isolates. Furthermore, we investigate the association between *homC* and the *bab* gene family.

## MATERIALS AND METHODS

### Bacterial Isolates and Culture Conditions

A total of 160 *H. pylori* clinical isolates: 80 South Korean isolates (KH) and 80 American isolates (AH) were used in this study. The isolates have been utilized in

previous studies, and detailed descriptions can be found in Kim *et al*. 2015 and Jones *et al*. 2009. Strains were cultured as previously described (Jones *et al*., 2009; Kim *et al*., 2015). Briefly, all bacterial isolates were incubated on horse blood agar plates at 37ºC under microaerophilic condition created in GasPak jars with Campypak generators (Mitsubishi Gas Chemical Company, Inc, Tokyo, Japan) (21).

### *homC* Gene Genotyping

*homC* genotyping was conducted on all 160 clinical isolates. PCR primers were designed using the published genome sequences of *H. pylori* strains G27, J99, 26695, HP51, HP52, and F30 (GenBank accession numbers CP001173, AE001439, CP003904, CP000012, CP001680, and AP011941, respectively) as reference strains. Chromosomal DNA was extracted from the *H. pylori* reference strains and the 160 clinical isolates using the G-spin total DNA extraction kit (Intron Biotechnology, inc., Seoul, South Korea). PCR amplifications using primers HC-f (GGCTTAACTTTTTATTCTCAA) and HC-r (AAATTGGTCAATTCTTCTGC) were conducted to identify and define the *homC* polymorphism. The PCR parameters were as follows: 95ºC for 2 min, followed by 30 cycles of 95ºC for 30 s, 47ºC for 30 s, and 72ºC for 40 s, and a final 72ºC extension for 5 min. The primers were designed to distinguish amplification of a ~422-bp short *homC* form ($homC^S$), a ~485-bp intermediate *homC* form ($homC^L$), and a ~524-bp long *homC* form ($homC^L$). Some PCR products were sequenced (data not shown) to validate expected DNA sequences. Sanger dideoxy sequencing was performed at Cosmo Genetech Co., Ltd. (Seoul, South Korea), and the resulting DNA sequences were

analyzed using Vector NTI version 9.1 (Invitrogen, Carlsbad, CA, USA) and Sequencher 5.1 (Gene Codes Corp., Ann Arbor, MI, USA).

### Statistical Analysis

The Pearson Chi square test was used to analyze the association between the *bab* genotype and the *homC* polymorphism. Data were analyzed using SPSS version 16 software (SPSS Inc., Chicago, IL) or SAS version 9.1 software (SAS Institute Inc., Cary, NC).

## RESULTS

### Comparative analysis of HomA, HomB, and HomC

Initial HomA, HomB, and HomC comparisons were made using amino acid sequences from the *H. pylori* J99 genome. These three Hom proteins shared 47% identity and varied in length (Figure 32); HomA and HomB were approximately 100 amino acids shorter than HomC (658 and 669 verses 752 residues, respectively). The N-terminal domain (ND) of HomA and HomB shared nearly identical amino acid sequences whereas the N-terminus of HomC was highly divergent from HomA and HomB. The middle domain (MD) of all three proteins differed significantly from one another and thus, lends itself as a means to distinguish the members of this family of proteins, specifically HomA from HomB. Finally, the C-terminal domain (CD) of HomC shares approximately 69% identity with HomA and HomB. Given these observations, we grouped the ND and MD of HomC together as the HomC-variable domains (HC-VD) and next sought to identify the level of conservation seen in HomC across other *H. pylori* strains.

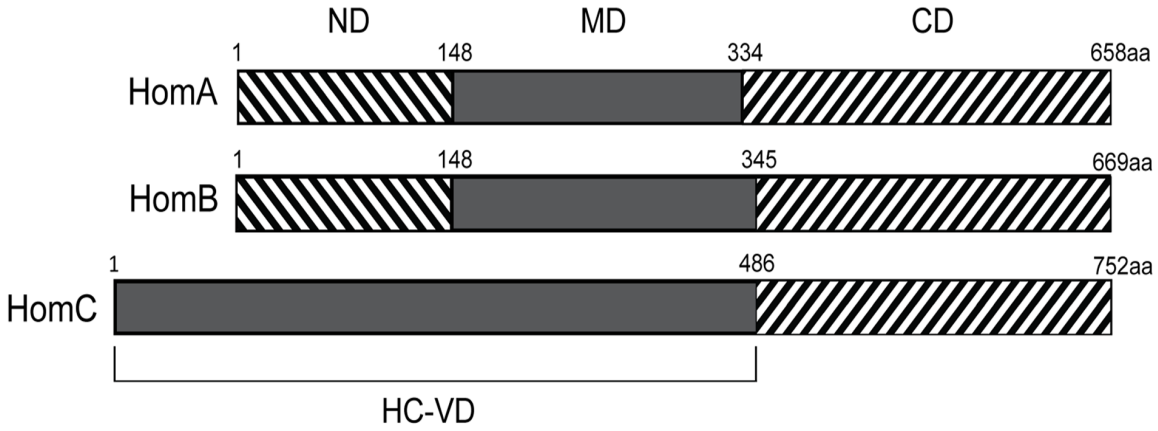



**Figure 32. Schematic comparison of HomA, HomB, and HomC.**

The representative HomA (jhp0649), HomB (jhp0870) and HomC (jhp1008) protein sequences were taken from the published J99 amino acid sequence found in Genebank (GenBank accession numbers AAD06225.1, AAD06437, AAD06586.1). ND indicates N-terminal domain, MD indicates middle domain, CD indicates C-terminal domain, and HC-VD indicates the HomC-variable domain. The HC-VD contains a significantly different amino acid sequence as compared to HomA and HomB. The hatched region of each box designates regions of homology. the ND lined segment designates > 90% identity between HomA and HomB. The CD lined segmant designates > 90% identity between HomA and HomB and > 60% among all three Hom proteins. The gray shading indicates regions of variable amino acid sequences among HomA, HomB and HomC.

### *homC* Polymorphism

To further characterize potential *homC* polymorphisms seen across strains, the *homC* sequences from G27, J99, 26695, HP51, HP52, and F30 were next compared. This analysis revealed two regions that stood out as profoundly different; these were represented as two insertion and/or deletion (Indel) regions, Indel-1 and Indel-2 within the HC-VD (Figure 33). To assess the presence of these polymorphisms within our populations of interest (AH and KH), we used conserved sequences found in the *homC* genes of G27, J99, 26695, HP51, HP52, and F30 to design a PCR primer set that flanked Indel-1 and 2. Representative PCR results are shown in Figure 34. As expected, PCR amplification using these primers primarily resulted in two different sized PCR fragments: a short fragment (~422 bp) and a long fragment (~524 bp), which we defined as $homC^S$ and $homC^L$, respectively. In agreement with the sequence analysis (Figure 33), amplification of *homC* from G27 and HP52 resulted in a 422-bp product corresponding to $homC^S$ whereas a 524-bp product was observed for J99 and HP51, each of which carry $homC^L$ (Figure 34). Amplicons indicating the presence of $homC^S$ (B179, K111) or $homC^L$ (B134, K6) were observed in both the AH and KH groups. Within the AH group, a single isolate, B138, gave an intermediate sized PCR amplicon referred to as $homC^M$ (Figure 33 and Figure 34). Of the 160 isolates tested (see below), only one isolate was shown to harbor $homC^M$. Sequence analysis of this isolate revealed that this strain contained a deletion at indel-1 and an insertion at indel-2 (Figure 33). Overall, we identified three *homC* polymorphisms, $homC^S$, $homC^L$, and $homC^M$ in the AH and KH populations (Figure 34).

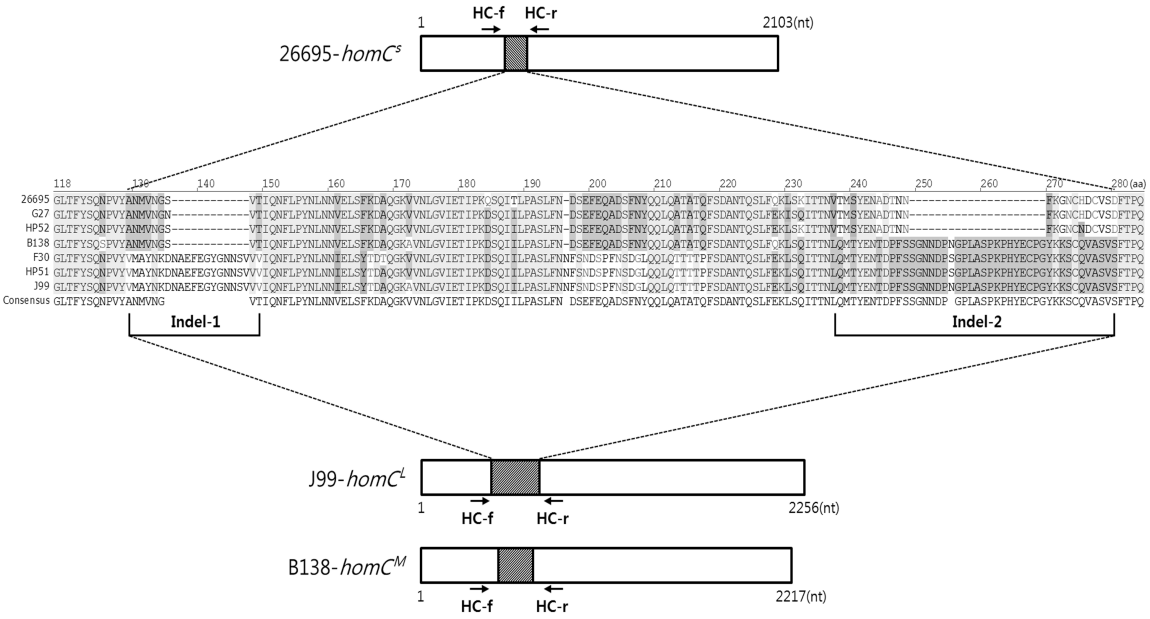



**Figure 33. Outline of *homC* genotyping in *H. pylori* strains.**

The schematic representation shows the aligned amino acid sequences from several representative strains. The identified indel regions, indel-1 and 2, are indicated by the brackets and the corresponding positions within *homC* are indicated. The short fragment, $homC^S$, *homC* of 26695 is depicted on the top of the alignment, and the long fragment, $homC^L$, *homC* of J99 is depicted in the middle, and the intermediate fragment, $homC^M$, *homC* of B138 is depicted at the bottom. The annealing positions (small arrows) and the primers (HC-f and HC-r) used in this study are indicated. The 2217-nt size of B138-$homC^M$ was prediced using sequences of the $homC^M$ PCR amplicon and J99-$homC^L$.

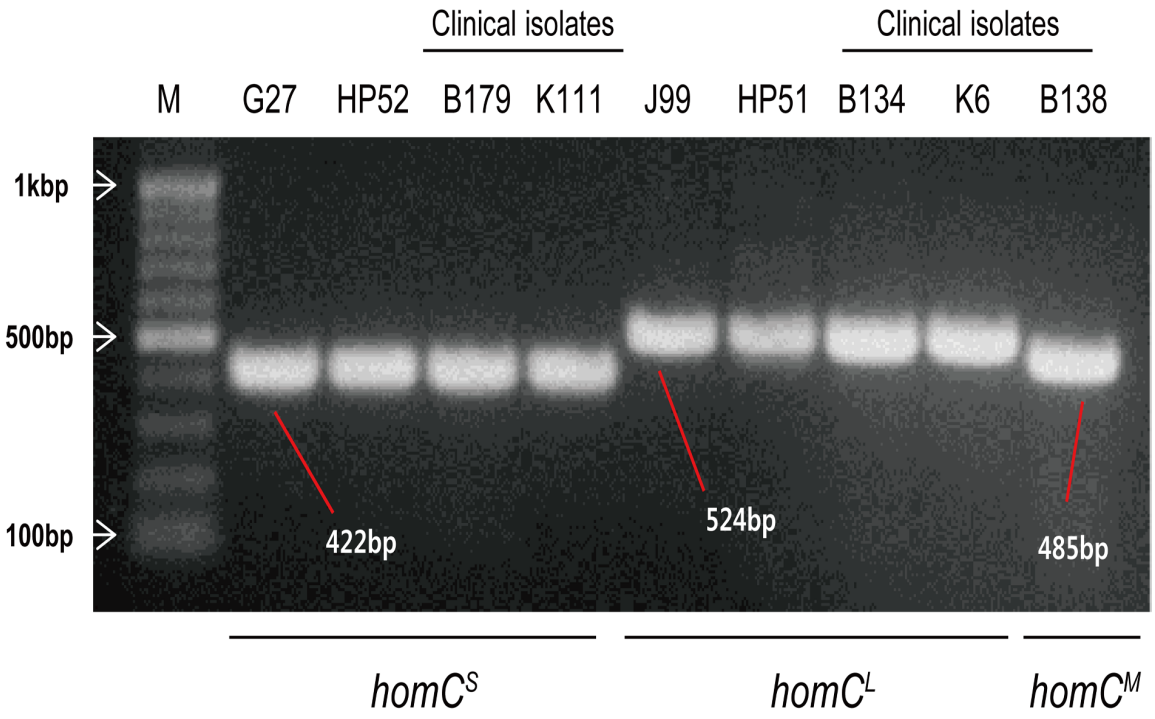



**Figure 34. Genotyping of *homC* variants by PCR.**

A representative gel image showing PCR amplification with the HomC genotyping primers in which examples of $homC^{S}$, $homC^{L}$, and $homC^{M}$ are shown. Results from both positive controls (G27, HP51, J99, and HP52) and clinical isolates (B179, K111, B134, K6, and B138) are shown.

**Distribution of *homC* Polymorphisms in AH and KH**

The distribution of the *homC* polymorphisms in the 80 AH and 80 KH isolates was assessed. All 160 clinical isolates were PCR positive; the majority of isolates (73.1%, 117/160) carried $homC^{L}$, 26.3% of isolates (42/160) carried $homC^{S}$, and 0.6% (1/160) carried $homC^{M}$. The prevalence of each *homC* variant is shown in Figure 35 . When the AH and KH groups were considered independently, within the AH group there was an almost equal distribution between the presence of $homC^{S}$ (40/80, 50%) and $homC^{L}$ (39/80, 48.8%). In addition, one AH isolate (1.3%) carried $homC^{M}$. In comparison, 97.5% of isolates (78/80) from the KH group carried $homC^{L}$: only two isolates (2.5%) contained $homC^{S}$. Thus, there was a significant difference in the distribution of *homC* polymorphisms between AH and KH ($P<0.0001$, Figure 35).

**Association Between *bab* Genotype and *homC* Polymorphism in AH**

Since we previously observed an association between *homA*/*homB* and the *bab* family (21), we reasoned that associations between *homC* and the *bab* family might also be important. Indeed, when we looked at all 160 strains, we observed significant associations between *homC* polymorphism and the *bab* genotype found at locus A ($P<0.0001$) and locus C ($P=0.005$) (data not shown). No significant association was seen with the *bab* genotype at locus B; however, this could be due to the fact that 139/160 strains harbored a *babB* at locus B. Since *babB* is known to harbor CT repeats, we next asked if there was an association between frame status of the 139 isolates and *homC* polymorphism. Indeed, we found that *homC* was associated with *babB* frame status at locus B ($P<0.001$) (data not shown).

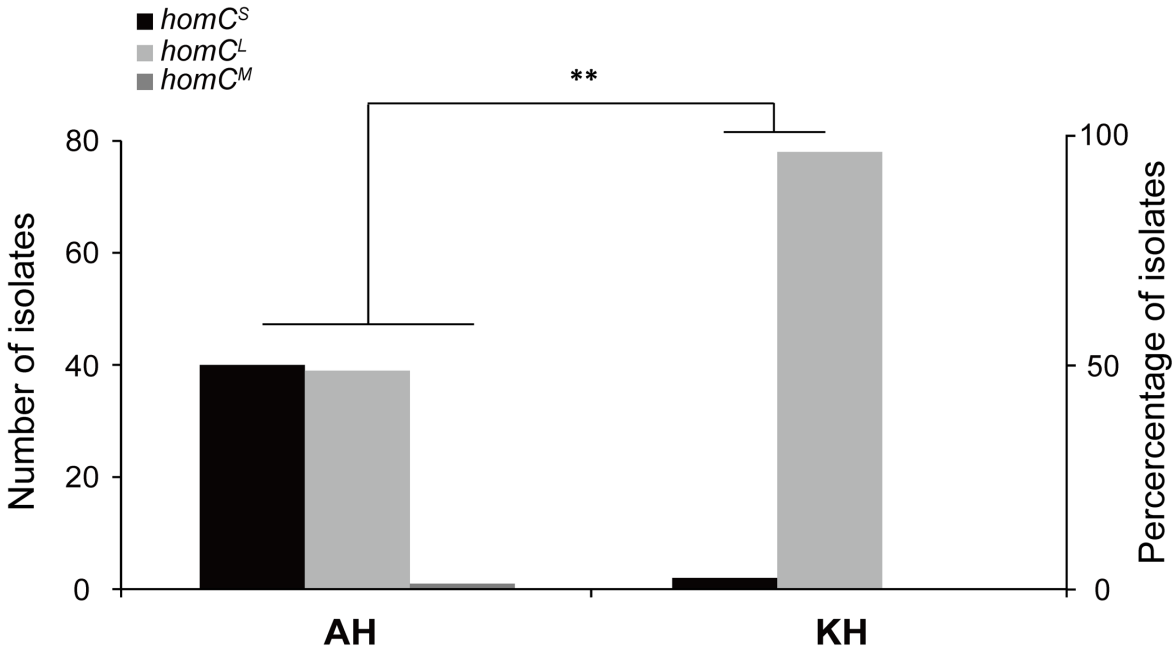



**Figure 35. Distribution of *homC* polymorphisms.**

The distribution of *homC* gene variants in each population is shown. The left y-axis indicates the number of isolates per population, and the right y-axis indicated the percentage of isolates per population. '**' indicates statistically significant differences; $P<0.0001$. *P*-values shown were calculated using Pearson $\chi^2$ square test.

Given the stark differences between *homC*$^{S}$ prevalence across the AH (40/80) and KH (2/80) groups, we next examined the aforementioned associations for each population individually to determine if the large number of KH *homC*$^{L}$ variants were skewing the results. When all of the same parameters were run with the KH samples alone, we found no statistically significant associations. This is not surprising given that the *homC*$^{L}$ polymorphism is found almost uniformly in this population (78/80). However, among the AH isolates, a significant association was identified between *bab* locus A and *homC* polymorphism ($P$<0.0001, Figure 36A). 73.7% of isolates containing *babA* at locus A carried *homC*$^{L}$. Conversely, isolates carrying *homC*$^{S}$ were more likely to carry one of the other *bab* paralogs at locus A (75.0%, 30/40).

Previous analysis showed that AH isolates are most likely to carry *babB* at locus B (69/80, 86.3%) and that *babB* can be either 'IN' frame (29/69, 42.0%) or 'OUT' of frame (40/69, 58.0%) (21). Analysis of the association between *homC* polymorphism and 'IN' frame *babB*, 'OUT' of frame *babB*, or the presence of another *bab* paralog at locus B revealed a significant association between *babB* frame status at locus B and *homC* polymorphism ($P$=0.003, Figure 36B); strains containg an 'IN' frame *babB* at locus B were more likely to carry *homC*$^{L}$ (75.9%, 22/29) while strains containg an 'OUT' of frame *babB* were more likely to carry *homC*$^{S}$ (62.5%, 25/40). Finally, analysis of the *bab* locus C genotype revealed a significant association between the presence (or absence) of a *bab* paralog at locus C and *homC* polymorphism; 100% (8/8) of the strains containing a *bab* paralog at locus C carried *homC*$^{S}$ ($P$=0.012, Figure 36C). As previously stated, no significant associations were found in the KH group; however, the predominate *homC*/*bab* genotype was *homC*$^{L}$, *babA* at locus A, *babB* 'IN' frame at locus B, and empty

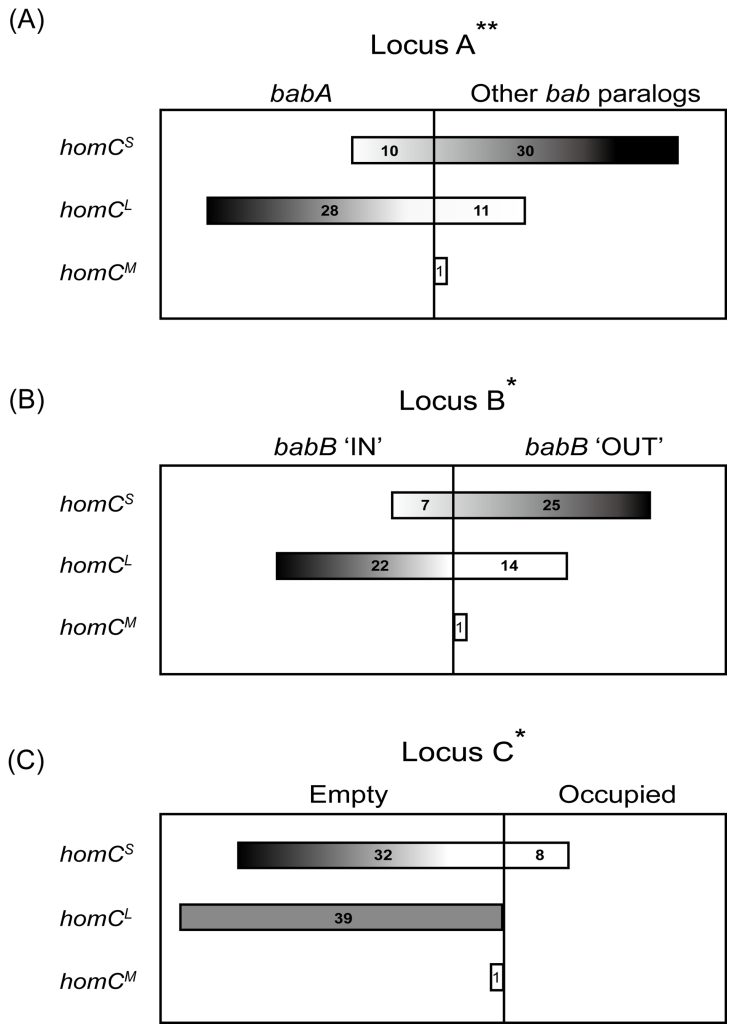



**Figure 36. Schematic depiction of the distribution of the *homC* polymorphism stratified by *bab* locus in AH.**

**A.** The distribution of *bab* genes (*babA* versus other *bab* paralog) at locus A across the three *homC* variants is shown. **B.** The frame status of *babB* ('IN' versus 'OUT') at locus B is shown across the *homC* variants. **C.** The presence of a *bab* gene (empty versus occupied) at locus C is shown within the *homC* variants. '*' indicates a statistically significant difference in distribution; $P<0.05$, '**' indicates a statistically significant difference in distribution; $P<0.0001$. *P*-values shown were calculated using Pearson $\chi^2$ square test.

locus C. It is of note that this *bab* genotype (*babA*/*babB* 'IN' frame/-) was also the most prevelant genotype associated with $homC^{L}$ in the AH isolates.

**Association Between *bab* Chimera and *homC* Polymorphism**

Our previous study revealed that approximately 20% (33/160) of our strains carried a chimeric *bab* gene (21); therefore, we next sought to determine whether strains that harbored a chimeric *bab* gene showed any association with a specific *homC* variant. Given the relatively low number of strains that carried a chimeric *bab*, this analysis was conducted independent of the *bab* locus. Analysis of the total collection revealed that the absence of a chimeric *bab* gene was signficantly associated with $homC^{L}$ ($P<0.0001$, Figure 37A). Given the unequal distribution of chimeric *bab* genes (27/80 AH strains verses 6/80 KH strains), we also examined independent populations. A significant relationship between the presence of a *bab* chimera and *homC* polymorphism was identified in the AH group ($P=0.002$, Fig. 37B). As shown in Figure 37B, $homC^{L}$ was more likely to be present in strains without a *bab* chimera (84.6%, 33/39), whereas strains containing $homC^{S}$ were equally likely to be found with or without a *bab* chimera (50.0%, 20/40). However, of the 27 strains that harbored a chimeric *bab* gene, 20 carried $homC^{S}$ (74.1%, 20/27). Finally, the one $homC^{M}$ variant was found in a strain that contained a *bab* chimera. Despite the association found in the AH group and combined group, all of these KH strains harboring chimeric genes carried $homC^{L}$ (100%, 6/6, data not shown). This data further supports the idea that greater levels of OMP variation are seen in the AH population, and further highlights the geographic differences that can be seen among OMP polymorphisms.

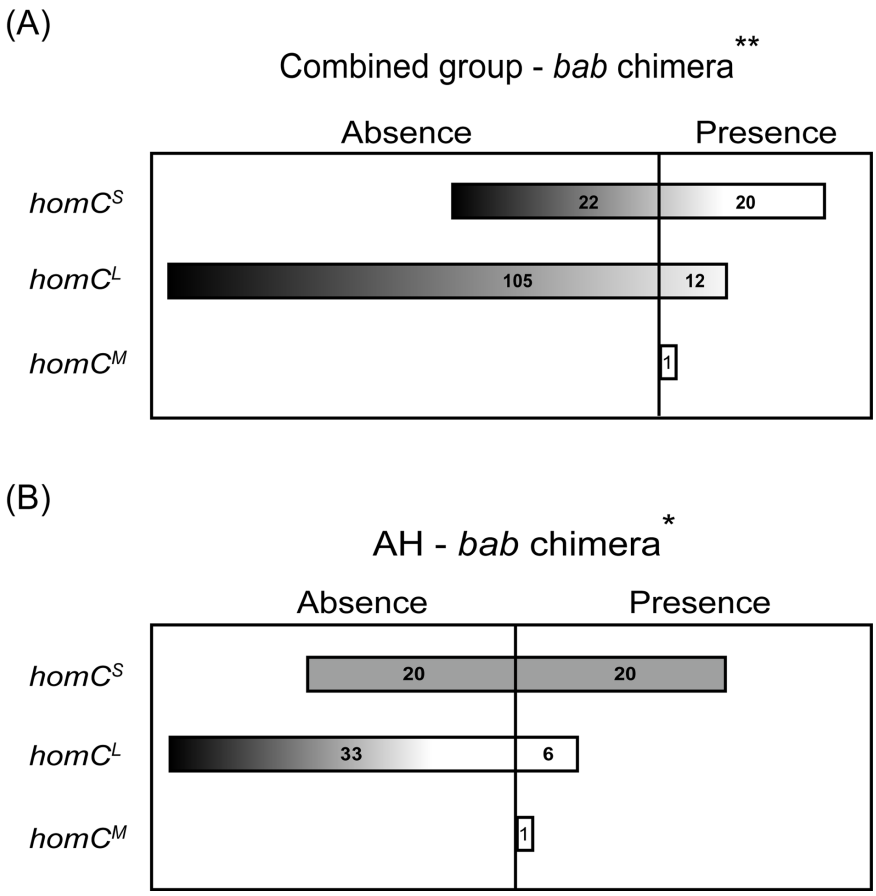



**Figure 37. Schematic depiction of the distribution of the *homC* polymorphism stratified by the presence of *bab* chimeras.**

**A.** The presence of the *bab* chimera stratified by *homC* polymorphism within the combined AH and KH groups. **B.** The presence of *bab* chimeras (absence versus presence) is shown across three *homC* variants within AH. '*' indicates a statistically significant difference in distribution; $P<0.05$, '**' indicates a statistically significant difference in distribution; $P<0.0001$. *P*-values shown were calculated using Pearson $\chi^2$ square test.

## DISCUSSION

The distribution of the *bab* and *hom* gene families has been investigated in several populations (7; 19; 21; 22; 24). However, previous studies regarding these two OMP families were primarily performed independently. Thus, it is difficult to identify potential relationships between the OMPs and their combined effect on disease progression. We recently investigated the *bab*, *homA,* and *homB* genotypes of isolates obtained from South Korean and American populations (19; 21). That analysis provided the first evidence of an association between the Hom and Bab OMP families. Herein, we expanded our analysis to include another member of the Hom family, *homC*. Our results revealed the presence of *homC* variation that differed geographically and showed a relationship with *bab* genotype. Specifically, we identified and genotyped three *homC* polymorphic forms: $homC^S$, $homC^L$, and $homC^M$ in a study group comprised of 160 *H. pylori* isolates. Half of these isolates originated from a Korean population that is known to have a high prevalence of *H. pylori* infection and a high rate of stomach cancer. The other half of the isolates came from an American population known to have much lower rates of infection and stomach cancer (21). Similar to previous findings (21), we found dramatic differences between the two populations; this highlights the geographic variation of *H. pylori* isolates, and supports the idea that genetic variations specific to differing geographies likely contribute to disease outcome.

As shown in our previous studies (21), the KH group showed less diversity than the AH strains. Indeed, 97.5% of KH isolates were $homC^L$ positive, whereas AH isolates showed almost equal distribution of $homC^S$ and $homC^L$. The presence of a single

predominant *homC* variant in the KH isolates may be indicative of selection for HomC$^{L}$ in the South Korean population. Since the function of HomC is unknown, it is difficult to speculate exactly what may drive this selection. However, given the location of HomC in the outer membrane, one possibility may be that host factors that are specific to or more active in the South Korean population are responsible for selection for the presence of *homC$^{L}$*.

When we considered both *homC* and *bab* genotypes, we found that the most common KH genotype was *homC$^{L}$*, *babA* at locus A, *babB* 'IN' at locus B, and an empty *bab* locus C. Despite the increased diversity of the AH isolates, HomC$^{L}$ was also more likely to be found in strains containing *babA* at locus A (73.7%), *babB* 'IN' at locus B (75.9%), and an empty *bab* locus C (100%). Taken together, these observations suggest that *homC$^{L}$* is most often associated with the specific *bab* genotype of *babA*/*babB* 'IN'/-, regardless of the population being examined. This association of *babA* carried at locus A with *homC$^{L}$* is intriguing given that BabA is an important bacterial adhesin that is associated with severe gastric disease (37). This is especially true given that *homC$^{L}$* is the predominant variant found in the South Korean population (Figure 35) where higher levels of *H. pylori* infection and gastric cancer are reported as compared to those seen in the Western population (17; 21). Thus, despite the fact that the function of *homC* is not yet known, it is intriguing to speculate that the different HomC variants may differentially affect *H. pylori* pathogenesis.

Finally, we noted that *bab* chimeras in the AH group were more likely to be found in combination with *homC$^{S}$* and *homC$^{M}$* than with *homC$^{L}$*. Furthermore, the single *homC$^{M}$* strain we identified was found to contain two chimeric *bab* genes. Given this association,

one might speculate that the *homC*$^{S}$, *homC*$^{M}$, and *bab* chimeras specifically arose as an adaptive mechanism in response to selective pressure found within the host. Conversely, this combination may suggest a common *H. pylori* strain lineage, whereas the *homC*$^{L}$ *babA*/*babB* 'IN'/- combination may represent a more distantly related cluster of strains. Regardless, this study has opened many new avenues for exploration of the Hom OMP family and the associations among the Hom family and other *H. pylori* virulence factors. Furthermore, given the presence of HomC in all 160 isolates studied and the association with other known virulence factors, it will be beneficial to investigate the function of HomC and to determine how it coordinates with other virulence factors during *H. pylori* pathogenesis.

## ACKNOWLEDGEMENTS

We thank Richard M. Peek, Jr. and Judith Romero-Gallo (Vanderbilt University) for providing the United States *H. pylori* clinical isolates obtained from patients at Vanderbilt University Medical Center.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2014R1A2A1A11051054).

## REFERENCES


1. 1994. Schistosomes, liver flukes and *Helicobacter pylori*. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7-14 June 1994. *IARC Monogr Eval Carcinog Risks Hum* 61:1-241

2. Ahuja V, Sharma MP. 2002. High recurrence rate of *Helicobacter pylori* infection in developing countries. *Gastroenterology* 123:653-4

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

4. Alm RA, Ling LS, Moir DT, King BL, Brown ED, et al. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 397:176-80

5. Anderson LA, Murphy SJ, Johnston BT, Watson RG, Ferguson HR, et al. 2008. Relationship between *Helicobacter pylori* infection and gastric atrophy and the stages of the oesophageal inflammation, metaplasia, adenocarcinoma sequence: results from the FINBAR case-control study. *Gut* 57:734-9

6. Ando T, Peek RM, Pride D, Levine SM, Takata T, et al. 2002. Polymorphisms of *Helicobacter pylori* HP0638 reflect geographic origin and correlate with *cagA* status. *J Clin Microbiol* 40:239-46

7. Armitano RI, Matteo MJ, Goldman C, Wonaga A, Viola LA, et al. 2013. *Helicobacter pylori* heterogeneity in patients with gastritis and peptic ulcer disease. *Infect Genet Evol* 16:377-85

8. Aspholm-Hurtig M, Dailide G, Lahmann M, Kalia A, Ilver D, et al. 2004. Functional adaptation of BabA, the *H. pylori* ABO blood group antigen binding adhesin. *Science* 305:519-22

9. Backstrom A, Lundberg C, Kersulyte D, Berg DE, Boren T, Arnqvist A. 2004. Metastability of *Helicobacter pylori bab* adhesin genes and dynamics in Lewis b antigen binding. *Proc Natl Acad Sci U S A* 101:16923-8

10. Boren T, Falk P, Roth KA, Larson G, Normark S. 1993. Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens. *Science* 262:1892-5

11. Gerhard M, Lehn N, Neumayer N, Boren T, Rad R, et al. 1999. Clinical relevance of the *Helicobacter pylori* gene for blood-group antigen-binding adhesin. *Proc Natl Acad Sci U S A* 96:12778-83

12. Goodwin AC, Weinberger DM, Ford CB, Nelson JC, Snider JD, et al. 2008. Expression of the *Helicobacter pylori* adhesin SabA is controlled via phase variation and the ArsRS signal transduction system. *Microbiology* 154:2231-40

13. Gwack J, Shin A, Kim CS, Ko KP, Kim Y, et al. 2006. CagA-producing *Helicobacter pylori* and increased risk of gastric cancer: a nested case-control study in Korea. *Br J Cancer* 95:639-41

14. Hatakeyama M. 2009. *Helicobacter pylori* and gastric carcinogenesis. *J Gastroenterol* 44:239-48

15. Jang S, Jones KR, Olsen CH, Joo YM, Yoo YJ, et al. 2010. Epidemiological link between gastric disease and polymorphisms in VacA and CagA. *J Clin Microbiol* 48:559-67

16. Jones KR, Jang S, Chang JY, Kim J, Chung IS, et al. 2011. Polymorphisms in the intermediate region of VacA impact *Helicobacter pylori*-induced disease development. *J Clin Microbiol* 49:101-10

17. Jones KR, Joo YM, Jang S, Yoo YJ, Lee HS, et al. 2009. Polymorphism in the CagA EPIYA motif impacts development of gastric cancer. *J Clin Microbiol* 47:959-68

18. Jung SW, Sugimoto M, Graham DY, Yamaoka Y. 2009. *homB* status of *Helicobacter pylori* as a novel marker to distinguish gastric cancer from duodenal ulcer. *J Clin Microbiol* 47:3241-5

19. Kang J, Jones KR, Jang S, Olsen CH, Yoo YJ, et al. 2012. The geographic origin of *Helicobacter pylori* influences the association of the *homB* gene with gastric cancer. *J Clin Microbiol* 50:1082-5

20. Kennemann L, Brenneke B, Andres S, Engstrand L, Meyer TF, et al. 2012. In vivo sequence variation in HopZ, a phase-variable outer membrane protein of *Helicobacter pylori*. *Infect Immun* 80:4364-73

21. Kim A, Servetas SL, Kang J, Kim J, Jang S, et al. 2015. *Helicobacter pylori bab* Paralog Distribution and Association with *cagA*, *vacA*, and *homA/B* Genotypes in American and South Korean Clinical Isolates. *PLoS One* 10:e0137078

22. Matteo MJ, Armitano RI, Romeo M, Wonaga A, Olmos M, Catalano M. 2011. *Helicobacter pylori bab* genes during chronic colonization. *Int J Mol Epidemiol Genet* 2:286-91

23. Oleastro M, Cordeiro R, Ferrand J, Nunes B, Lehours P, et al. 2008. Evaluation of the clinical significance of *homB*, a novel candidate marker of *Helicobacter pylori* strains associated with peptic ulcer disease. *J Infect Dis* 198:1379-87

24. Oleastro M, Cordeiro R, Menard A, Yamaoka Y, Queiroz D, et al. 2009. Allelic diversity and phylogeny of *homB*, a novel co-virulence marker of *Helicobacter pylori*. *BMC Microbiol* 9:248

25. Oleastro M, Cordeiro R, Yamaoka Y, Queiroz D, Megraud F, et al. 2009. Disease association with two *Helicobacter pylori* duplicate outer membrane protein genes, *homB* and *homA*. *Gut Pathog* 1:12

26. Oleastro M, Menard A. 2013. The Role of *Helicobacter pylori* Outer Membrane Proteins in Adherence and Pathogenesis. *Biology (Basel)* 2:1110-34

27. Oleastro M, Monteiro L, Lehours P, Megraud F, Menard A. 2006. Identification of markers for *Helicobacter pylori* strains isolated from children with peptic ulcer disease by suppressive subtractive hybridization. *Infect Immun* 74:4064-74

28. Pride DT, Blaser MJ. 2002. Concerted evolution between duplicated genetic elements in *Helicobacter pylori*. *J Mol Biol* 316:629-42

29. Senkovich OA, Yin J, Ekshyyan V, Conant C, Traylor J, et al. 2011. *Helicobacter pylori* AlpA and AlpB bind host laminin and influence gastric inflammation in gerbils. *Infect Immun* 79:3106-16

30. Sheu SM, Sheu BS, Chiang WC, Kao CY, Wu HM, et al. 2012. *H. pylori* clinical isolates have diverse *babAB* genotype distributions over different topographic sites of stomach with correlation to clinical disease outcomes. *BMC Microbiol* 12:89

31. Shin A, Shin HR, Kang D, Park SK, Kim CS, Yoo KY. 2005. A nested case-control study of the association of *Helicobacter pylori* infection with gastric adenocarcinoma in Korea. *Br J Cancer* 92:1273-5

32. Solnick JV, 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

33. Suerbaum S, Michetti P. 2002. *Helicobacter pylori* infection. *N Engl J Med* 347:1175-86

34. Talebi Bezmin Abadi A, Taghvaei T, Mohabbati Mobarez A, Vaira G, Vaira D. 2013. High correlation of *babA* 2-positive strains of *Helicobacter pylori* with the presence of gastric cancer. *Intern Emerg Med* 8:497-501

35. Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, et al. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388:539-47

36. Torres LE, Melian K, Moreno A, Alonso J, Sabatier CA, et al. 2009. Prevalence of *vacA*, *cagA* and *babA2* genes in Cuban *Helicobacter pylori* isolates. *World J Gastroenterol* 15:204-10

37. Yu J, Leung WK, Go MY, Chan MC, To KF, et al. 2002. Relationship between *Helicobacter pylori babA2* status with gastric epithelial cell turnover and premalignant gastric lesions. *Gut* 51:480-4