

2011

An Investigation of the Polymorphic *arsS* Gene of *Helicobacter pylori*

Daniel Ross Hallinger
College of William & Mary - Arts & Sciences

Follow this and additional works at: <https://scholarworks.wm.edu/etd>



Part of the [Microbiology Commons](#), and the [Molecular Biology Commons](#)

Recommended Citation

Hallinger, Daniel Ross, "An Investigation of the Polymorphic *arsS* Gene of *Helicobacter pylori*" (2011). *Dissertations, Theses, and Masters Projects*. William & Mary. Paper 1539626912.
<https://dx.doi.org/doi:10.21220/s2-qacc-1410>

This Thesis is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M ScholarWorks. It has been accepted for inclusion in Dissertations, Theses, and Masters Projects by an authorized administrator of W&M ScholarWorks. For more information, please contact scholarworks@wm.edu.

An Investigation of the Polymorphic *arsS* Gene of *Helicobacter pylori*

Daniel Ross Hallinger

Poquoson, Virginia

Bachelor of Science, Virginia Polytechnic Institute and State University, 2008

A Thesis presented to the Graduate Faculty
of the College of William and Mary in Candidacy for the Degree of
Master of Science

Department of Biology

The College of William and Mary
January, 2011

APPROVAL PAGE

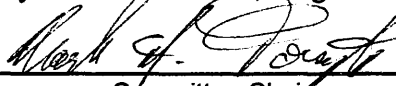
This Thesis is submitted in partial fulfillment of
the requirements for the degree of

Master of Science



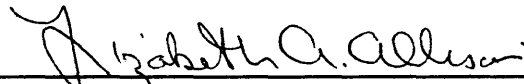
Daniel Ross Hallinger

Approved by the Committee, August 2010



Committee Chair

Associate Professor Mark H. Forsyth, Biology
The College of William and Mary



Professor Lizabeth A. Allison, Biology
The College of William and Mary



Assistant Professor Oliver Kerscher, Biology
The College of William and Mary

COMPLIANCE PAGE

Research approved by

W&M Institutional Biosafety Committee

Protocol number(s): IBC-5724-mhfors

IBC-6016-mhfors

Date(s) of approval: 2009-12-18

2009-05-19

ABSTRACT PAGE

Two-component signal transduction systems are vital for the survival of most bacterial species when exposed to variable environmental conditions. These systems are characteristically composed of a histidine kinase and a cognate response regulator. Histidine kinases perceive environmental changes to activate associated response regulators through phosphorylation events. The ArsRS two-component signal transduction system of the human gastric pathogen *Helicobacter pylori* is critical for its adaptation to acidic conditions. Specifically, ArsS will autophosphorylate in response to an acidic stimulus and subsequently phosphorylate ArsR. Non-phosphorylated and phosphorylated ArsR may regulate gene expression differentially to allow for the survival of the organism.

The *arsS* gene, encoding the histidine kinase ArsS, is subject to frequent mutation due to a homopolymeric cytosine tract in its 3' terminus. Together, the differences in cytosine tract length and the existence of a stop codon for each *arsS* open reading frame allow for the generation of various, functional ArsS isoforms. Here, we hypothesize that *H. pylori* populations consist of cells that encode for a number of *arsS* alleles with variant homopolymeric cytosine tract lengths that allow for the propagation of different ArsS isoforms.

In order to determine whether different *arsS* alleles existed in populations of *H. pylori*, amplified fragment length polymorphism polymerase chain reactions (AFLP-PCR) were utilized to amplify a genomic region of *arsS* that encodes roughly 300 base pairs including the cytosine tract. The size and relative frequency of resulting AFLP-PCR amplicons were quantified via amplified fragment length polymorphism (AFLP) analyses. To verify that the amplicons differed in length due to differences in cytosine tract lengths, sequencing was performed on plasmid-cloned amplicons produced from the same *arsS* region. As a control, a non-polymorphic region of *arsS* was utilized for similar AFLP and sequencing analyses.

AFLP and sequencing data suggest that different *arsS* alleles exist within *H. pylori* populations and, in some instances, within single strains derived from three gastric regions of individual patients. Additionally, the relative frequencies of *arsS* alleles and, thus, *arsS* open reading frames were able to be determined. Data suggest that a predominant *arsS* open reading frame can vary among *H. pylori* populations from different patients and among populations within a single patient. Interestingly, *H. pylori* populations from two of twelve patients significantly ($p < 0.05$) encoded different predominant *arsS* open reading frames in various anatomical regions of their stomachs.

Sequencing analyses also allowed for ArsS isoforms to be identified via *in silico* translations of the vector-cloned amplicons. Here, data suggest that the *arsS* open reading frames could be translated into ArsS isoforms with variable C-terminal regions. These data also identified another ArsS isoform that had not been previously recognized in the literature. Additionally, sequencing data allowed for the determination that many of our populations consisted of multiple strains, while some were monoclonal. Interestingly, each patient appeared to be infected with their own predominant strain in at least two of their gastric regions. Collectively, our data enhance the limited knowledge base concerning the *arsS* gene and the ArsS protein of the carcinogenic bacterium, *Helicobacter pylori*.

Table of Contents

Dedication	vi
Acknowledgements	vii

Chapter 1

Introduction	1 - 20
History of <i>Helicobacter pylori</i>	1
General <i>Helicobacter pylori</i> Biology	2
Acid Acclimation of <i>Helicobacter pylori</i>	3
Two-component Signal Transduction Systems	6
Characterization of ArsRS.....	8
Genome Plasticity	10
Phase Variation	14
This Study	19

Chapter 2

Materials and Methods	21 - 32
Bacterial Strains and Growth Conditions	21
General Techniques.....	22
Amplified Fragment Length Polymorphism (AFLP) Analyses of <i>arsS</i> Regions	23
Sequencing Analyses of <i>arsS</i> Regions	28

Chapter 3

Results	33 - 68
<i>arsS</i> Poly C Region AFLP-PCR Amplicons Vary in Length Among and Within Clinical <i>Helicobacter pylori</i> Populations.....	33

Predominant <i>arsS</i> Poly C Region AFLP-PCR Amplicons Vary in Length Among Clinical <i>Helicobacter pylori</i> Populations	35
Variant <i>arsS</i> Open Reading Frames are Encoded Within Clinical <i>Helicobacter pylori</i> Populations	37
Predominant <i>arsS</i> Open Reading Frames are Encoded Among Clinical <i>Helicobacter pylori</i> Populations	40
<i>arsS</i> Control Region AFLP-PCR Amplicons Vary in Length Among and Within Clinical <i>Helicobacter pylori</i> Populations.....	41
Predominant <i>arsS</i> Control Region AFLP-PCR Amplicons Vary in Length Among Clinical <i>Helicobacter pylori</i> Populations.....	42
Cloned <i>arsS</i> Poly C Region Sequences Encode Variable Polycytosine Tract Lengths Among and Within Clinical <i>Helicobacter pylori</i> Populations.....	43
Cloned <i>arsS</i> Poly C Region Sequences Reveal Predominant Poly Cytosine Tract Lengths Among Clinical <i>Helicobacter pylori</i> Populations.....	46
Cloned <i>arsS</i> Poly C and Control Region Sequences Vary in Length Among Clinical <i>Helicobacter pylori</i> Populations Due to Deletions in Primer Binding Regions	46
Cloned <i>arsS</i> Poly C Region Sequences Indicate Different <i>Helicobacter pylori</i> Strains Within Populations	47
Cloned <i>arsS</i> Poly C Region Sequences Indicate Predominant <i>Helicobacter pylori</i> Strains <u>Within</u> Populations	49
Cloned <i>arsS</i> Poly C Region Sequences Indicate Predominant <i>Helicobacter pylori</i> Strains <u>Among</u> Populations of an Individual Patient	52
Cloned <i>arsS</i> Control Region Sequences Indicate Different <i>Helicobacter pylori</i> Strains Within Populations	55
Cloned <i>arsS</i> Control Region Sequences Indicate Predominant <i>Helicobacter pylori</i> Strains Within Populations	56

Cloned <i>arsS</i> Control Region Sequences Indicate Predominant <i>Helicobacter pylori</i> Strains Among Populations of an Individual Patient.....	58
The <i>arsS</i> Poly C Region is Variable Outside of the Polycytosine Tract Among <i>Helicobacter pylori</i> Populations.....	60
Different <i>arsS</i> Open Reading Frames Arise Due to Changes in Polycytosine Tract Length and the Polymorphic Deletion.....	61
Different <i>arsS</i> Open Reading Frames are Present Within and Among <i>Helicobacter pylori</i> Populations Allowing for Different ArsS Isoforms	62
The C-terminal Peptide Sequence of ArsS Isoforms is Variable.....	66

Chapter 4

Discussion	69 - 81
Conclusion	74
Future Directions.....	76

Appendices and References

Appendices	82 - 121
Appendix 1: General Broth and Agar Media.....	82
Appendix 2A: AFLP Data from Poly C Regions of <i>H. pylori</i> Populations.....	83
Appendix 2B: AFLP Data from Control Regions of <i>H. pylori</i> Populations.....	119
References	122 - 129

List of Tables

Chapter 2: Materials and Methods

2.1	Clinical populations of <i>Helicobacter pylori</i>	21
-----	--	----

2.2	Oligonucleotide primers	25
2.3	Calculating AFLP fragment length frequencies	27

Chapter 3: Results

3.1	Summary of poly C region AFLP data from <i>H. pylori</i> populations	35
3.2	Summary of control region AFLP data from <i>H. pylori</i> populations	43

List of Figures

Chapter 1: Introduction

1.1	Generic model for histidine kinase autophosphorylation	6
1.2	Generic model for response regulator phosphorylation	7
1.3	A model for <i>Salmonella</i> phase variation	15
1.4	Generic model for slipped-strand mispairing	16

Chapter 2: Materials and Methods

2.1	Example of AFLP raw data	26
-----	--------------------------------	----

Chapter 3: Results

3.1	<i>arsS</i> poly C region AFLP data from the antrum <i>H. pylori</i> population of patient B256.....	34
3.2	Extrapolation of AFLP data into open reading frame data.....	39
3.3	<i>arsS</i> control region AFLP data from the antrum <i>H. pylori</i> population of patient B284.....	42
3.4	Differences in polycytosine tract lengths among all <i>H. pylori</i> populations.....	45
3.5	Number of sequence polymorphisms outside of the polycytosine tract among poly C region clones from <i>H. pylori</i> populations	48
3.6	Estimated number of strains within <i>H. pylori</i> populations via poly C region clones	51

3.7	Primary and secondary predominant strains of <i>H. pylori</i> populations determined with poly C region clones.....	54
3.8	Number of sequence differences among control region clones within <i>H. pylori</i> populations.....	56
3.9	Estimated number of strains within <i>H. pylori</i> populations via control region clones.....	57
3.10	Primary and secondary predominant strains of <i>H. pylori</i> populations determined with control region clones.....	59
3.11	Presence of a deletion among <i>arsS</i> poly C region clones from each gastric population of twelve patients.....	61
3.12	Schematic of the <i>arsRS</i> operon.....	62
3.13	Differences in ArsS isoform distribution among all <i>H. pylori</i> populations.....	65
3.14	Variable C-terminal amino acid sequence of the ArsS isoforms.....	67

Dedicated to my family for all of their love and support.

Acknowledgements

I would like to thank my family for all they have provided me over the course of my education at the College of William and Mary. I would especially like to thank my father, Scott, for being a superior role model and the greatest dad a son could ever have. My girlfriend, Madeline, also deserves a great deal of recognition, as she has been immensely influential in my life with her continuous love and unsurpassed support. I also greatly appreciate the guidance and direction that Dr. Lizabeth Allison and Dr. Oliver Kerscher have provided me the past two years. Furthermore, I would like to express gratitude to Lidia Epp for all of her assistance and, at times, patience. Members of the Forsyth Lab are also due recognition because of their friendship and words of encouragement. I would particularly like to thank Sam for being a great darts opponent, but a better friend. Lastly, but certainly not least, I want to recognize my primary advisor Dr. Mark Forsyth, as he has provided me with a wealth of knowledge that is indubitably priceless. Additionally, his character, sense of humor, and Hawaiian t-shirts will forever remain in my memories of William and Mary. Mark has been a great friend and leader; and because of this, I will always value the time that I spent in the Forsyth Laboratory.

Chapter 1

Introduction

History of *Helicobacter pylori*

Prior to 1983, the human stomach was believed to be a nearly sterile environment as gastric acid was thought to act as a protective barrier against microbial infection (44). Additionally, the underlying causes for the development of gastric disorders, such as peptic ulcers, were linked to diet, tobacco, alcohol, stress, and hyperchlorhydria (hyperacidity) (34, 44). In fact, acid suppression therapies, such as prescribed H₂ receptor antagonists, were found to effectively control symptoms of gastritis and help heal peptic ulcers (34, 44). Thus, the idea that gastric disorders arose because of personality type, dietary habits, or genetic predisposition seemed likely because lowering stomach acid secretion seemed effective against ulcers (34). As a result, the clinical adage “no acid, no ulcer” was becoming increasingly established in the minds of gastrointestinal specialists (44).

Contrary to excess hydrochloric acid being the sole cause of gastric disorders, there were several reports describing observations of spiral organisms in the stomach linings of patients diagnosed with peptic ulcers or gastric cancer (34). However, this “spirochaete” organism was deemed a postmortem contaminant by a leading United States gastroenterologist who was unable to observe any bacteria while examining over 1,000 gastric biopsies (34). A major breakthrough for the proponents of bacterial-influenced stomach disorders occurred in 1982 when *Helicobacter pylori*, originally termed *Campylobacter pylori*, was cultured from the gastric biopsies of patients suffering

from chronic gastritis and peptic ulcers (34, 44). This discovery sparked an immense interest in linking ulceration to bacterial infection (44).

Despite isolating bacteria from stomach biopsies of infected patients, a majority of scientists and physicians still did not accept that *H. pylori* might play a causative role in gastric disease progression (34). Many argued that Koch's postulates, frequently used to identify microbes as etiologic agents of specific diseases, could not be fulfilled because purified bacterial culture was not shown to reproduce disease in piglets (34). However, as there was no suitable animal model at the time, Dr. Barry Marshall ingested a pure culture of the bacteria and subsequently developed gastritis, which was confirmed endoscopically and histologically (34, 44). In light of confirming a microbial cause for peptic ulcers, antibiotics were shown to effectively eradicate *H. pylori* from patients with the disease, which also greatly reduced reoccurrence of ulceration (44). Accordingly, a 1994 National Institutes of Health (NIH) consensus conference agreed that *H. pylori*-induced ulcers should be treated with antimicrobial agents, which helped to widely establish a bacterial cause for peptic ulcer disease (44). Also in 1994, the World Health Organization (WHO) and the International Agency for Research on Cancer recognized the carcinogenic capabilities of *H. pylori* by designating it as the first class I bacterial carcinogen (67). Today, *H. pylori* is an acknowledged widespread human pathogen that infects nearly 50% of the world's population and it is linked to a variety of gastric disorders, including peptic ulcer disease and gastric cancers (22, 44, 48, 49, 67).

General *Helicobacter pylori* Biology

H. pylori is a gram-negative bacterium that is physically distinguishable by its spiraled cell shape and lophotrichous flagella (44, 53). Considering that the organism

requires oxygen, but at lower concentrations than those relative to the atmosphere, it is considered a microaerophile (44, 53). Being that *H. pylori* thrives in environments where pH is near neutrality, the organism has also been deemed neutralophilic (47, 71). This microbe is an etiologic agent of peptic ulcer disease, chronic active-type B gastritis, and non-ulcer dyspepsia (44, 46, 53). If untreated, infection persists for decades, increasing the risk of developing gastric malignancies such as mucosa-associated lymphoid tissue (MALT) lymphoma and adenocarcinoma (43, 46, 48). There is also some speculation that *H. pylori* may play a factor in the development of asthma, coronary artery disease, colonic adenomas, childhood growth retardation, diabetes mellitus, rosacea, and sudden infant death syndrome (SIDS), but these associations remain controversial (8, 44, 62). At least one positive association exists as epidemiological reports suggest *H. pylori* infection may be linked to decreased incidence of esophageal diseases, such as Barrett's esophagus and esophageal adenocarcinoma (8). Interestingly, there are two subsets of *H. pylori* strains based upon the presence or absence of a 40-kilobase genetic element known as the cytotoxin-associated gene pathogenicity island, *cag* PAI (5, 10, 12, 48). Type I strains, which carry the *cag* PAI, show increased virulence in patients and are typically associated with more severe gastric diseases than type II strains, which lack the *cag* PAI (5, 10, 12, 48). Accordingly, patients infected with type II strains of the bacterium tend to be asymptomatic carriers as a clear correlation between the *cag* PAI and severity of disease has been observed (5, 12).

Acid Acclimation of *Helicobacter pylori*

H. pylori is unique in that it exclusively colonizes the mucin layer of the gastric epithelium (44, 47, 71). Other bacterial neutrophiles, such as *Salmonella typhimurium*,

Vibrio cholera, *Yersinia enterocolitica*, and pathogenic *Escherichia coli* strains are able to survive transit through the human stomach, but fail to colonize this environment (36, 50). Interestingly, other *Helicobacter* species have also been found to colonize stomachs of other animals, such as several large wild feline species, domestic cats, and domestic dogs to name only a few (16, 60). Not surprisingly, *H. pylori* faces considerable pH changes during its colonization and establishment within the gastric mucosa, as luminal pH levels can vary from ~5.0 during digestion to ~1.0 during fasting (48, 49). Fittingly, many studies show that *H. pylori* has unique acid acclimation mechanisms that keep periplasmic and cytoplasmic pH levels near neutrality (48, 50, 56, 57, 71). These buffering mechanisms allow *H. pylori* to withstand severe acid shock and to grow at moderately low pH levels (47, 48). Furthermore, acid acclimation is reliant upon gene regulation, as several studies have shown that the transcription of 101 to 280 genes may be affected in response to low pH levels (47, 48, 70).

Central to the acid acclimation of *H. pylori* are the genes of the urease operon whose pH-induced transcription has been supported by transcriptome studies (2, 37, 48). Two subunits, UreA and UreB, compose the urease apoenzyme, which requires nickel for enzymatic activity (2, 48). This metal-containing enzyme hydrolyzes urea into carbon dioxide (CO₂) and ammonia (NH₃). CO₂ is subsequently converted to bicarbonate (HCO₃⁻) by a periplasmic α -carbonic anhydrase (33, 48). Both HCO₃⁻ and NH₃ play important roles in buffering the cytoplasm and periplasm of the cell (33). *H. pylori* can also control the influx of urea via a pH-gated inner membrane channel, UreI (48, 54, 55). This channel protein has two important functions in that it regulates the amount of urea entering the cell in response to acidic pH and that it inhibits the alkalization of the cell

under neutral conditions, as urease is still substantially expressed in such circumstances (33, 48). The necessity of urease and UreI for colonization of several infection models, such as piglets and gerbils, indicates the importance of this system (29, 38, 48). Aside from the urease system, the expression of ammonia-producing aliphatic amidases, AmiE and AmiF, have also been linked to the exposure of *H. pylori* to acidic pH (48, 70).

Compared to other bacterial species, *H. pylori* has relatively few transcriptional regulators, perhaps due to its exclusive association with the human stomach as persisting in such a restricted niche may reduce the number of required regulators (9, 48, 52, 66). Related to acid adaptation, this organism has at least three transcriptional regulators that respond to different primary signals (48). Two of which are metal-dependent transcriptional regulators NikR and Fur (47). NikR of *H. pylori* is an ortholog of an *E. coli* nickel (Ni^{2+}) responsive regulator, yet these orthologs have different functions in their respective organisms (48). In *H. pylori*, NikR is a regulator protein that can activate or repress transcription in response to the availability of Ni^{2+} (48). This regulator has been shown to control genes encoding the urease apoenzyme as well as genes involved with Ni^{2+} uptake and storage, iron III (Fe^{3+}) uptake and storage, stress response, motility, and outer membrane proteins (48). The other metal-dependent regulator, Fur, is important for regulating iron uptake and storage proteins (48). Even though both regulators are involved in the acid acclimation of *H. pylori*, neither is absolutely essential for the colonization of a gerbil model (11, 48). Aside from NikR and Fur, two-component signal transduction (TCST) also plays a substantial role in the acid acclimation abilities of *H. pylori* (47, 56, 71).

Two-component Signal Transduction Systems

Bacterial species live in a multitude of different environments and therefore need the means to sense variable surroundings and adapt accordingly. One of the ways this is achieved is through the utilization of two-component signal transduction systems. These environmental sensing systems are widespread in prokaryotes and can regulate cellular functions in response to specific conditions (5, 53, 56). Typically, these systems consist of a sensor protein that can recognize a signal and transduce it to an intracellular response regulator protein (17, 48, 53). The sensor proteins, known as histidine kinases (HKs), can recognize environmental stimuli through their variable N-terminal input domains that may be located extracellularly or periplasmically (Figure 1.1) (17, 48, 53). In the presence of the appropriate stimulus, an autophosphorylation may occur at a conserved histidine residue in the cytoplasmic, C-terminal transmitter domain of the HK (Figure 1.1) (17, 48, 53). Autophosphorylation of the HK is actually a bimolecular reaction that involves two HK molecules to form a homodimer, as one HK monomer catalyzes the phosphorylation of the second HK monomer (61).

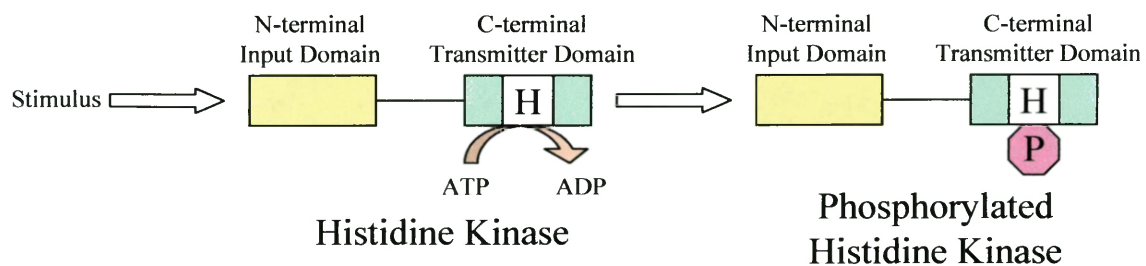


Figure 1.1: Generic model for histidine kinase autophosphorylation. When the N-terminal input domain of a histidine kinase (HK) senses a stimulus, an autophosphorylation event occurs at a conserved histidine residue in its C-terminal transmitter domain. In reality, this is a bimolecular reaction in which two HK monomers form a homodimer. Here, one HK monomer would catalyze the phosphorylation of another HK monomer.

Phosphorylated HK allows for the phosphorylation of a cognate response regulator (RR) (Figure 1.2) (17, 53). The RR catalyzes the transfer of the phosphoryl group from the HK as it becomes stoichiometrically favorable (61). Phosphorylation of the RR occurs at an aspartic acid residue located in a conserved N-terminal receiver domain of this protein (Figure 1.2) (17, 53, 61). Once phosphorylated, the variable C-terminal effector domain of the RR becomes active, which results in a particular response (Figure 1.2) (17, 53, 61). Typically, the C-terminal domain of the RR can bind to DNA, which allows for the regulation of transcription (17, 53). However, not all RR C-terminal domains can bind to DNA, as some have enzymatic function and some lack a C-terminal domain entirely (61).

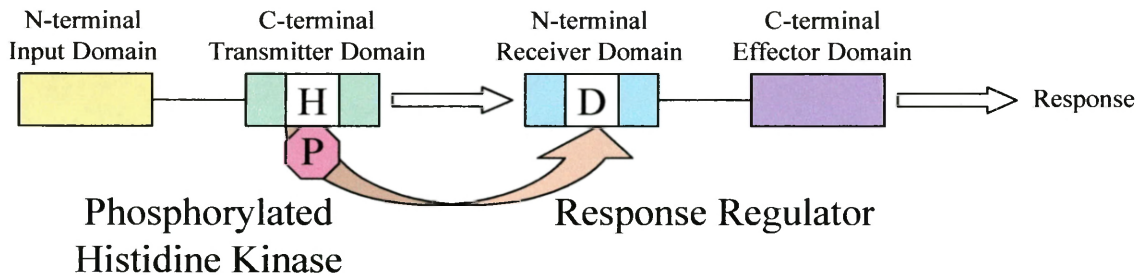


Figure 1.2: Generic model for response regulator phosphorylation. Phosphorylation of a response regulator (RR) occurs at a conserved aspartic acid residue within its N-terminal receiver domain. The RR catalyzes the transfer of a phosphoryl group from a phosphorylated histidine kinase (HK). Once the RR is phosphorylated, its C-terminal effector domain becomes active and a particular response may occur.

In terms of TCST systems, not all bacteria are created equally, but rather proportionally. For instance, *E. coli*, which is found in a number of different environments, encodes for 62 two-component proteins that are involved in complex and diverse tasks such as metabolism and motility (73, 74). However, *Streptococcus pneumoniae*, which is mainly a human pathogen, has only 27 two-component proteins (65, 68, 73). Conversely, *H. pylori* has a very limited repertoire of TCSTs compared to

these and other bacterial species, likely due to the fact that it is ecologically restricted to the human stomach and because it lacks competition with other microbes in this environment (5). In fact, *H. pylori* has only four HK and RR pairs as well as two orphan RRs, which do not seem to have an associated HK (5, 57). One of the two-component systems, CheA-CheY, appears to be dedicated to the regulation of chemotaxis, which ultimately affects the movement of the bacterium based on attractants or repellents in the surrounding environment (57). Interestingly, an additional system seems to be devoted to motility as genes for flagellar synthesis and several hypothetical proteins are regulated by a TCST system deemed FlgRS (39, 57). The final two *H. pylori* TCST systems appear to regulate genes in response to specific environmental factors. One of these has been implicated in the regulation of genes when environmental concentrations of copper ions are increased, and therefore has been designated as the copper resistance determinant system, CrdRS (68). The other has been shown to regulate genes in response to acid, and thus called the acid responsive signaling system, ArsRS (5).

Characterization of ArsRS

The two-component acid responsive signaling system of *H. pylori* is composed of the HK, ArsS, and cognate RR, ArsR (5, 47, 48). ArsS has been described as an orthodox HK composed of 414 amino acids (5, 47). This HK is considered orthodox because it functions as a periplasmic membrane receptor; as most, but not all, orthodox HKs operate similarly (5, 61). Appropriately, two short transmembrane segments, amino acids 8 to 22 and amino acids 134 to 153, of the ArsS N-terminal region suggest that it spans the cytoplasmic membrane, which locates the acid-responsive input domain of 111 amino acids in the periplasm of *H. pylori* cells (5, 45, 47, 48). Furthermore, ArsS

belongs to the IIIA class of sensor proteins, which generally means that this HK pairs with a RR that is related to OmpR of *E. coli* (5, 17). The RR, ArsR is composed of 225 amino acids and is OmpR-like in that it is a transcription factor that can function as a repressor or an activator to differentially regulate the expression of genes (5, 47, 61).

The N-terminal input domain of the HK ArsS senses periplasmic pH (45, 47, 49). In response to acidic conditions, an autophosphorylation at a conserved histidine residue within the ArsS C-terminal transmitter domain will occur through the dimerization of two ArsS proteins (Figure 1.1) (47, 61). Phosphorylated ArsS can subsequently phosphorylate the cognate RR, ArsR (Figure 1.2) (5, 47). ArsR likely regulates genes differentially in phosphorylated and non-phosphorylated states (5, 14). These two phosphorylation states of ArsR have been hypothesized to regulate genes differentially because gene knockouts of *arsR* are not viable *in vitro*, yet knockouts of *arsS* are (5). Furthermore, an *arsR* mutant strain of *H. pylori* that was incapable of phosphorylation, due to a substitution of the aspartic acid residue, was found to be viable *in vitro* (53). However, *arsS* knockouts inhibit the ability of *H. pylori* to colonize a mouse model (42). Together, these findings suggest that viability of the organism and differential gene regulation are linked to the two phosphorylation states of ArsR (5, 42, 53). Thus, non-phosphorylated and phosphorylated forms of ArsR are likely necessary *in vivo*.

There have been many attempts to characterize the genes under the control of ArsRS, also known as the ArsRS regulon (20, 31, 32, 46, 47, 48, 71, 72). One of the first studies found that phosphorylated ArsR negatively autoregulates itself, meaning that it binds at or near the *arsR* promoter to block RNA polymerase from initiating transcription (14). In this study, ArsR was also found to act as a transcriptional activator of several

genes with unknown function (14). Additionally, it was determined that *arsR*, *arsS*, and two other downstream genes were organized in an operon, or a series of genes that are co-transcribed on a single mRNA transcript (14). Two other studies, by the same group, subsequently demonstrated that the expression of both urease operons are under control of phosphorylated ArsR due to changes in pH (45, 46). This was determined by examining changes in gene regulation when *H. pylori* was exposed to neutral and acidic environmental pH. The study ultimately supported the hypothesis that ArsS undergoes autophosphorylation in response to acidic pH (45). Thus, the findings from these studies suggest that a functional ArsS sensory histidine kinase is necessary for at least one of the major acid acclimation mechanisms of *H. pylori*. ArsR has since been found to regulate the expression of genes that directly or indirectly play a part in other acid acclimation mechanisms as well, including genes for the periplasmic α -carbonic anhydrase (*hp1186*), aliphatic amidases (*amiE* and *amiF*), arginase (*rocF*), iron transportation (*hp0889* and *tonB2*), and a nickel-binding accessory protein (*hypA*) (6, 20, 47, 71, 72). Interestingly, variant ArsRS regulons have been identified and may diverge due to differences in individual strains of *H. pylori*. This assertion has been supported by several comparative transcriptional profiling studies for the ArsRS regulon of different *H. pylori* strains (20, 31, 32, 47, 71). Thus, the possibility exists that these differences in ArsRS regulons were observed merely because different strains of *H. pylori* were utilized in each of these transcriptional studies (20, 31, 32, 47, 71).

Genome Plasticity

Helicobacter pylori strain-specific ArsRS regulons would not be surprising as *H. pylori* strains have remarkable genetic variation, especially among the coding sequences

of particular genes (1, 24). Multiple techniques and methods, including whole genome sequencing, have been utilized in an attempt to characterize the population structure of specific *H. pylori* isolates and their differences in physiology (30). In fact, several studies have suggested that this organism is one of the most genetically diverse bacteria known. This is due in part to a more rapid rate of mutation and recombination compared to other microbes (63, 69). For instance, *Salmonella enterica* Typhimurium and *Salmonella enterica* Typhi, the causal agent of typhoid fever in mice and humans respectively, show about a 98-99% sequence similarity among these serological variants (serovars) (69). On the other hand, bacterial species from two separate genera, *S. enterica* Typhimurium and *E. coli*, show roughly an 85% sequence similarity (69). However, *H. pylori* typically exhibits a 5% difference among unrelated strains at the nucleotide level (69).

In fact, a comparison of the first two *H. pylori* strains whose genomes were sequenced in total, 26695 and J99, showed that there are 95 potential protein encoding open reading frames present in strain 26695 that are absent in strain J99 (3). Together, these sequencing projects have also recognized that *H. pylori* strains contain mobile genetic elements, such as plasmids and transposons, which may also lead to genomic differences among these strains (3, 66). Thus, genomic disparities among *H. pylori* isolates have been suggested to be at least partly due to horizontal transfer.

Horizontal transfers may occur readily as *H. pylori* strains appear to lack two essential double-strand break repair proteins involved in strand exchange during DNA recombination events (66). Thus, the horizontal transfer of DNA among *H. pylori* strains has been hypothesized to readily occur when patients are infected with multiple strains of

the bacterium (69). Multiple strain infection potentially sets up a scenario allowing for the transfer of DNA from strain-to-strain followed by genetic recombination within individual strains to subsequently alter their genomes (69). Thus, such recombination events could lead to a highly polymorphic population of *H. pylori* within an individual stomach (69).

Recombination alone is not likely to be the sole source of such variation in *H. pylori* genomes, especially since the mutation rates of this organism are predicted to be very high (69). Mutations can accumulate in *H. pylori* strains, as this organism appears to lack proteins necessary for the SOS response and proteins involved in mismatch repair (66, 69). These findings were mainly achieved through comparative analyses with the genome of *E. coli*. Compared to *E. coli*, *H. pylori* appears to lack three major components of the SOS response, which may repair nucleotide and recombination errors caused by physical or chemical mutagens (66, 69). Furthermore when compared to *E. coli*, *H. pylori* seems to lack two essential proteins involved in mismatch repair, which is involved in recognizing and repairing insertions, deletions, and the erroneous incorporation of nucleotides during DNA replication and recombination (66, 69).

The lack of functionality regarding these major DNA repair pathways coupled with the oxidative and acidic environment of the human stomach may assist in the frequent mutation rates of *H. pylori* (7, 30, 69). In fact, some have determined that the *H. pylori* mutation rate is so high that it may be close to being harmful for the bacterium (16, 30). However, *H. pylori* cells are naturally competent, or able to transport naked DNA from adjacent *H. pylori* cells and integrate it into their genomes (30). Therefore, the natural competence of *H. pylori* cells has been hypothesized to allow individual cells to

offset lethal mutations and furthermore add to diversification of the species (30). Thus, mutations and recombination events, along with natural selection, likely play key roles in allowing *H. pylori* to adapt to individual human stomachs.

Not surprisingly, there is much variation among the species of the *Helicobacter* genus. Interestingly, the closest relative of *H. pylori* appears to be *Helicobacter acinonychis*, which is found in large felines, such as tigers, cheetahs, and lions (16). There are large parts of the *H. pylori* and *H. acinonychis* genomes that are homologous, but their sequence differences are what led one group to hypothesize that *H. pylori* made a host jump due to a tragic event involving a big cat and an early human leading to a unique speciation scenario (16). If such an event did occur, current sequence data would indicate that the rate of genetic rearrangements and host-specific adaptations within *H. acinonychis* were 4-times greater than the rate of *H. pylori* in humans (30).

A specific means in which mutations may arise in the genomes of *Helicobacter* species is through long stretches of repeated nucleotides or polymeric tracts (15, 23). These sorts of repeated sequences may serve to increase antigenic variation, especially when disparities arise in the nucleotide sequence of genes encoding for proteins that can be recognized by the host (15, 23). For instance, a host-recognizable protein may be altered in its amino acid sequence by mutations originating in the polymeric tracts of its protein encoding sequence. Thus, a mutation occurring in the coding sequence of the gene may alter the amino acid sequence of the protein and render it undetectable by the host. Polymeric tracts may also act as transcriptional and translational regulatory mechanisms when present within the promoters regions or coding sequences of functional genes (15). Thus, variations occurring in these repeated regions may allow for

the evasion of host responses and/or adaptation to specific niches as long as these repeats are encoded by genes that can be expressed into proteins associated with such tasks (15). Accordingly, these sorts of variations involving polymeric tracts within the promoter and coding regions of genes are commonly exhibited in the genomes of *H. pylori*, *H. hepaticus*, and *H. canadensis* (15, 59).

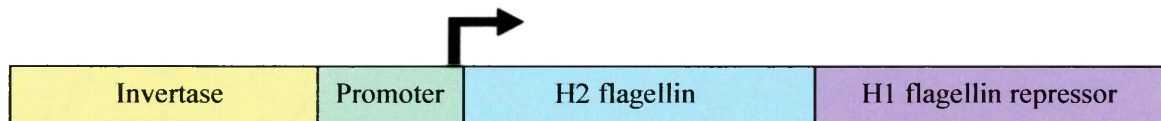
Phase Variation

Phase variation has been typically defined as a rapid means by which microorganisms can regulate gene expression in an “on or off” manner (25, 41). This gene “switch” is usually reversible and random, but arises relatively more frequently than normal mutations (25). One of the most common means of phase variation occurs by genetic recombination events. Phase variation occurring via genetic recombination is typically site-specific, meaning that the event takes place in a particular region of the genome, but it can also arise due to homologous recombination events (25, 41). A DNA inversion in a genomic region that encompasses a promoter is one example of a reversible genetic recombination event resulting in phase variation. Two of the best-characterized examples of such an inversion have been documented in the upstream regions of genes encoding for type 1 fimbriae of pathogenic *Escherichia coli* and flagella of *Salmonella* species (25, 58).

In *Salmonella*, phase variation causes two distinct phenotypes by variably switching expression of the flagellar protein from H1 flagellin to H2 flagellin and vice versa (Figure 1.3) (58). This “switch” in protein appeared unusual in that the rate of this mutation was much higher than the normal mutation rates of *Salmonella* and that it was entirely reversible (58). The key to the phase variation was that an invertible sequence

encoded an invertase, which is an enzyme that promotes the inversion of sequences, and a promoter of two other genes (58). The genes under the control of this invertible promoter encoded for the H2 flagellin protein and H1 flagellin repressor protein (58). Thus, when the invertible sequence was in one orientation, H2 flagellin protein and H1 flagellin repressor protein could be expressed such that H1 flagellin was not expressed (Figure 1.3A). In the other orientation, the promoter on the invertible element could neither promote the expression of the H2 flagellin protein nor H1 flagellin repressor protein (Figure 1.3B). Thus, this sequence orientation allowed for the expression of H1 flagellin (58).

A



B

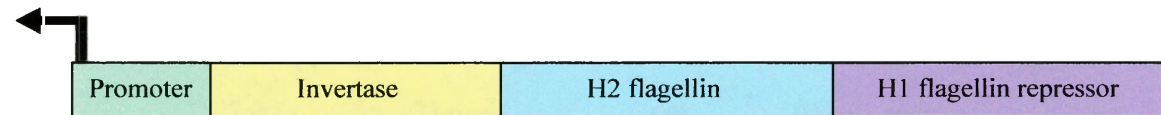


Figure 1.3: A model for *Salmonella* phase variation. Different phenotypes of *Salmonella* arise due to an invertible sequence consisting of a promoter and an invertase, which is an enzyme that promotes the inversion these particular genetic elements. This specific example of phase variation renders the flagellar phenotype of *Salmonella* to be either of the H1 or H2 type. (A) When the invertible sequence is in one orientation, the promoter can regulate the expression of H2 flagellin and the H1 flagellin repressor. Thus, in this orientation the cell produces H2 flagellin protein and represses the expression of H1 flagellin, which ultimately allows for the H2 phenotype. (B) In the opposite orientation, the expression of H2 flagellin and H1 flagellin repressor cannot be promoted. Thus, in this orientation the cell produces H1 flagellin because it cannot be repressed, which allows for the H1 phenotype.

Phase variation may also occur via slipped-strand mispairing due to repetitive DNA sequences in the genome. Slipped-strand mispairing occurs when template and nascent DNA strands do not align correctly during DNA replication, allowing for the

insertion or deletion of bases (Figure 1.4). These hypermutable repetitions may be homopolymeric tracts of single nucleotides or heteropolymeric tracts that contain two or more repeated nucleotides. Short sequence repeats (SSRs), as they are also known, have been considered as a mechanism of phase variation when they are located in the upstream regions of genes or within a gene itself, as frameshift mutations in these regions could regulate transcription or translation respectively (25). For example, repeats in a promoter region could modulate the binding affinity of RNA polymerase, whereas repeats embedded in the 5' end of a gene could lead to frameshifts and truncated proteins. Interestingly, phase variation due to slipped-strand mispairing events caused by repetitive nucleotides sequences have been observed in a number of bacterial species, including *Helicobacter pylori* (41).

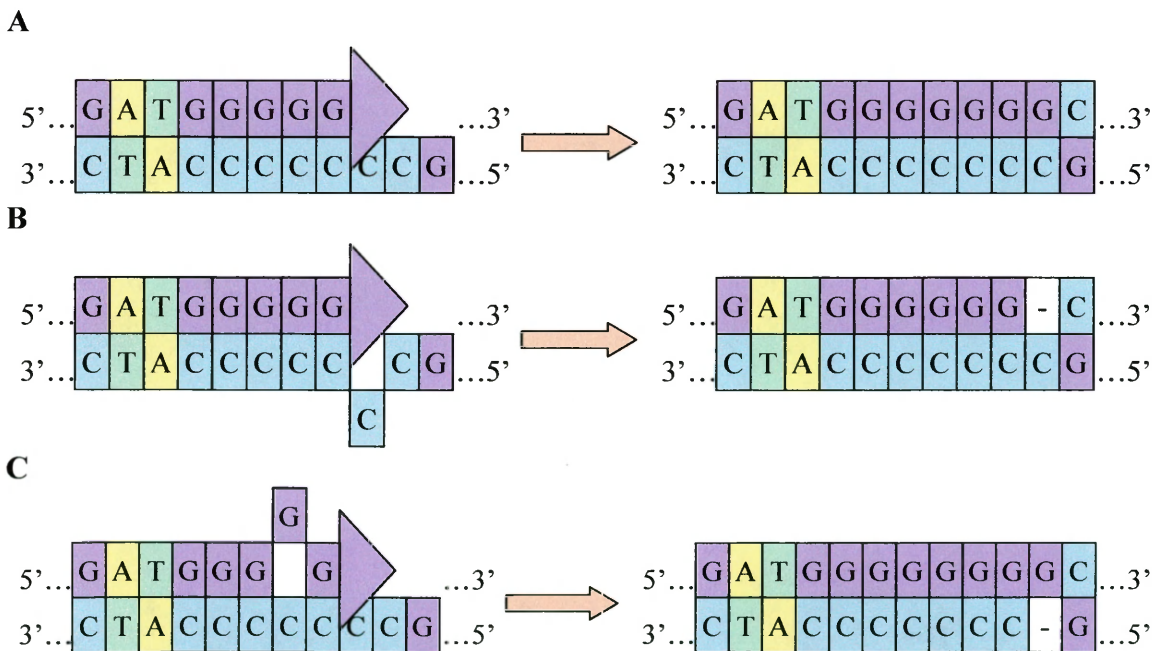


Figure 1.4: Generic model for slipped-strand mispairing. Slipped-strand mispairing occurs when template and nascent DNA strands do not align correctly during DNA replication. Mutations resulting from slipped-strand mispairing events cause deletions or insertions in the nascent strand. If located within gene encoding sequences, these slippages may cause frameshift mutations to occur. Here, three scenarios are depicted involving normal replication (A), a slippage in the template strand (B), and a slippage in

the nascent strand (C). Colored blocks represent different nucleotides and white blocks represent gaps, which are just an indication of a deletion or an insertion. (A) The correct alignment of template and nascent strands allows for normal replication to occur. (B) A slipped-strand mispairing event in the template strand causes for a deletion in the nascent strand, which is indicated by a gap. (C) A slipped-strand mispairing event in the nascent strand causes for an insertion in this strand, which is indicated by an extra guanine.

H. pylori encodes for a number of genes that may be phase variable due to slipped-strand mispairing events, including genes encoding for enzymes that are involved in the biosynthesis of outer membrane structures such as lipopolysaccharides (30). The coding sequences of outer membrane proteins, such as porins and adhesins, have also been shown to be phase variable in *H. pylori* strains (22, 30, 51, 75). The phase variability of these genes due to repeated nucleotide sequences, or SSRs, likely give *H. pylori* an advantage in the selection or evolution of particular phenotypes when environmental conditions or host-bacterial interactions change (30). Interestingly, the lengths of SSRs among *H. pylori* strains 26695 and J99 have been found to be different, suggesting that gene expression in these strains may deviate due to strain-specific phase variation (3, 66).

For instance, the *H. pylori* adhesin gene *sabA* may be under transcriptional and translational regulation by phase variation involving repeated nucleotide tracts. Near the promoter region of *sabA* there is a homopolymeric thymine repeat and within the coding region of this gene there is a heteropolymeric cytosine-thymine repeat (22, 75). For transcriptional control, the length of the homopolymeric thymine repeat is hypothesized to influence the binding affinity of RNA polymerase, changing the rate and/or efficiency of transcription initiation for *sabA* (Forsyth, unpublished). For translational control, the number of heteropolymeric cytosine-thymine repeats influences the reading frame of the transcript and can thus regulate the expression of full length SabA (22). Given the level

of divergence at the nucleotide level for individual strains of *H. pylori*, *sabA* is likely expressed differently among strains (69).

Polymeric tracts are found in other species of *Helicobacter* as well (15, 59). Most studies to date appear to have focused upon polymeric tracts near or within promoter regions and within the 5' regions of protein encoding genes because of their potential to affect gene regulation through phase variation. However, a recent study with *Helicobacter canadensis* reflected upon homopolymeric tracts near the 3' region of genes (59). In this study, they found eight coding regions that had polyguanine tracts near the end of the predicted genes (59). As a side note, most of the genes that were strongly considered to be phase variable, because of the proximity of the homopolymeric tract in their 5' region, were also due to polyguanine repeat sequences (59). This study suggested that the C-terminal region of particular proteins may have been altered due to frameshift mutations caused by polyguanine repeats in the 3' terminus of their coding sequences (59). Based on data from the NCBI Conserved Domain Database (CDD), these authors concluded that frameshifts due to such mutations may lead to structural differences in the resulting proteins (59). For instance, polyguanine-induced frameshifts, truncating the amino acid sequence of a particular enzyme of *H. canadensis*, were hypothesized to ultimately affect the structure of the protein because of the possible loss of an α -helix at its C-terminus (59). Another polyguanine-induced frameshift was hypothesized to affect the structure of a flagellar protein as changes in amino acid sequence length would possibly affect a β -sheet in its C-terminus (59). Thus, the authors suggested that polymeric repeats in the 3' regions of coding sequences likely have effects on protein

structure (59). Thus, these allegations should prompt similar investigations in other *Helicobacter* species, especially the human gastric pathogen *H. pylori*.

This Study

In *H. pylori*, the gene that encodes for the sensory histidine kinase ArsS exhibits sequence variation due to a 3' homopolymeric cytosine repeat that differs in length among distinct strains (3, 5, 66). In fact, potentially different lengths of the *arsS* homopolymeric cytosine tract were previously observed in strain J99 when whole genome shotgun sequencing was initially performed, but this research group stated that these mutations may have occurred while culturing the bacteria *in vitro* (3). Frameshifts corresponding to variation in polymeric repeats have previously been shown to be a cause of phase variation in bacteria, especially when located in or near promoter regions or within the 5' regions of coding sequences. However, homopolymeric repeats located in the 3' region of *H. canadensis* coding sequences have been hypothesized to have potential effects on protein structure (59). In *arsS* of *H. pylori*, different alleles, attributed to cytosine repeat length, have been shown to be capable of expressing alternate functional ArsS isoforms *in vitro* (5). These investigators determined that each ArsS isoform was able to autophosphorylate and subsequently donate phosphoryl groups to the cognate response regulator protein, ArsR (5). Thus, this group has suggested that the C-terminal protein sequence of ArsS does not interfere with its kinase activity or with its phosphotransfer reaction to ArsR (5).

In spite of these contentions, this thesis research examined the 3' region of the *arsS* histidine kinase gene to determine whether different polycytosine tract lengths could be detected within and among *H. pylori* populations. Additionally, we aimed to

determine whether single strains of *H. pylori* could potentially express individual ArsS isoforms due to various *arsS* polycytosine tract lengths. Here, we utilized thirty-six clinical populations of *H. pylori* from three anatomically distinct gastric regions of twelve patients. Genomic DNA extracted from these *H. pylori* populations was used to screen a region of *arsS* with amplified fragment length polymorphism (AFLP) analyses to determine if differences in cytosine repeat lengths existed within and among the clinical populations. Our results suggested that variable lengths of *arsS* could be amplified and detected among and within these *H. pylori* populations. AFLP data also indicated the relative frequency for each of the variable *arsS* lengths within each clinical population. To verify that the differing lengths detected via AFLP were due to differences in the polycytosine tract length of *arsS*, we sequenced clones encoding for the same *arsS* region investigated in these AFLP analyses.

Together, AFLP and sequencing analyses conferred that various *arsS* alleles, due to differences in the length of a 3' polycytosine tract, existed within and among each clinical *H. pylori* population. Thus, AFLP data indicative of variant *arsS* alleles were utilized for the extrapolation of allele frequencies into open reading frame frequencies, which were representative of ArsS isoforms. In addition to confirming that differences in polycytosine tract lengths existed, sequencing data were also used to estimate the predominant polycytosine length of each population, the number of strains present in each population, and if predominant strains existed within polyclonal populations. Interestingly, sequencing data also shed light on another way *arsS* is genetically variable, as a new isoform of ArsS was recognized.

Chapter 2

Materials and Methods

Bacterial Strains and Growth Conditions

Clinical populations of *H. pylori* (Table 2.1) used in this thesis research were collected via gastric biopsy from the antrum, cardia, and corpus of twelve patients that visited gastrointestinal clinics at the Veterans Administration Hospital or the Vanderbilt University Medical Center, both located in Nashville, Tennessee. These populations were cultured and frozen stocks were given to Dr. Mark H. Forsyth as gifts from Dr. Richard M. Peek, Jr. and Judith Romero-Gallo. At William and Mary, these *H. pylori* populations were cultured on Trypticase Soy Agar II plates supplemented with 5% sheep blood (BBL). These bacterial cultures were incubated in a humidified environment at 37°C and 5% CO₂. The *H. pylori* populations were then passed onto fresh blood agar plates every 24 to 48 hours to ensure adequate cellular growth for the production of an additional frozen stock and the extraction of genomic DNA (gDNA).

Table 2.1: Clinical populations of *Helicobacter pylori*

Patient Number	Gastric Geography	Description^a	Source
B215	Antrum, Cardia, Corpus	3 Populations	R.M. Peek
B221	Antrum, Cardia, Corpus	3 Populations	R.M. Peek
B253	Antrum, Cardia, Corpus	3 Populations	R.M. Peek
B256	Antrum, Cardia, Corpus	3 Populations	R.M. Peek
B266	Antrum, Cardia, Corpus	3 Populations	R.M. Peek
B268	Antrum, Cardia, Corpus	3 Populations	R.M. Peek
B284	Antrum, Cardia, Corpus	3 Populations	R.M. Peek
B292	Antrum, Cardia, Corpus	3 Populations	R.M. Peek
B294	Antrum, Cardia, Corpus	3 Populations	R.M. Peek
B295	Antrum, Cardia, Corpus	3 Populations	R.M. Peek
B300	Antrum, Cardia, Corpus	3 Populations	R.M. Peek
B301	Antrum, Cardia, Corpus	3 Populations	R.M. Peek

^a Collected via gastric biopsy at the Veterans Administration Hospital (Nashville, TN) or the Vanderbilt University Medical Center (Nashville, TN).

General Techniques

Production of Bacterial Frozen Stocks

The clinical *H. pylori* populations that had been passed either once or twice from original frozen stocks were grown overnight for the production of additional frozen stocks. The following day, the cultures were harvested and suspended in sulfite-free brucella broth (SFBB) supplemented with 15% (v/v) glycerol. These suspensions were kept in cryogenic vials at -80°C for long-term preservation.

Preparation of Helicobacter pylori genomic DNA (gDNA)

Preparation of *H. pylori* gDNA was adapted from a cetyltrimethylammonium bromide (CTAB) extraction protocol that the Forsyth laboratory uses regularly (35, 41). A 48-hour *H. pylori* culture that had been typically passed once, but no more than five times from original frozen stock, was harvested from blood agar and suspended into sterile 0.9% NaCl solution. These cells were centrifuged at 6,000 revolutions per minute (rpm) for 10 minutes, resuspended in elution buffer (EB; 10mM Tris, pH 8.0), and subsequently lysed with 0.5% sodium dodecyl sulfate (10% SDS stock) at 37°C for 30 minutes. gDNA was extracted by the addition of 0.7M NaCl (5M NaCl stock) and 0.125 volume CTAB-NaCl (10% CTAB-0.7M NaCl stock) at 65°C for 10 minutes. gDNA was purified from bacterial lysates with a series of sequential 1:1 volume extractions with chloroform:isoamyl alcohol (24:1) (CHCl_3), phenol:chloroform:isoamyl alcohol (25:24:1) (P:C:I), and again CHCl_3 . gDNA was precipitated with 0.86 volume of isopropanol (99% v/v) and the gDNA pellet was air dried to be resuspended in 100 μL EB. gDNA concentration was measured with a NanoDrop 1000 (Thermo Scientific).

These gDNA extractions were performed three times for each clinical population. As a general note, each of these extractions is referred to as gDNA template 1, 2, or 3.

DNA Amplification

Genomic regions were amplified via Expand High Fidelity PCR System (Roche). In general, each 50 μ L reaction was prepared as follows: 1X Expand High Fidelity buffer with 1.5mM MgCl₂ (Roche), 0.8mM deoxynucleotide triphosphates (0.2mM dATP, 0.2mM dCTP, 0.2mM dGTP, and 0.2mM dTTP), 2.6 U of Expand High Fidelity enzyme mix (Roche), 10 μ M of forward and reverse primers, 200ng of gDNA, and the appropriate volume of sterile deionized water (sdH₂O). Amplification was performed with an ABI 2720 Thermal Cycler (Applied Biosystems) or a MultiGene Thermal Cycler (LabNet International) and conditions are listed where necessary. Amplicons generated via PCR were always verified with agarose gel electrophoresis.

Agarose Gel Electrophoresis

All agarose gels were prepared by dissolving agarose (1.0% w/v or 1.2% w/v) in 0.5X TAE buffer. Gels were supplemented with ethidium bromide (0.05% v/v) to allow visualization under ultraviolet light with the Gel Logic 100 Imaging System (Kodak) and Molecular Imaging Software version 4.5 (Kodak). Electrophoresis voltage and running time were dependent upon desired resolution and degree of DNA separation.

Amplified Fragment Length Polymorphism (AFLP) Analyses of *arsS* Regions

Amplified Fragment Length Polymorphism Polymerase Chain Reactions (AFLP-PCR)

Three separate gDNA templates from each clinical *H. pylori* population (Table 2.1) were utilized to amplify a genomic region of roughly 300 base pairs encoding portions of the 3' *arsS* terminus and the 5' terminus of a downstream gene, *hemB*. This

region also encoded for the suspected polymorphic cytosine tract of *arsS*, thus we refer to it as the poly C region. In general, the Expand High Fidelity PCR system was utilized as previously described with a 5' fluorescently labeled forward primer, *arsS* F-1 FAM, and an unlabeled reverse primer, *hemB* R-1 (Table 2.2). Amplification conditions consisted of a 2-minute hot start at 94°C, 30 cycles of 94°C for 30 seconds, 60°C for 30 seconds, and 72°C for 30 seconds, followed by a final extension at 72°C for 7 minutes. AFLP-PCR was performed three times for each gDNA template for a total of nine trials for each clinical population.

As a control, one set of gDNA templates for *H. pylori* populations from patients B284 and B294 (Table 2.1) were utilized to amplify a genomic region of 248 base pairs that encoded an upstream region of *arsS* relative to the suspected polymorphic cytosine tract. Considering this genomic region is not polymorphic in length, we refer to it as the control region. To control for various gDNA templates, gDNA templates 2, 3, and 1 were used from the antrum, cardia, and corpus *H. pylori* populations of patient B284, respectively. Only template 1 gDNA was used for the antrum, cardia, and corpus *H. pylori* populations of patient B294. In general, the Expand High Fidelity PCR system was utilized as previously described with a fluorescently labeled forward primer, HK0165 Fwd FAM, and an unlabeled reverse primer, HP0165 Rev (Table 2.2). Control region amplicons were achieved with identical amplification conditions as used for the poly C region AFLP-PCR. Here, three AFLP-PCR trials were executed for each gDNA template utilized.

Table 2.2: Oligonucleotide primers

Oligonucleotide Name	Sequence	Modification
arsS F-1	5'-CTTCTAACCCCAGCCAAGCCCATGG-3'	
arsS F-1-FAM	5'-CTTCTAACCCCAGCCAAGCCCATGG-3'	5' 6-FAM
hemB R-1	5'-CGCTGCTTCGTAATCTTCTCAATCG-3'	
HK0165 Fwd	5'-CGCACCCCTATCACTAAGGGCAAG-3'	
HK0165 Fwd-FAM	5'-CGCACCCCTATCACTAAGGGCAAG-3'	5' 6-FAM
HK0165 Rev	5'-GAAGAGGATACATGGATAGGGC-3'	

Amplified Fragment Length Polymorphism (AFLP) Analyses

AFLP-PCR product was diluted 1:100 in sdH₂O (3μL AFLP-PCR product into 297μL sdH₂O). 1μL of 1:100 diluted AFLP-PCR sample was added to 12μL Hi-Di Formamide (Applied Biosystems) with 0.25μL GeneScan ROX500 Size Standard (Applied Biosystems) in a 96-well plate. By loading the entire sample plate into a thermal cycler, the samples were then denatured at 95°C for 3 minutes. All samples were processed at the College of William and Mary.

Differing AFLP-PCR amplicon sizes were detected with an ABI 3130 Genetic Analyzer (Applied Biosystems) and the resulting AFLP data were interpreted with GeneMapper version 4.0 (Applied Biosystems). An example of raw AFLP data is shown in Figure 2.1. Here, differences in amplicon lengths are measured in base pairs, which are indicated on the x-axis (Figure 2.1). Due to these amplicons being tagged with fluorescent primers, their quantities were measured with relative fluorescence units (RFU), which are indicated on the y-axis (Figure 2.1). Thus, using data from the x-axis (fragment length in base pairs) and the y-axis (fluorescence intensity in RFU), GeneMapper generated an area under the curve for each fragment length (Figure 2.1). All of the AFLP data indicative of various fragment lengths and their respective areas were utilized in this project.

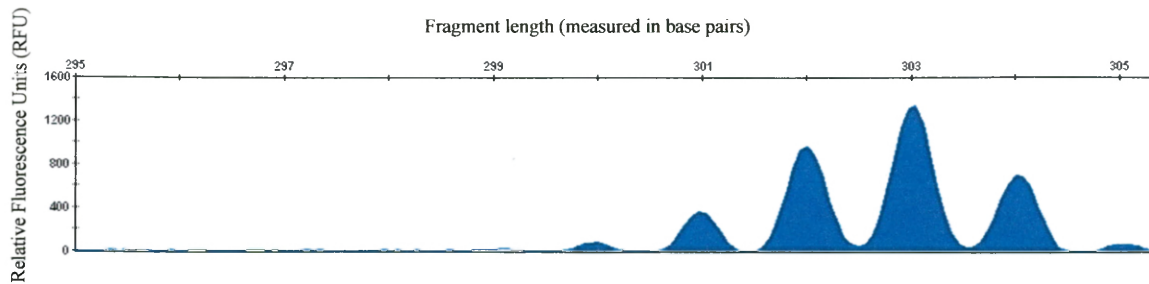


Figure 2.1: Example of AFLP raw data. AFLP-PCR amplicons were detected with an ABI 3130 Genetic Analyzer and the resulting AFLP data were interpreted with GeneMapper software. Here, amplicon lengths, measured in base pairs, are represented on the x-axis. The quantity of each particular amplicon was indicated with relative fluorescence units on the y-axis. Together, GeneMapper generated an area under the curve by utilizing amplicon lengths and their relative fluorescent units. These areas allowed for the relative frequency of each amplicon to be calculated. For instance, fragment length 303 was the most common amplicon generated from this particular AFLP-PCR trial.

AFLP Data Analysis

AFLP data produced with GeneMapper software were analyzed with Microsoft Excel X for Mac (Microsoft). Within each AFLP-PCR trial, the total area for each fragment length was summed (Table 2.3) to ultimately determine their relative frequencies. Thus, fragment length frequencies were calculated by dividing individual fragment length area by the total area of all fragment lengths. Frequencies were then multiplied by one hundred to find frequency percentages. These analyses were performed for each of the nine poly C region AFLP-PCR trials generated from the gDNA of each clinical *H. pylori* population. These were also performed for the three control region AFLP-PCR trials generated from the gDNA of clinical populations from patients B284 and B294.

Table 2.3: Calculating AFLP fragment length frequencies

Fragment Length (bp) ^a	Area Under the Curve ^b	Frequency (%) ^c
300	387	1.70
301	2011	8.85
302	6276	27.61
303	9164	40.32
304	4598	20.23
305	293	1.29
Total	22729	100

^a Fragment lengths, measured in base pairs, were generated with AFLP-PCR from the gDNA of clinical *H. pylori* populations. These amplicons were detected with an ABI 3130 Genetic Analyzer and interpreted with GeneMapper software.

^b GeneMapper produced an area under the curve for each amplicon by utilizing AFLP data representative of fragment lengths and their relative fluorescence units. These areas were summed to calculate the relative frequency of each fragment length.

^c The relative frequency of each amplicon length was calculated by dividing individual fragment length area by the total area of all fragment lengths. These figures were then multiplied by one hundred to determine final frequency percentage.

To integrate the AFLP data from AFLP-PCR trials performed for each *H. pylori* population, the frequencies of corresponding fragment lengths were averaged to find overall fragment frequencies. Standard deviations for overall fragment frequencies were also calculated by using the frequencies of corresponding fragment lengths from each of the AFLP trials.

Furthermore, each poly C region fragment length was considered to be an individual allele of *arsS*. Thus, fragment lengths that varied by three base pairs, could be considered alleles that encoded for the same open reading frame (ORF). For instance, fragment lengths 300 and 303, 301 and 304, and 302 and 305 would encode for the same ORFs, respectively. Thus, overall poly C region fragment length data were extrapolated into potential ORFs by summing the frequencies of fragment lengths that were divergent by three or six base pairs. The standard deviation of each *arsS* ORF was calculated by summing the variances for each fragment used in the determination of ORF frequency

followed by taking the square root of this figure ($\sigma_{x+y \dots +n} = \sqrt{(\sigma_x^2 + \sigma_y^2 \dots \sigma_n^2)}$). To determine if particular *arsS* ORFs were significant within gastric populations, a Student's T-test was performed to test the three ORFs amongst one another.

Sequencing Analyses of *arsS* Regions

Generation of arsS Poly C and Control Region Amplicons for Cloning and Sequencing

gDNA from clinical *H. pylori* populations (Table 2.1) was utilized to amplify a genomic region of *arsS* that encodes the suspected polymorphic cytosine tract as explained in the protocol for AFLP-PCR. As a control, one set of gDNA templates for gastric populations from patient B284 and B294 (Table 2.1) were utilized to amplify a genomic region that encodes an *arsS* region upstream of the suspected polymorphic cytosine tract, as previously described in the protocol for AFLP-PCR as well. Poly C region amplicons were amplified with unlabeled forward, *arsS* F-1, and reverse, hemB R-1, primers (Table 2.2). Control region amplicons were also produced with unlabeled forward, HK0165 Fwd, and reverse, HP0165 Rev, primers (Table 2.2). The PCR system and amplification conditions were identical to those described in the protocol for AFLP-PCR.

Cloning of arsS Poly C and Control Region Amplicons

arsS poly C and control region amplicons were cloned with pCR-Blunt II-TOPO (Invitrogen) vector by the sequential addition of 2 μ L of fresh (0 – 2 days) PCR product, 0.2M NaCl with 0.01M MgCl₂ salt solution, 10ng of pCR-Blunt II-TOPO, and sdH₂O in 6 μ L reactions. Cloning reactions were incubated at 22°C (ambient room temperature) for 10 minutes. Alternatively, amplicons were similarly cloned with pCR2.1-TOPO (Invitrogen) vector.

Transformation of arsS Poly C and Control Region Plasmids into Escherichia coli DH5 α

Transformation of the *arsS* poly C and control region plasmids into OneShot Chemically Competent *E. coli* DH5 α (Invitrogen) was initiated by gently stirring 2 μ L of the cloning reactions into a 100 μ L culture of *E. coli* DH5 α . This cell and plasmid mixture was incubated on ice for at least 30 minutes, as increasing this incubation time tended to increase overall transformation efficiency. To induce the cellular uptake of plasmid DNA, the culture was heat shocked at 42°C for 30 seconds and immediately allowed to recover on ice for at least 1 minute. To allow for plasmid generation, outgrowth was initiated with the addition of 250 μ L of super optimal broth with catabolite repression (SOC). The cultures were then incubated at 37°C with horizontal shaking of at least 200 rpm in an Innova 43 shaker (New Brunswick) for 1.25 hours. Successful transformants were selected by spread plating 25 μ L of outgrowth onto Luria-Bertani (LB) agar complemented with kanamycin (50 μ g/mL) followed by an overnight incubation at 37°C.

arsS Poly C and Control Region Plasmid Preparations

An *E. coli* DH5 α colony grown on selective media was aseptically inoculated into 5mL LB broth with kanamycin (50 μ g/mL). This culture was incubated overnight at 37°C in a rotator at 70% power (Glas-col). The following day, the culture was centrifuged at 4,000 rpm for 10 minutes. Plasmids were then purified via the QIAprep Spin Miniprep Kit (Qiagen) or the IBI High-Speed Plasmid Mini Kit (MidSci). Purified plasmid DNA was stored at -20°C until sequencing reactions were performed. In total, ten poly C region clones from each gastric population, or 30 per patient, were prepared in

this manner. As a control, ten control region clones from the populations of patient B284 and B294 were also prepared in this manner.

Screening for Successful Poly C and Control Region Plasmids

To verify the successful cloning and transformation of poly C and control region plasmids, 5 μ L of purified plasmid was digested with 15U *EcoRI* (Promega) in 1X Buffer H (Promega) and sdH₂O in 12.5 μ L reactions. Restriction digests were incubated in a 37°C water bath for at least 1.5 hours. The entire digest was subjected to agarose gel electrophoresis in order to determine successful cloning of the desired fragment into the vector.

Poly C and Control Region Plasmids Sequencing Reactions

Forward and reverse sequences for each poly C or control region plasmid were generated with BigDye Terminator version 3.1 (Applied Biosystems). In general, each 20 μ L sequencing reaction was prepared as follows: 1X Big Dye Terminator v1.1/v3.1 Sequencing Buffer, 0.25X Big Dye Terminator v3.1 Sequencing Reaction Mix, 10 μ M of T7 (20mer; Promega) or Sp6 (19mer; Promega) promoter primer, 10 μ L of purified plasmid DNA, and the appropriate volume of sdH₂O. Sequencing reactions were performed in a thermal cycler and conditions consisted of a hot start at 94°C for 5 minutes followed by 26 cycles of 94°C for 45 seconds, 50°C for 30 seconds, and 60°C for 4 minutes.

DTR Gel Purification of BigDye v3.1 Sequencing Reactions

Excess dye was removed via Performa Dye Terminator Removal (DTR) Gel Filtration Cartridges (Edge Biosystems). Each cartridge was centrifuged at 3,000 rpm for 2 minutes to dry the filter and subsequently transferred to a fresh centrifuge tube, in

which the 20 μ L sequencing reaction was added. The DTR column with sequencing reaction was centrifuged again at 3,000 rpm for 2 minutes and resulting filtrate was vacuum dried for 1.25 hours.

Analysis of Sequencing Reactions

Dried, purified sequencing reactions were resuspended in 12 μ L Hi-Di Formamide and transferred to a 96-well plate. Sample plates were placed in a thermal cycler for 3 minutes at 95°C to allow for sample denaturation. Sequencing samples were processed at the College of William and Mary with an ABI 3130 Genetic Analyzer. Base calling was accomplished with SequencingAnalysis version 5.2 (Applied Biosystems) software. Sequencing analyses were performed with MacVector version 9.0 (Accelrys, Inc.), 4Peaks version 1.7.2 (Mek&Tosj.com), and WebLogo version 3.0 (Threeplusone.com).

Helicobacter pylori Strain Designation from Sequencing Data

From the nine *H. pylori* genome sequences recognized by the National Center for Biotechnology Information (NCBI), we found that no two strains were identical in the variable poly C region of roughly 300 base pairs or the control region of 248 base pairs (3, 4, 19, 26, 27, 28, 40, 64, 66). The most closely related *H. pylori* strains regarding the poly C region were 26695 and HPAG1, which differed by two nucleotides outside of the polycytosine tract (40, 66). For the control region, the most closely related *H. pylori* strains were 51 and Shi470, which exhibited a single nucleotide polymorphism in this region (26, 27). Thus, one base pair differences, when disregarding polycytosine tract length polymorphisms, were our metric for the denotation of individual *H. pylori* strains within specific gastric populations. However, if deletions were present in the primer

binding regions of the poly C or control region cloned sequences, these polymorphisms were not considered evidence for a difference in *H. pylori* strain.

Chapter 3

Results

***arsS* Poly C Region AFLP-PCR Amplicons Vary in Length Among and Within Clinical *Helicobacter pylori* Populations**

Amplified fragment length polymorphism polymerase chain reactions (AFLP-PCR) were utilized to determine whether homopolymeric cytosine tracts of *arsS* differed in length among and within populations of *H. pylori*. An experimental and a control region of *arsS* were amplified in these reactions and analyzed via AFLP. The experimental poly C region encoded portions of the 3' *arsS* terminus and the 5' terminus of an essential downstream gene, *hemB*, (Forsyth, unpublished) as well as the cytosine tract of *arsS*. Our data suggest that different fragment lengths of the *arsS* poly C region could be amplified via AFLP-PCR from each of the clinical *H. pylori* populations investigated and detected via AFLP analyses.

By averaging corresponding AFLP data generated from the nine AFLP-PCR trials performed, bar graphs representing *arsS* poly C region amplicon lengths and relative frequencies were produced. For example, AFLP data generated from the population infecting the antrum of patient B256 are shown in Figure 3.1. Differences in *arsS* poly C region fragment lengths, measured in base pairs, are represented on the x-axis. The relative frequency of each fragment length, measured in percentage, is represented on the y-axis. Thus, the *arsS* poly C region fragment length of 303 base pairs is amplified most readily from the gDNA of this particular *H. pylori* population, as this amplicon is representative of 38.34% of all fragments produced (Figure 3.1). Furthermore, fragment lengths of 302 and 304 base pairs are the next two most prevalent fragments generated as

they represent 29.13% and 18.16% of all amplicons, respectively (Figure 3.1). Standard deviations calculated from the relative fragment frequencies of all AFLP trials are represented by error bars (Figure 3.1). Standard deviations represent relative fragment length frequency variations among the nine AFLP trials. Similar bar graphs representing the *arsS* poly C region AFLP data from all clinical *H. pylori* populations are located in Appendix 2A, AFLP Data from Poly C Regions of *H. pylori* Populations.

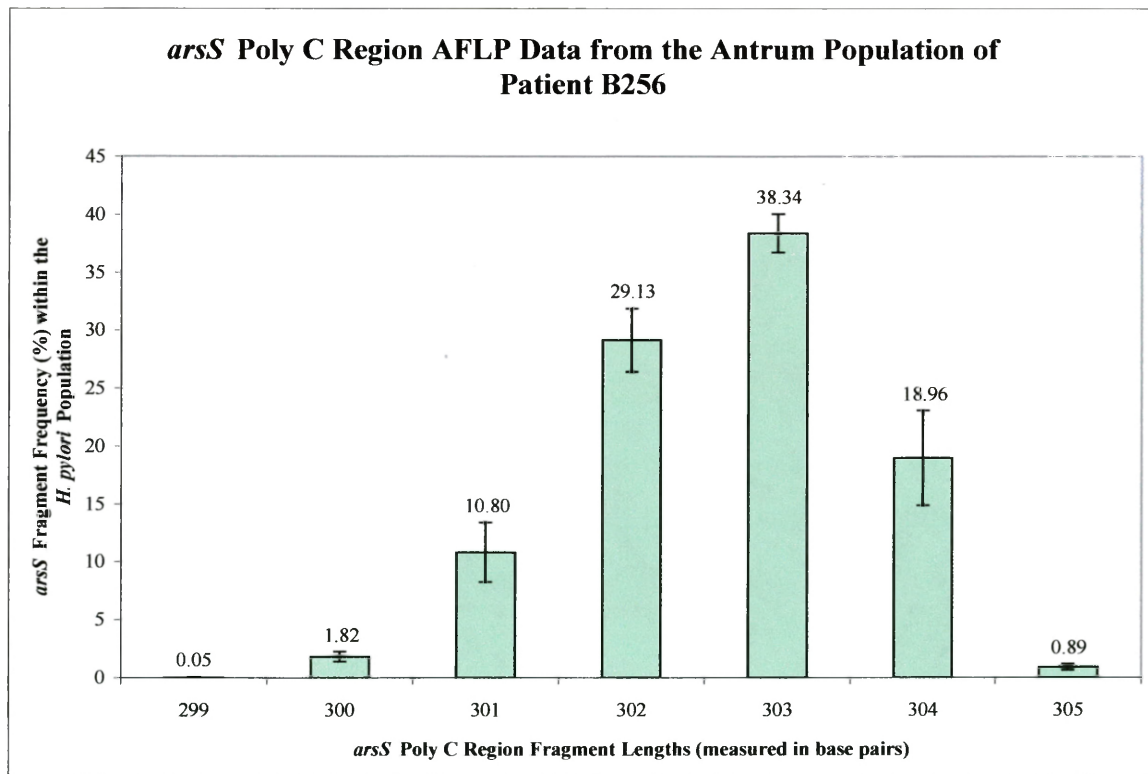


Figure 3.1: *arsS* poly C region AFLP data from the antrum *H. pylori* population of patient B256. This bar graph was developed with AFLP data generated from the *H. pylori* population derived from the antrum of patient B256. The x-axis represents various *arsS* fragment lengths, measured in base pairs, which were produced via AFLP-PCR. The y-axis represents the relative frequency of each fragment length produced from this particular population. Error bars indicate the standard deviations of the various fragment lengths, which are representative of relative fragment length frequency variations among the nine AFLP trials.

Predominant *arsS* Poly C Region AFLP-PCR Amplicons Vary in Length Among Clinical *Helicobacter pylori* Populations

AFLP results indicated that the presence of multiple *arsS* poly C region amplicons was a characteristic of all *H. pylori* populations, but the most predominant fragment length was not always similar among populations (Table 3.1). The total number of different poly C region fragments detected via AFLP ranged from six to nine indicating that multiple amplicons could be generated from the gDNA of each clinical *H. pylori* population. Among all thirty-six *H. pylori* populations, fragment length 302 was the most predominant in thirteen populations (36.1%) and fragment length 301 was the most predominant in nine populations (25.0%). Furthermore, populations among the different gastric regions of five patients exhibited the same predominant poly C region fragment length (Table 3.1). However, the populations among the different gastric regions of the other seven patients had variable predominant poly C region fragment lengths (Table 3.1).

Table 3.1: Summary of poly C region AFLP data from *H. pylori* populations

Patient Number and Gastric Region of <i>H. pylori</i> Population^a	Number of <i>arsS</i> Poly C Region Fragments Detected^b	Predominant <i>arsS</i> Poly C Region Fragment Length (bp) Detected^c	Predominant <i>arsS</i> Frame Derived from Fragment Data^d	Frequency of <i>arsS</i> Frame (%)^e	Significance of <i>arsS</i> Frame (p-Value)^f
B215 Antrum	7	301	1	44.88	<0.001
B215 Cardia	7	301	1	44.7	<0.001
B215 Corpus	6	301	1	45.38	<0.001
B221 Antrum	6	302	2	45.79	<0.001
B221 Cardia	6	302	2	44.46	<0.001
B221 Corpus	6	302	2	44.17	<0.001

B253 Antrum	7	301	1	45.02	<0.001
B253 Cardia	7	301	1	45.48	<0.001
B253 Corpus	6	301	1	45.93	<0.001
B256 Antrum	7	303	3	40.17	<0.001
B256 Cardia	7	303	3	42.22	<0.001
B256 Corpus	7	302	2	37.96	<0.070
B266 Antrum	6	300	3	43.12	<0.005
B266 Cardia	6	300	3	43.82	<0.001
B266 Corpus	8	301	1	37.44	<0.250
B268 Antrum	6	302	2	45.79	<0.001
B268 Cardia	7	303	3	41.41	<0.001
B268 Corpus	6	302	2	47.1	<0.001
B284 Antrum	8	304	1	36.85	<0.001
B284 Cardia	9	303	2	34.06	<0.300
B284 Corpus	9	304	1	37.23	<0.001
B292 Antrum	7	302	2	40.85	<0.001
B292 Cardia	8	302	2	40.86	<0.001
B292 Corpus	7	302	2	40.91	<0.001
B294 Antrum	7	299	2	39.04	<0.010
B294 Cardia	6	300	3	43.38	<0.001
B294 Corpus	6	300	3	43.14	<0.001
B295 Antrum	9	302	2	36.76	<0.500
B295 Cardia	7	304	1	37.84	<0.010
B295 Corpus	6	301	1	44.05	<0.001
B300 Antrum	6	300	3	42.89	<0.001
B300 Cardia	7	301	1	37.19	<0.250
B300 Corpus	6	300	3	43.63	<0.001
B301 Antrum	6	302	2	45.47	<0.001
B301 Cardia	6	302	2	43.51	<0.001
B301 Corpus	7	302	2	43.33	<0.001

^a The patient number and their gastric region from which the *H. pylori* population was derived. Populations were collected from the gastric antrum, cardia, and corpus of twelve patients.

^b The total number of *arsS* poly C region fragment lengths generated from any of the nine AFLP-PCR trials and detected via AFLP.

^c The predominant poly C region fragment length determined from fragment length frequency averages calculated with data from all AFLP trials.

^d The predominant *arsS* open reading frame as determined with the extrapolation of poly C region fragment length data. Poly C region fragment lengths that varied by three or six base pairs were summed to determine *arsS* open reading frame frequencies.

^e The predominant *arsS* open reading frame frequency.

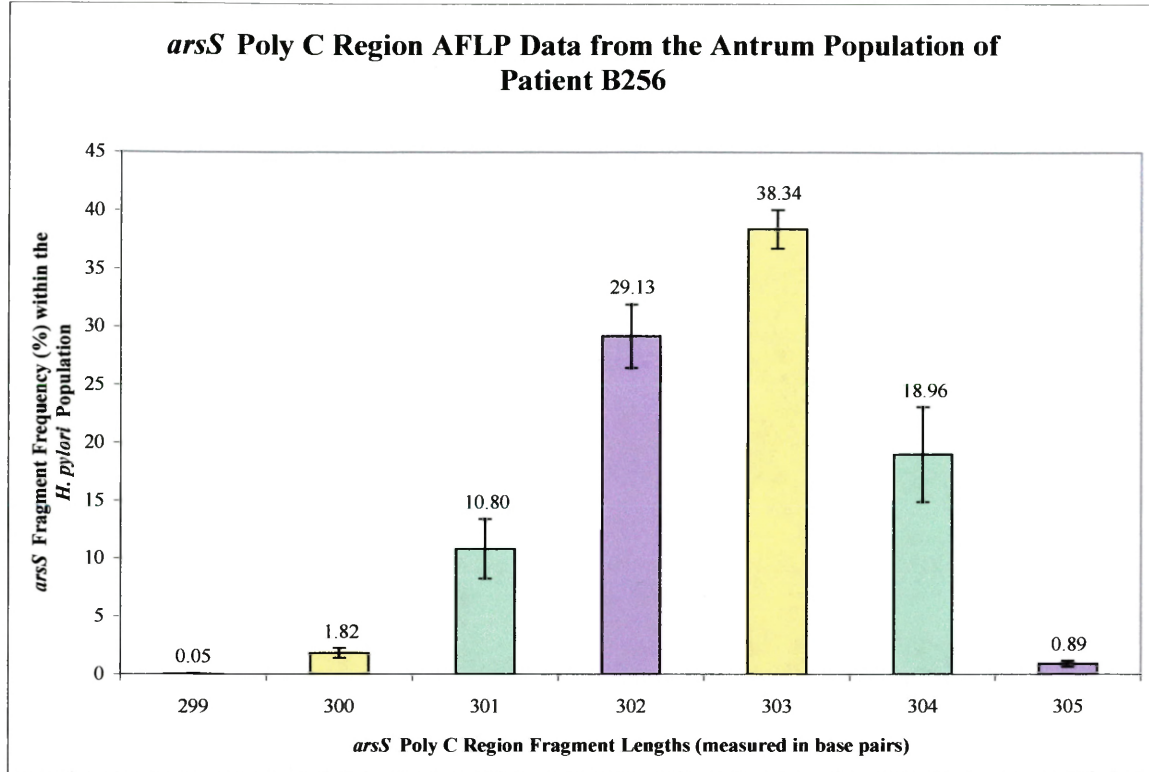
^f The largest p-value of the predominant *arsS* open reading frame when tested against the other two open reading frames. This statistic was determined with a Student's T-test. Blue and purple rows indicate gastric populations from individual patients that encoded different predominant *arsS* open reading frames among each other. This was significant ($p < 0.05$) in the populations of two patients (purple rows). White rows indicate gastric populations from individual patients that encoded the same predominant *arsS* open reading frame among each other.

Variant *arsS* Open Reading Frames are Encoded Within Clinical *Helicobacter pylori* Populations

As we determined that various *arsS* poly C region fragment lengths could be amplified from all *H. pylori* populations, we hypothesized that these fragment lengths could represent different open reading frames of *arsS*. Considering that the genetic code is translated into protein via codons, or three adjacent nucleotides, fragment lengths that varied by three and six base pairs could be considered to be of the same *arsS* open reading frame. On the contrary, fragments that varied in size by one or two and four or five base pairs could be considered to be of different *arsS* open reading frames. Thus, *arsS* open reading frame frequencies were calculated by summing the frequencies of fragment lengths that varied by three and six base pairs. In Figure 3.2, the extrapolation of *arsS* open reading frame frequency from poly C region fragment length frequency is color-coded. Therefore, fragment lengths corresponding to a particular open reading

frame are of the same color. For example, when poly C region AFLP data are converted into *arsS* open reading frame data, frame 3 predominates in the *H. pylori* population from the antrum of patient B256 (Figure 3.2B). However, *arsS* open reading frames 1 and 2 are present, albeit at lower relative frequencies. In fact, similar data analyses from all populations indicated that each potential open reading frame of *arsS* was always present. The graphical representations of *arsS* open reading frame frequencies from all clinical *H. pylori* populations are located in Appendix 2A, AFLP Data from Poly C Regions of *H. pylori* Populations.

A



B

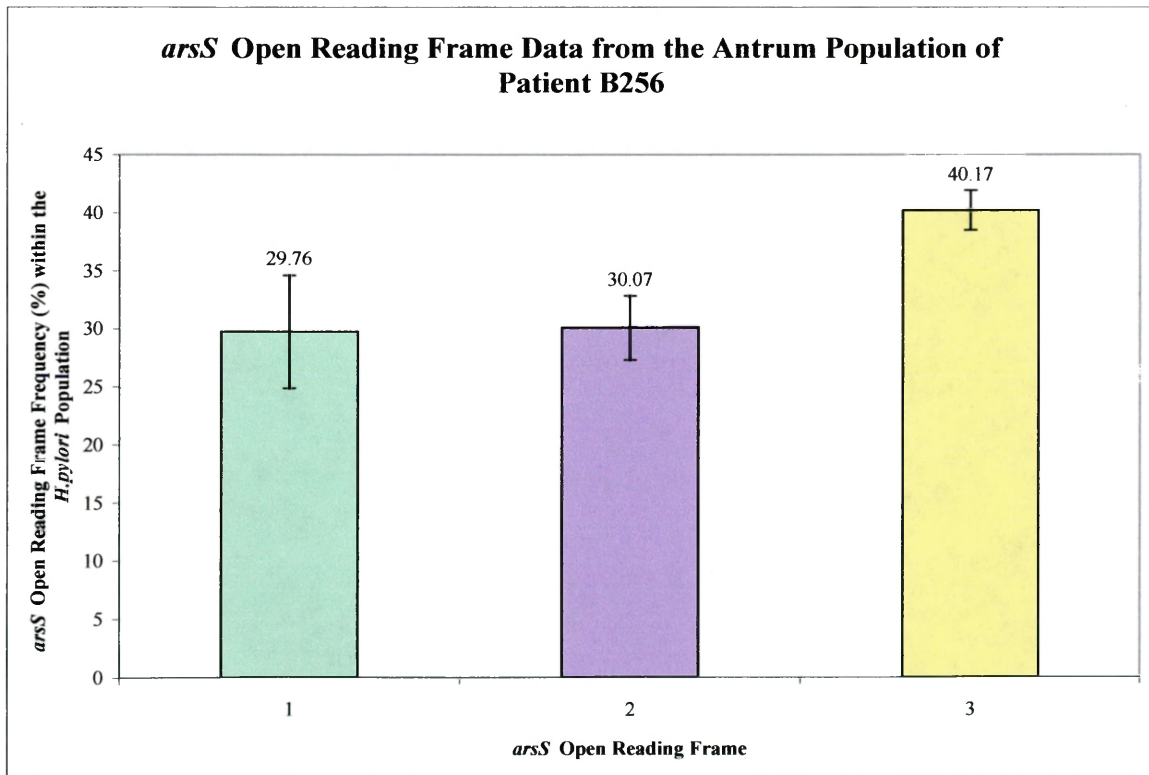


Figure 3.2: Extrapolation of AFLP data into open reading frame data. *arsS* poly C

region fragment length data were converted into *arsS* open reading frame data by summing the frequency of fragment lengths that varied by three or six base pairs. Data representative of fragment frequency (A) were summed according to color for their conversion into data representative of open reading frame frequency (B). Poly C region fragment lengths (A) and *arsS* open reading frames (B) are indicated on the x-axes. The frequency of fragment lengths (A) and open reading frames (B) are located on the y-axes. Error bars indicate their standard deviations.

Predominant *arsS* Open Reading Frames are Encoded Among Clinical *Helicobacter pylori* Populations

As different predominant poly C region fragment lengths were detected among the *H. pylori* populations, variant predominant *arsS* open reading frames were determined as well (Table 3.1). Thus, if the predominant fragment length varied among populations in a particular patient, then the associated open reading frame differed as well. The predominant *arsS* open reading frame remained the same for populations from all gastric regions of five of the twelve patients in this study. These patients are represented by white rows in Table 3.1. However, in seven patients, predominant open reading frames differed among the populations of their gastric regions. These patients are represented by blue or purple rows in Table 3.1. In five of these seven patients, the *arsS* open reading frame variation was not significant ($p < 0.05$) among their populations, which is signified by blue rows in Table 3.1. Conversely, variations in predominant *arsS* open reading frame among the gastric regions of an individual patient were significant ($p < 0.05$) in two cases, which are represented by purple rows in Table 3.1. For instance, the antrum and corpus *H. pylori* populations of patient B268 had more cells that encoded the second open reading frame of *arsS* than did the population from the cardia. On the contrary, the cardia population had more cells that encoded for the third frame of *arsS*. While not always significant ($p < 0.05$), 58.3% of the patients exhibited populations that encoded

different predominant *arsS* open reading frames among their gastric regions. However, populations from 41.7% of the patients significantly ($p < 0.05$) encoded for the same *arsS* open reading frame among their gastric regions.

***arsS* Control Region AFLP-PCR Amplicons Vary in Length Among and Within Clinical *Helicobacter pylori* Populations**

As a control, an *arsS* region of 248 base pairs and upstream of the cytosine tract was amplified via AFLP-PCR for AFLP analysis. *H. pylori* populations from patients B284 and B294 were used for the generation of control region AFLP-PCR. This region was not expected to be polymorphic in nucleotide length based on NCBI sequence data for *H. pylori* strains. However, our data indicate that amplicons of different length could be generated from this region of *arsS* with our current methods. For example, three fragment lengths were detected from the antrum *H. pylori* population of patient B284 (Figure 3.3). Even though multiple fragment lengths were detected, the frequencies of control region fragments were much different than those of the poly C region (Figure 3.1). Shown in Figure 3.3, fragment length of 249 base pairs represents 76.89% of all amplicons generated. The fragment lengths of 248 and 247 base pairs represent 20.74% and 2.37% of all amplicons, respectively (Figure 3.3). Similar bar graphs representing the remaining control region AFLP data from patients B284 and B294 are located in Appendix 2B, AFLP Data from Control Regions of *H. pylori* Populations.

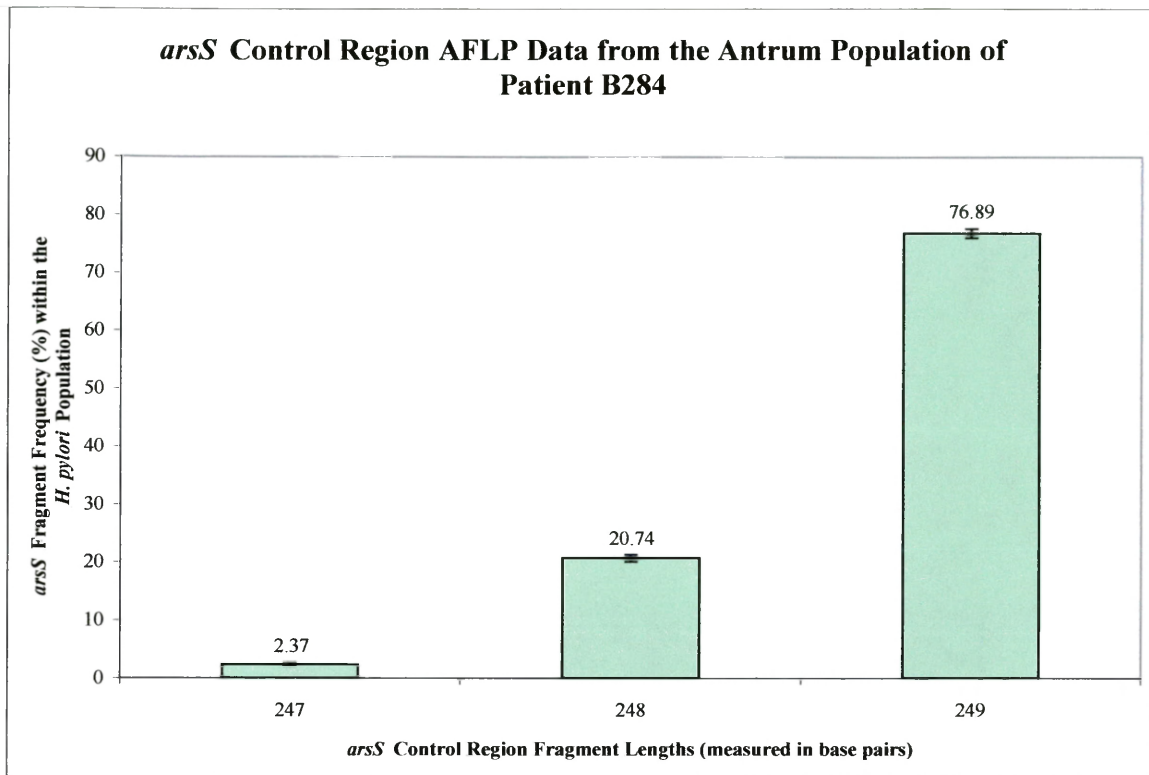


Figure 3.3: *arsS* control region AFLP data from the antrum *H. pylori* population of patient B284. This bar graph was developed with AFLP data generated from the *H. pylori* population derived from the antrum of patient B284. The x-axis represents various *arsS* fragment lengths, measured in base pairs, which were produced via AFLP-PCR. The y-axis represents the relative frequency of each fragment length produced from this particular population. Error bars indicate the standard deviations of the various fragment lengths, which are representative of relative fragment length frequency variations among the nine AFLP trials.

Predominant *arsS* Control Region AFLP-PCR Amplicons Vary in Length Among Clinical *Helicobacter pylori* Populations

The number of control region amplicons detected from any of the investigated populations ranged from three to four (Table 3.2). All gastric populations of patient B284 and the antrum population of patient B294 exhibited three control region amplicons. However, four fragments could be detected with AFLP data from the cardia and corpus populations of patient B294 (Table 3.2). Even though the control region was not suspected of being polymorphic in length, different predominant control region

amplicon lengths were detected among populations from different gastric regions of both patients. A fragment length of 249 base pairs was determined to be predominant in the antrum and corpus populations of patient B284 as well as the antrum population of B294. However, a predominant fragment length of 250 base pairs was detected from the other three populations (Table 3.2)

Table 3.2: Summary of control region AFLP data from *H. pylori* populations

Patient Number and Gastric Region of <i>H. pylori</i> Population^a	Number of <i>arsS</i> Control Region Fragments Detected^b	Predominant <i>arsS</i> Control Region Fragment Length (bp) Detected^c
B284 Antrum	3	249
B284 Cardia	3	250
B284 Corpus	3	249
B294 Antrum	3	249
B294 Cardia	4	250
B294 Corpus	4	250

^a The patient number and their gastric region from which the *H. pylori* population was derived. Populations were collected from the gastric antrum, cardia, and corpus of these patients.

^b The total number of *arsS* control region fragment lengths generated from any of the three AFLP-PCR trials and detected via AFLP.

^c The predominant control region fragment length determined from fragment length frequency averages calculated with data from all AFLP trials.

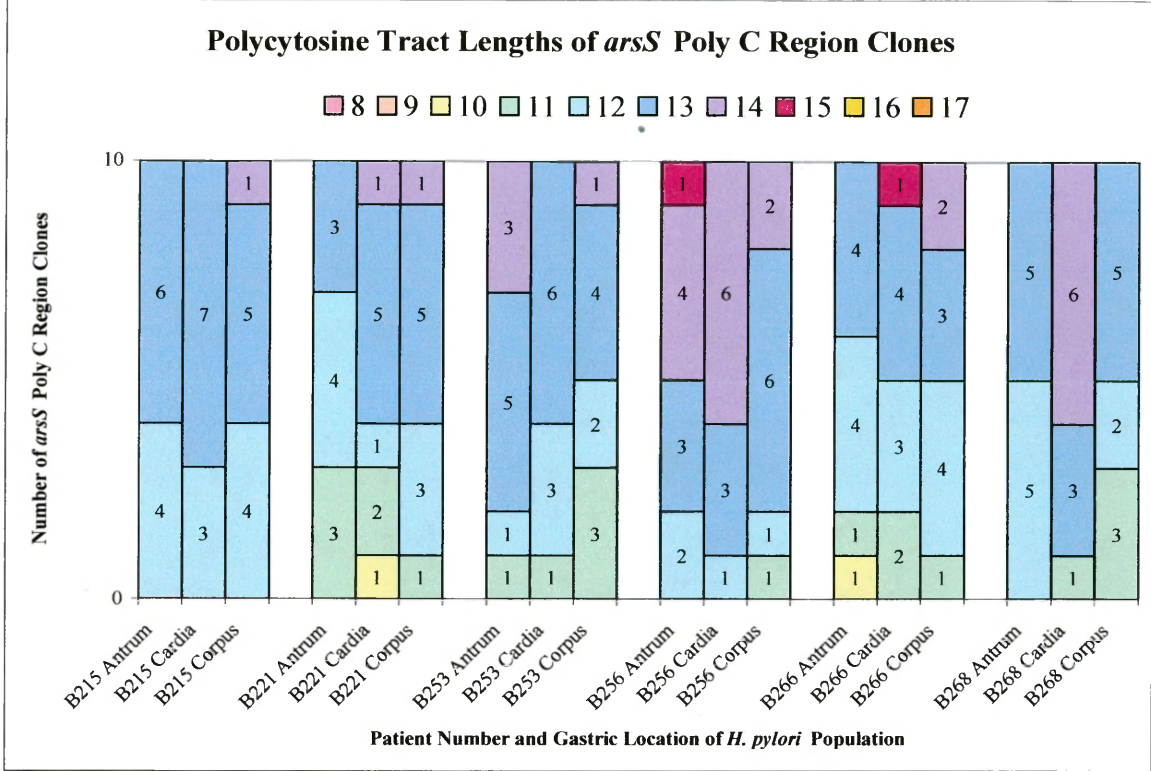
Cloned *arsS* Poly C Region Sequences Encode Variable Polycytosine Tract Lengths Among and Within Clinical *Helicobacter pylori* Populations

Plasmid clones encoding the poly C region were generated to verify that differing AFLP-PCR amplicons were due to polymorphic cytosine tract lengths encoded in the 3' terminus of *arsS*. Ten poly C region plasmid clones using gDNA templates from each population were produced and both strands were sequenced. Poly C region sequences confirmed that cytosine tract lengths of *arsS* varied among and within the clinical *H. pylori* populations (Figure 3.4). Among all populations, the length of the polycytosine

tracts ranged from eight to seventeen consecutive cytosines. Differently colored bars represent the variable tract lengths in Figure 3.4. Within each population, the number of plasmid clones with specific cytosine tract lengths also varied, which are represented by stacked columns (Figure 3.4).

Poly C region sequence data indicate that at least two cytosine tract lengths are present among *H. pylori* cells within each population (Figure 3.4). Furthermore, up to six different polycytosine tract lengths were observed from the cardia *H. pylori* population of patient B284 and the antrum population of patient B295 (Figure 3.4B). Of the thirty-six *H. pylori* populations in this study, twelve (33.3%) had *H. pylori* possessing three different polycytosine tract lengths, eleven (30.6%) had cells with four variant lengths, and seven (19.4%) populations exhibited five different tract lengths (Figure 3.4). We detected only two variable polycytosine tract lengths in just four *H. pylori* populations (11.1%) (Figure 3.4).

A



B

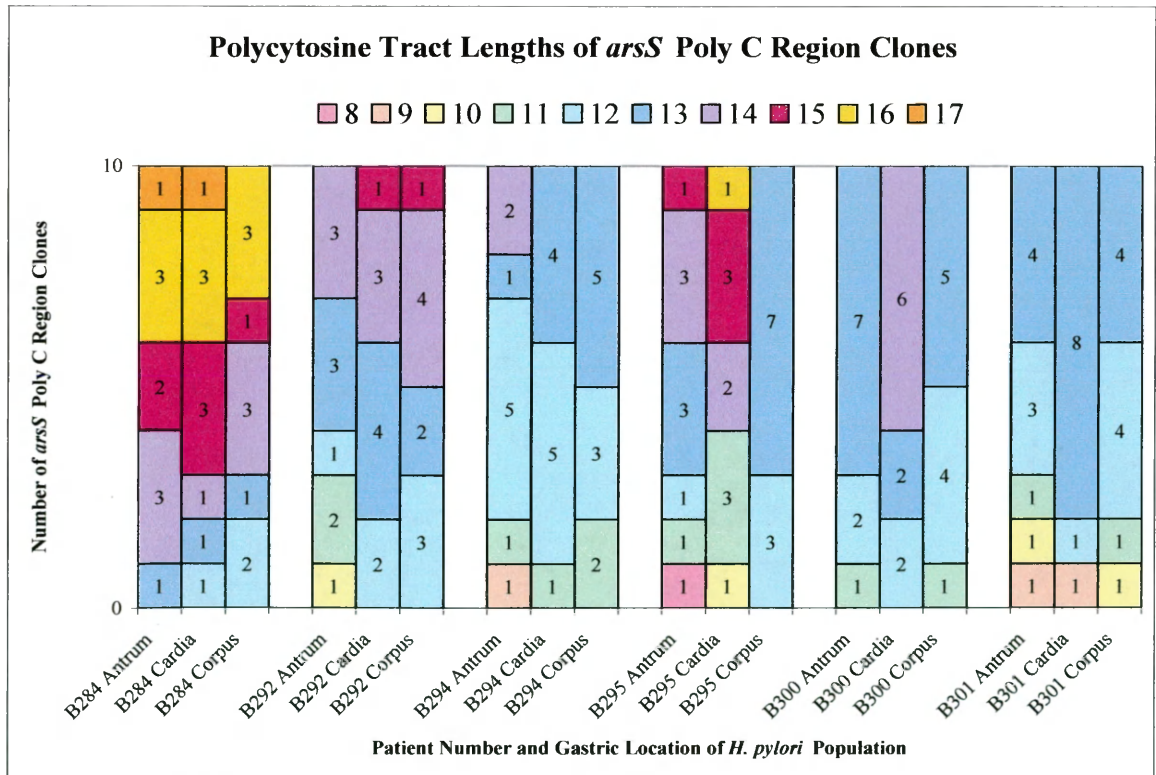


Figure 3.4: Differences in polycytosine tract lengths among all *H. pylori* populations.

This stacked bar graph represents the number of clones (stacked columns) within each *H. pylori* population that exhibited differences in polycytosine tract length (different bar colors). For A and B, the x-axes represent the patient number and gastric regions from which the *H. pylori* populations were derived. The y-axis represents the differing number of clones with variable tract lengths. Among populations, polycytosine tracts were observed to deviate in length by eight to seventeen consecutive cytosines. Within each population, sequence data indicate that at least two and as many as six distinct polycytosine tract lengths could be detected. (A) Polycytosine tract lengths of *arsS* poly C region clones generated from the gastric populations of patients B215, B221, B253, B256, B266, and B268. (B) Polycytosine tract lengths of *arsS* poly C region clones generated from the gastric populations of patients B284, B292, B294, B295, B300, and B301.

Cloned *arsS* Poly C Region Sequences Reveal Predominant Polycytosine Tract Lengths Among Clinical *Helicobacter pylori* Populations

Poly C region clones varied in their polycytosine tract lengths among *H. pylori* populations as well as their predominant tract lengths. Examination of the data presented in Figure 3.4 shows that cytosine tract lengths of twelve, thirteen, and fourteen are clearly the most frequently represented among the populations. Of 360 poly C region clones, a cytosine tract length of thirteen was identified in 144, or 40% of all plasmid clones. Furthermore, a cytosine tract length of twelve was recognized in 89 cloned *arsS* poly C region sequences, or 24.7% of all clones. Sequence data also indicated that 57 clones had cytosine tract lengths of fourteen, which represents 15.8% of all clones. Thus, 80.5% of all cloned sequences possessed one of these three predominant polycytosine tract lengths, whereas all other tract lengths contributed to 20.4% of all clones.

Cloned *arsS* Poly C and Control Region Sequences Vary in Length Among Clinical *Helicobacter pylori* Populations Due to Deletions in Primer Binding Regions

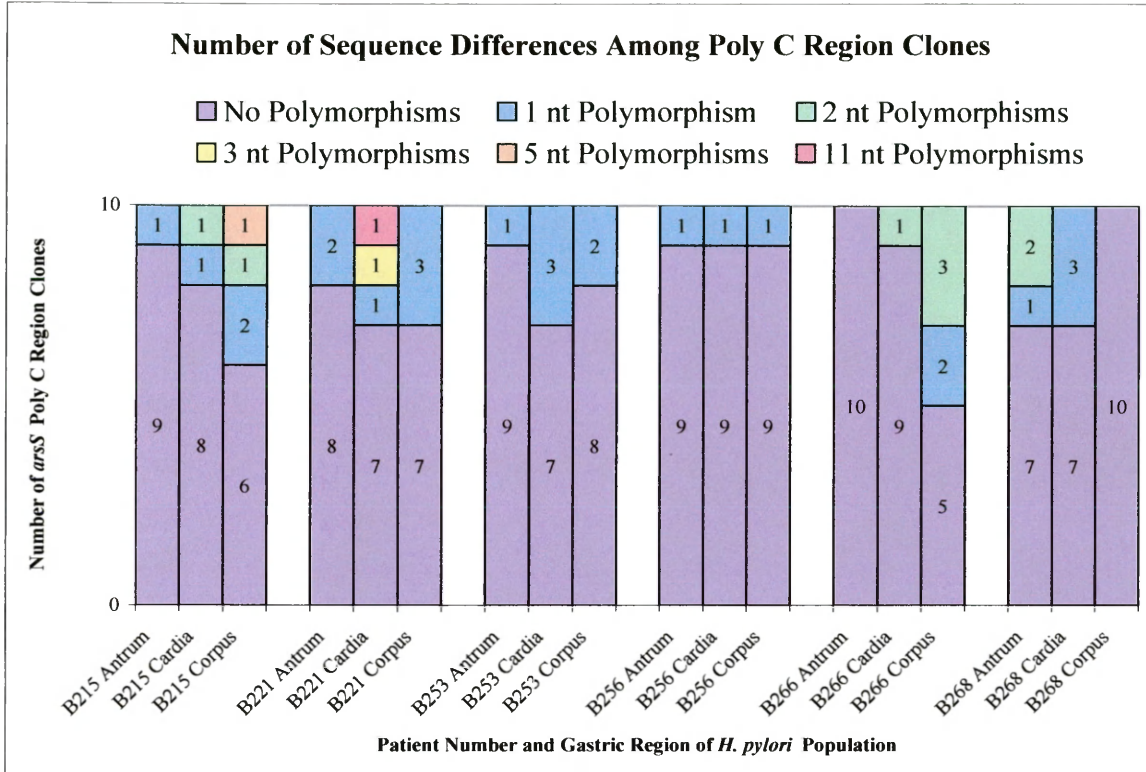
Sequencing data indicated that single nucleotide deletions could be detected in the primer binding regions of the amplicons utilized to generate clones encoding the poly C

and control regions of *arsS*. Though rare, these deletions were present in five (1.4%) of the 360 cloned poly C region sequences and in two (3.3%) of the 60 cloned control region sequences. These deletions likely affected AFLP sizing data, but were not considered evidence for divergent strains in sequencing analyses. Instead, these differences were speculated to be due to primer synthesis errors.

Cloned *arsS* Poly C Region Sequences Indicate Different *Helicobacter pylori* Strains Within Populations

When ClustalW alignments were performed to compare cloned poly C region sequences within single populations, our data indicate that different strains of *H. pylori* were present. Given that one nucleotide polymorphisms, disregarding variant polycytosine tract lengths, were our metric for strain determination; nucleotide sequences that were identical within a gastric population were considered to be of the same strain and sequences that encoded nucleotide substitutions were considered to be of distinct strains. In fact, we frequently observed clones that encoded no sequence polymorphisms, as at least five to as many as ten clones were homologous within individual populations. We also clearly observed cloned sequences that differed in nucleotide sequence by one base pair to as many as eleven base pairs. Thus, we found that twenty-nine (80.6%) of the thirty-six populations contained multiple strains of *H. pylori*. These are represented by stacked columns of various colors (Figure 3.5). On the contrary, sequence data indicate that the remaining seven (19.4%) *H. pylori* populations consisted of only one strain, which are indicated by columns of the same color in Figure 3.5.

A



B

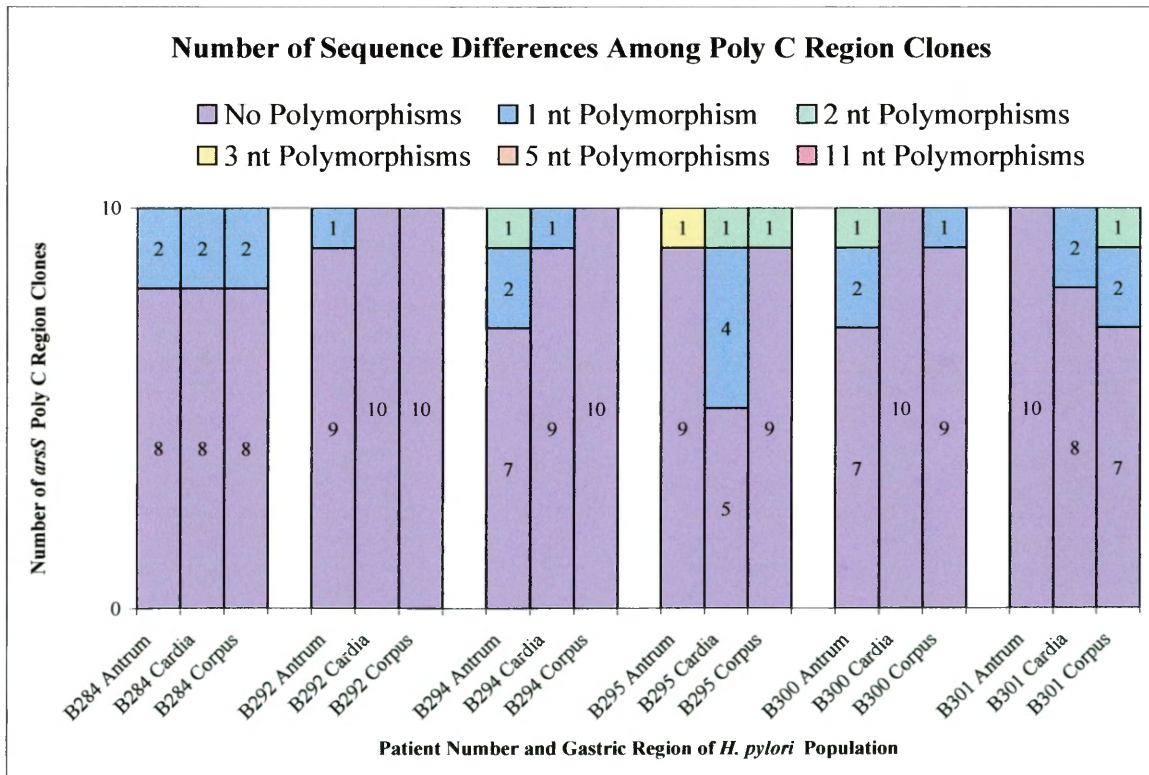


Figure 3.5: Number of sequence polymorphisms outside of the polycytosine tract

among poly C region plasmid clones from *H. pylori* populations. This stacked bar graph represents the number of clones (stacked columns) within each *H. pylori* population that exhibited differences in nucleotide sequence. Here, differently colored bars indicate the number of sequence polymorphisms relative to the sequence observed most often. For A and B, the x-axes represent the patient number and gastric regions from which the *H. pylori* populations were derived. The y-axes represent the differing number of clones with variable nucleotide polymorphisms. Within each population, data indicate that there are poly C region sequences that exist with and without polymorphisms when disregarding the polycytosine tract length. For example, five clones generated from the corpus population of patient B266 were identical. In comparison, two of the other five clones had a single nucleotide polymorphism and the remaining three cloned sequences possessed two polymorphisms. (A) Number of sequence differences among *arsS* poly C region clones generated from the gastric populations of patients B215, B221, B253, B256, B266, and B268. (B) Number of sequence differences among *arsS* poly C region clones generated from the gastric populations of patients B284, B292, B294, B295, B300, and B301.

Cloned *arsS* Poly C Region Sequences Indicate Predominant *Helicobacter pylori*

Strains *Within* Populations

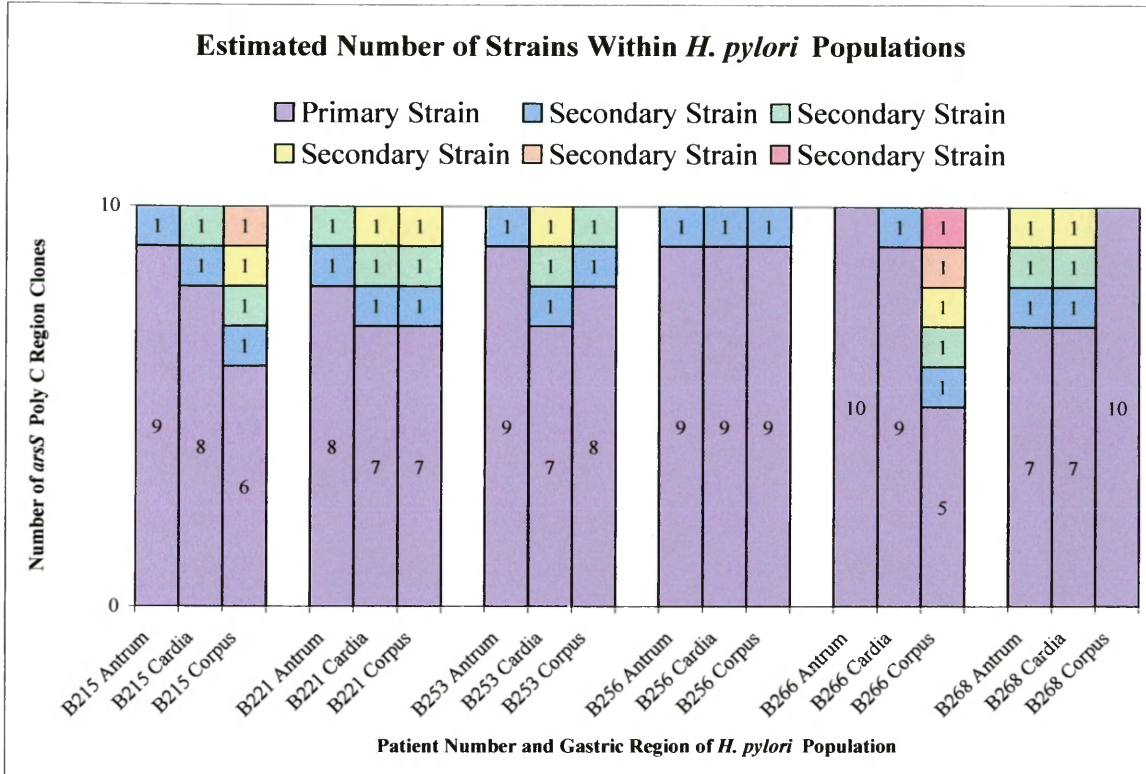
Among poly C regions clones generated from the same *H. pylori* population, data clearly indicate that there are a large number of cloned *arsS* poly C region sequences that do not possess polymorphisms when disregarding polycytosine tract lengths (Figure 3.5). Thus, we interpret this vast quantity of identical clones as evidence that indicates each clinical population appears to have a predominant strain, even when multiple strains are present.

Less obvious in Figure 3.5 are the estimated number of strains within each population. This is due to the observation that none of the deviant poly C region sequences matched one another within a single population. For instance, the corpus population of patient B266 is estimated to have six different strains of *H. pylori*. This number of strains was contrived by considering the five clones without polymorphisms as one strain and the five clones with polymorphisms as five separate strains (Figure 3.5).

An alternative representation of the total number of strains detected within each *H. pylori* population is shown in, Figure 3.6.

Figure 3.6 indicates that one primary strain and as many as six secondary strains could be detected among clones generated from the same *H. pylori* populations. In seven of the populations, the *H. pylori* infection appears to be monoclonal, or consisting of a single strain. However, populations with a multiple number of strains were detected more frequently. This sort of strain variation was observed in twenty-nine of the thirty-six populations in the current study (Figure 3.6). Interestingly, none of the patients appear to have a monoclonal infection in all three of their gastric regions sampled. For example, patient B292 appears to have two monoclonal populations located in their gastric cardia and corpus, but possesses a polyclonal population in the gastric antrum. Furthermore, seven patients lack monoclonal populations altogether, whereas the five other patients have at least one monoclonal population (Figure 3.6). Whether the infection was monoclonal or polyclonal, predominant strains were still obviously apparent within all clinical populations.

A



B

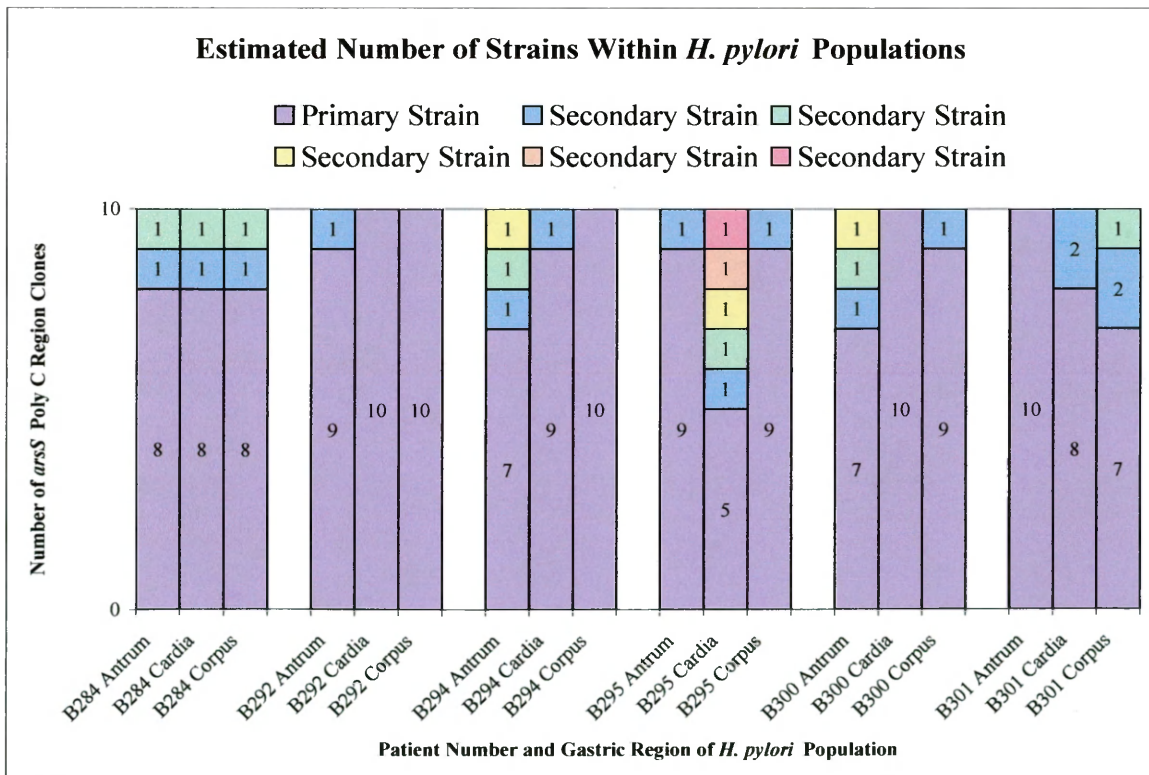


Figure 3.6: Estimated number of strains within *H. pylori* populations via poly C region

clones. This stacked bar graph represents the number of clones (stacked columns) within each *H. pylori* population that were considered to be of different strains. Various colored bars indicate the primary strain (purple) and secondary strains (assorted colors). Interestingly, none of the secondary strains were detected twice within or among *H. pylori* populations. For A and B, the x-axes represent the patient number and gastric regions from which the *H. pylori* populations were derived. The y-axes represent the differing number of clones representing different strains. Within each population, sequence data indicate that there are predominant strains (primary strain), which are represented more frequently than the others (secondary strains). Thus, data suggest that seven of the populations consist of only the primary strain. However, data also suggest that secondary strains are observed in twenty-nine populations. (A) Estimated number of strains contrived with data from *arsS* poly C region clones generated from the gastric populations of patients B215, B221, B253, B256, B266, and B268. (B) Estimated number of strains contrived with data from *arsS* poly C region clones generated from the gastric populations of patients B284, B292, B294, B295, B300, and B301.

Cloned *arsS* Poly C Region Sequences Indicate Predominant *Helicobacter pylori*

Strains Among Populations of an Individual Patient

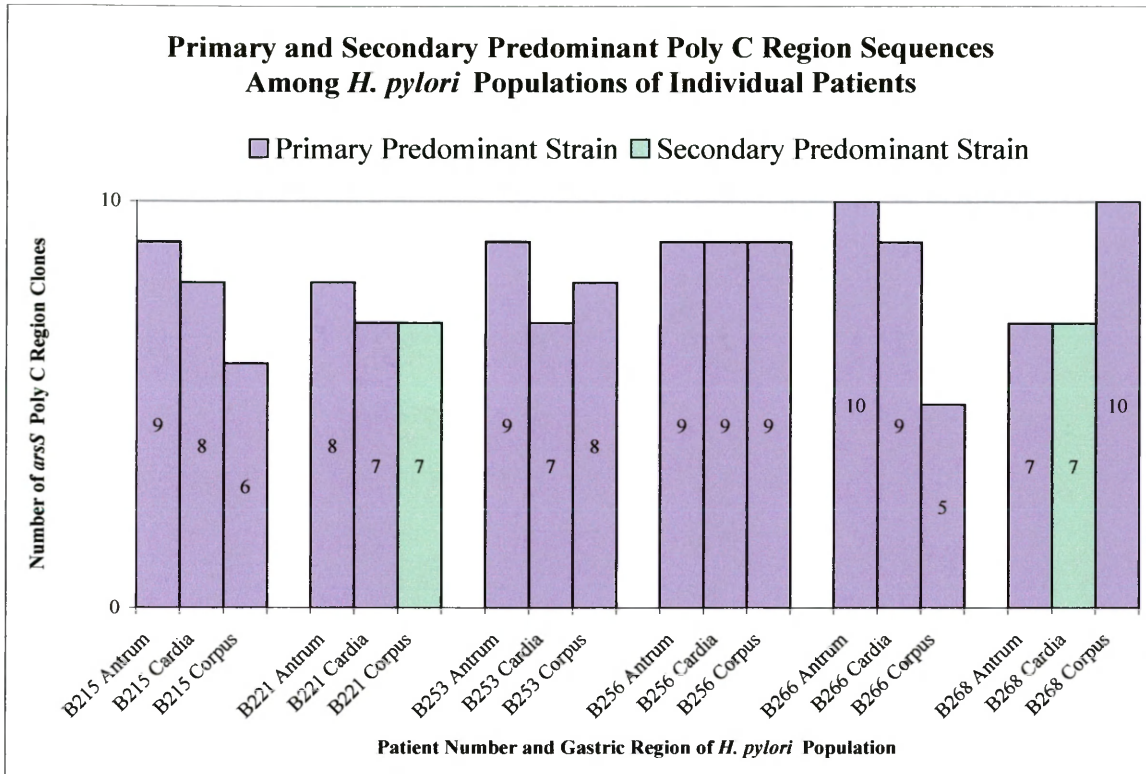
Now that we had established that predominant strains were evident within each of the clinical populations, we set out to determine if there were similar predominant strains among the gastric regions of all patients. Thus, we performed ClustalW alignments with the cloned poly C region sequences that were previously shown to be predominant within single populations (Figure 3.6). Interestingly, we found that if a predominant *H. pylori* strain was detected in one patient that strain would not be found in another patient, no matter the gastric location of the strain. For instance, if an *H. pylori* strain was detected in a population from patient B301, the same strain could not be detected in any of the other eleven patients. However, we did find that the same predominant strains could be detected among the gastric regions of an individual patient.

Each patient was infected with *H. pylori* populations that had an identical predominant strain observed in at least two of their gastric regions. As an example, data from populations of patient B253 indicate that the predominant strain of the corpus was

also the same predominant strain of the antrum and cardia. Keeping in mind that only strains within a patient were identical, predominance of the same strain among all gastric regions was observed in nine (75%) of the twelve patients in this study.

In each of the three remaining patients, we found that two of their gastric regions contained the same predominant strain, but their third gastric region had a dissimilar predominant strain. For example, patient B221 had the same predominant *H. pylori* strain in their antrum and cardia populations, but had a dissimilar predominant strain in their corpus population. The correlations concerning predominant strains and dissimilar predominant strains are shown in Figure 3.7. This stacked bar graph is only representative of clones that were previously determined to be from predominant strains; secondary strains have been removed for easier interpretation. Here, purple bars represent the number of cloned sequences that were of the same predominant strain from at least two gastric regions within an individual patient, which are denoted as the primary predominant strain. Furthermore, green bars indicate the number of clones that were of a dissimilar predominant strain in a particular gastric region of an individual patient, which we have indicated as the secondary predominant strain (Figure 3.7).

A



B

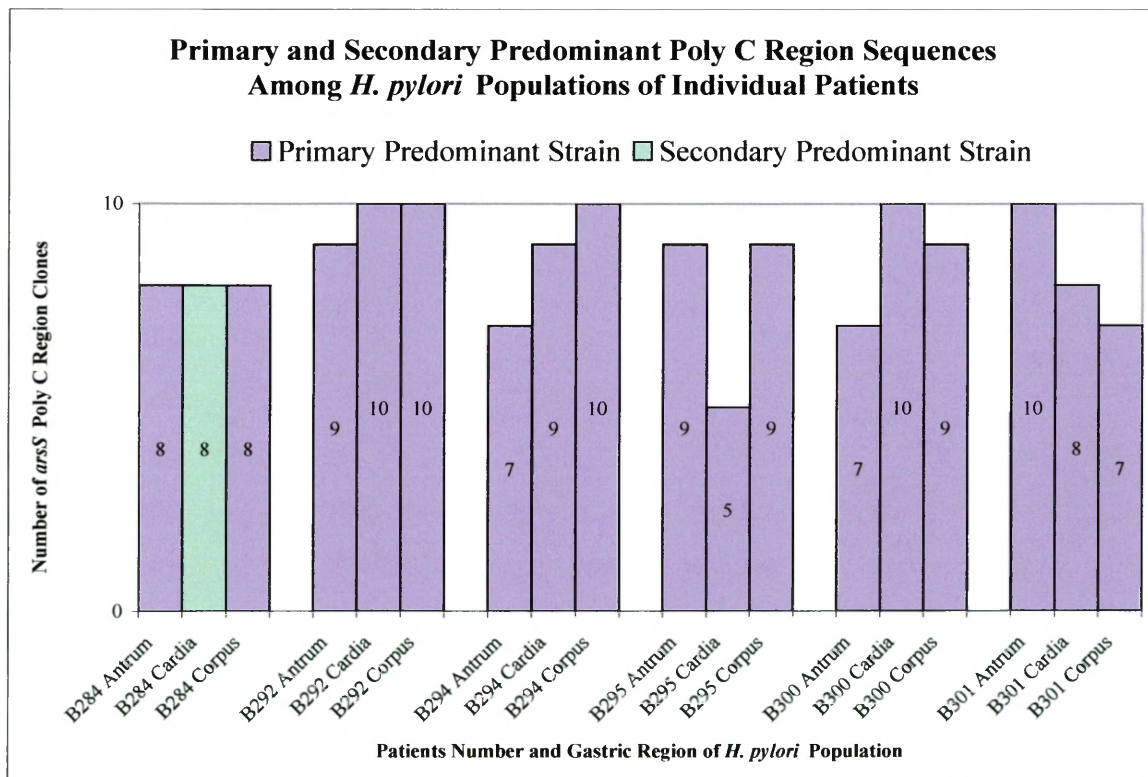


Figure 3.7: Primary and secondary predominant strains of *H. pylori* populations

determined with poly C region clones. This bar graph represents the number of clones within each *H. pylori* population that originated from predominant strains. Here, differently colored bars indicate whether the predominance was primary or secondary, in respect to populations from particular patients. For A and B, the x-axes represent the patient number and gastric regions from which the *H. pylori* populations were derived. The y-axes represent the differing number of clones representing predominant strains. Among the populations of an individual patient, sequence data indicate that there is a primary predominant strain (purple), which is represented in at least two of their gastric locations. Furthermore, secondary predominant strains (green) arise in one gastric region population of particular patients. Thus, data suggest that nine of the patients have only primary predominant strains among their gastric populations. However, data also suggest that secondary predominant strains are present in one gastric population of three patients. (A) Primary and secondary predominant strains determined with data from *arsS* poly C region clones generated from the gastric populations of patients B215, B221, B253, B256, B266, and B268. (B) Primary and secondary predominant strains determined with data from *arsS* poly C region clones generated from the gastric populations of patients B284, B292, B294, B295, B300, and B301.

Cloned *arsS* Control Region Sequences Indicate Different *Helicobacter pylori* Strains Within Populations

Similar to data from poly C region clones, we observed the existence of different strains in our *arsS* control region cloned sequences as well. Again, polymorphisms in the cloned control region sequences allowed us to detect different strains within specific populations. Figure 3.8 shows the number of sequence differences in control region clones with respect to the cloned sequence observed most often. More frequently, we noticed that sequences from particular populations were without polymorphisms, which were indicative of one strain. On the contrary, one to as many as twenty-one nucleotide polymorphisms could be detected from the clones of the six populations. Thus, data suggest that four of these populations consist of multiple *H. pylori* strains, which are indicated by stacked columns of various colors (Figure 3.8). Furthermore, data suggest that only one strain could be detected in two of the populations, which are represented by columns of one color (Figure 3.8).

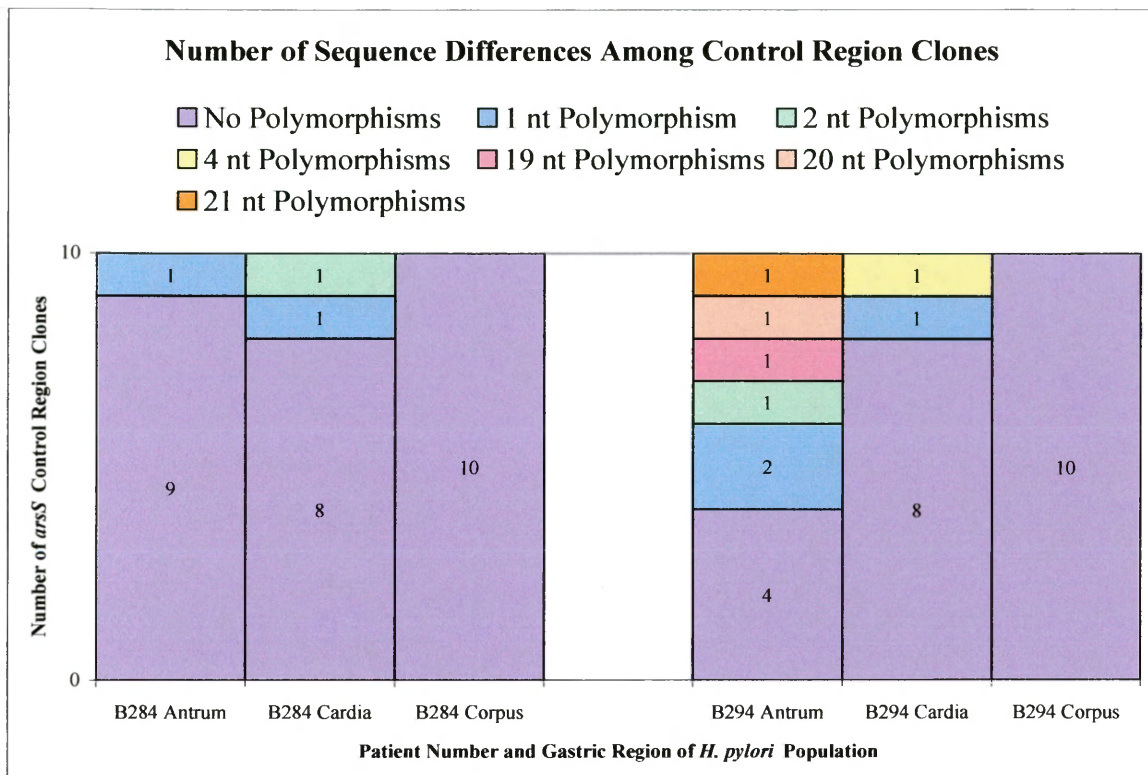


Figure 3.8: Number of sequence differences among control region clones within *H. pylori* populations. This stacked bar graph represents the number of clones (stacked columns) within each *H. pylori* population that exhibited differences in nucleotide sequence. Here, differently colored bars indicate the number of sequence polymorphisms relative to the sequence observed most often. The x-axis represents the patient number and gastric regions from which the *H. pylori* populations were derived. The y-axis represents the differing number of clones with variable nucleotide polymorphisms. Within each population, data indicate that there are sequences that exist with and without polymorphisms. Specifically, this graph shows the number of sequence differences among *arsS* control region clones generated from the gastric populations of patients B284 and B294.

Cloned *arsS* Control Region Sequences Indicate Predominant *Helicobacter pylori* Strains Within Populations

Since cloned control region sequences suggested that four populations consisted of multiple strains, we set out to estimate how many strains were present within each population as we did for cloned poly C region sequencing data. ClustalW alignments suggested that one primary strain to as many as six secondary strains could be detected

from control region clones (Figure 3.9). Monoclonal populations were observed in the corpus populations from patient B284 and B294 (Figure 3.9). Interestingly, a monoclonal corpus population of patient B294 was also suggested by cloned poly C region sequence data (Figure 3.6). However, cloned poly C region sequence data suggested that the corpus B284 population was not monoclonal (Figure 3.6).

One primary strain and five secondary strains were detected in the antrum population of patient B294 (Figure 3.9). Interestingly, two clones from the antrum population of patient B294 were representative of the same secondary strain. This was a characteristic that had not been previously observed in cloned poly C region data (Figure 3.7).

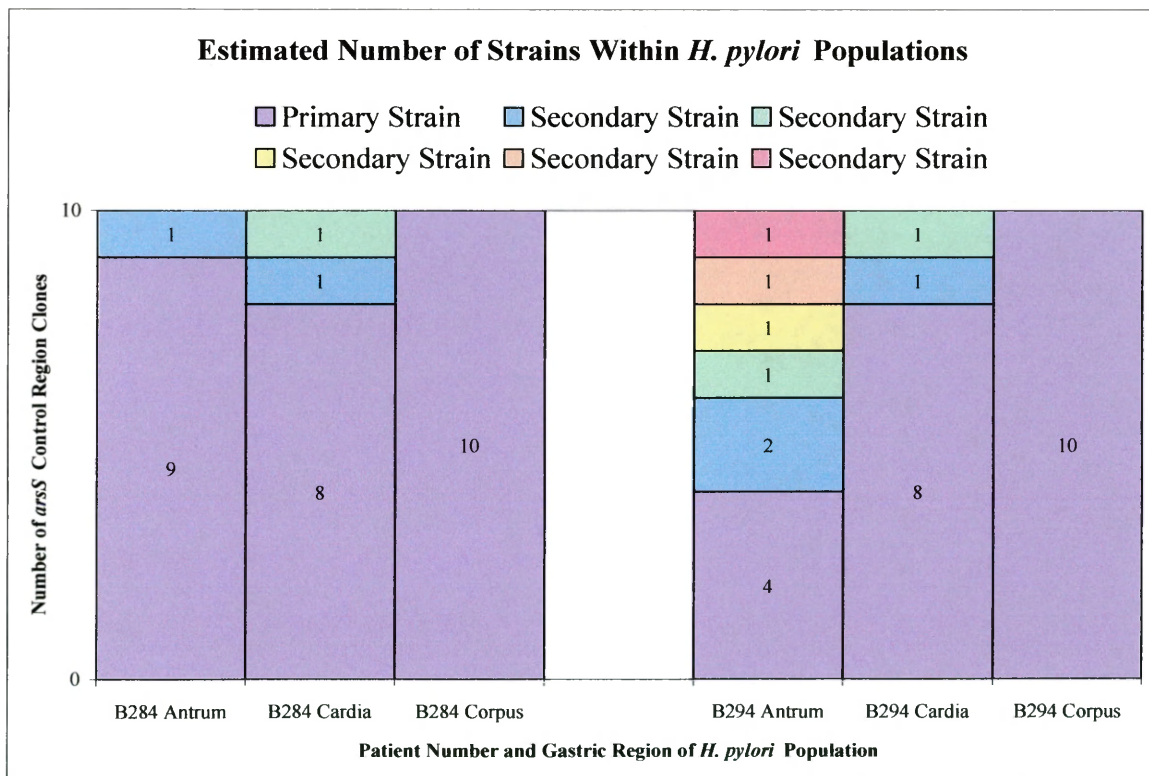


Figure 3.9: Estimated number of strains within *H. pylori* populations via control region clones. This stacked bar graph represents the number of clones (stacked columns) within each *H. pylori* population that were considered to be of different strains. Here, differently colored bars indicate the primary strain (purple) and secondary strains

(various colors). The x-axis represents the patient number and gastric regions from which the *H. pylori* populations were derived. The y-axis represents the differing number of clones representing different strains. Within each population, sequence data indicate that there are predominant strains (primary strain), which are represented more frequently than the others (secondary strains). Thus, data suggest that two of the populations consist of only the primary strain. However, data also suggest that secondary strains are observed in four populations. Specifically, this graph shows the estimated number of strains contrived with data from *arsS* control region clones generated from the gastric populations of patients B292 and B294.

Cloned *arsS* Control Region Sequences Indicate Predominant *Helicobacter pylori* Strains Among Populations of an Individual Patient

Predominant strains were apparent from the graph representing the estimated number of strains within each control region population (Figure 3.9). Thus, we set out to determine the relatedness of control region clones amongst populations as we did for poly C region clones. Again we found that primary predominant strains existed in at least two gastric regions of the patients. However, one gastric region of both patients was colonized with secondary predominant strains (Figure 3.9). For instance, the cardia population of patient B284 consisted of a secondary predominant strain that was different than the primary predominant strain of the antrum and corpus (Figure 3.10). Cloned control region sequence data suggesting a secondary predominant strain in the cardia of patient B284 was congruent with cloned poly C region sequence data (Figure 3.7)

A primary predominant strain was also detected in the cardia and corpus populations of patient B294. Interestingly, this primary predominant strain and a secondary predominant strain were found in the antrum population of patient B294. Here, clones generated from the secondary predominant strain actually outnumbered the clones produced from the primary strain (Figure 3.10). Poly C region data did not previously indicate a secondary predominant strain in any of the populations of patient

B294 (Figure 3.7). Thus, cloned poly C region sequences for predominant strains of the antrum of patient B294 population were always identical, yet cloned control region primary and secondary predominant sequences were not (Figure 3.10).

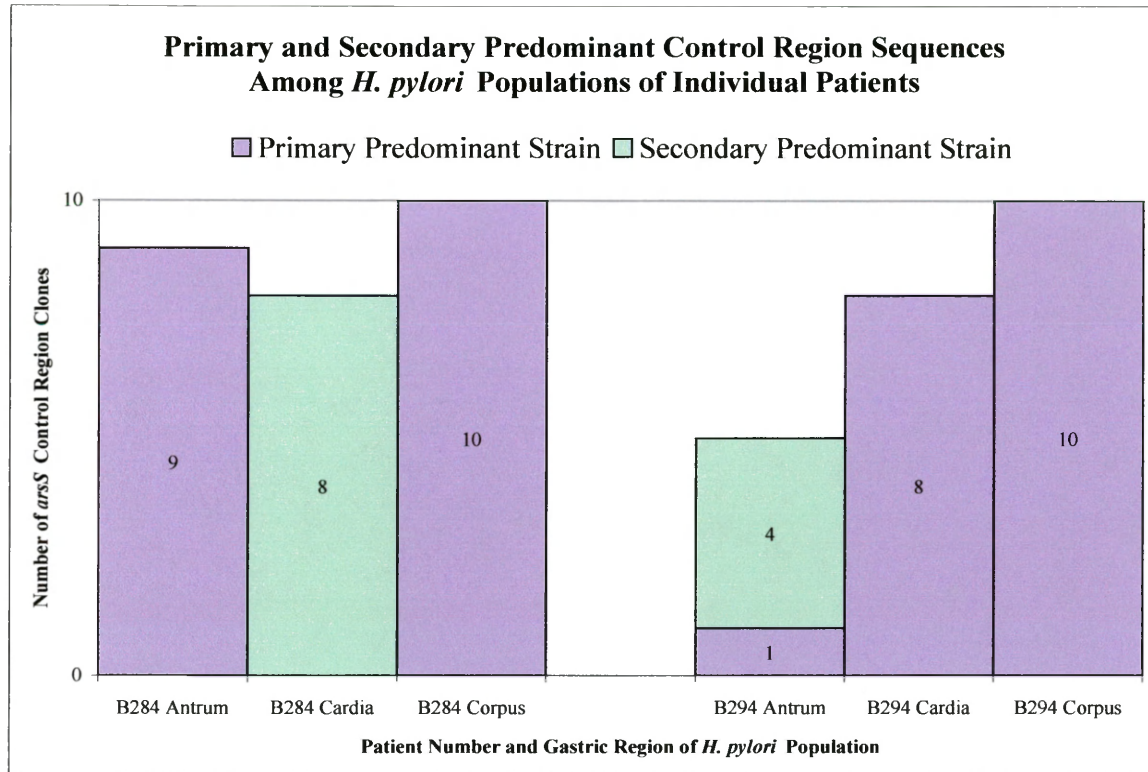


Figure 3.10: Primary and secondary predominant strains of *H. pylori* populations determined with control region clones. This bar graph represents the number of clones within each *H. pylori* population that originated from predominant strains. Here, differently colored bars indicate whether the predominance was primary or secondary, in respect to populations from the particular patients. The x-axis represents the patient number and gastric regions from which the *H. pylori* populations were derived. The y-axis represents the differing number of clones representing predominant strains. Among the populations of an individual patient, sequence data indicate that there is a primary predominant strain (purple), which is represented in at least two of their gastric locations. Furthermore, secondary predominant strains (green) arise in one gastric region population in both of these patients. Thus, data suggest that each of these patients has a primary and a secondary predominant strain among their gastric regions. Furthermore, a primary and a secondary predominant strain were detected in the antrum of patient B294. This specific graph represents primary and secondary predominant strains determined with data from *arsS* control region clones generated from the gastric populations of patients B284 and B294.

The *arsS* Poly C Region is Variable Outside of the Polycytosine Tract Among *Helicobacter pylori* Populations

When ClustalW alignments were executed with each of the 360 cloned poly C region sequences, another polymorphism in this region was detected. This difference in sequence was due to a thymine deletion downstream of the polycytosine tract in 241 (66.9%) of the *arsS* poly C region cloned sequences from all *H. pylori* populations. This polymorphism tended to be specific to clones generated from populations of particular patients. In fact, each of the thirty clones generated from eight of the patients possessed this single nucleotide deletion (Figure 3.11). Virtually none of the cloned sequences from any of the other four patients possessed the thymine deletion (Figure 3.11). However, one (0.28%) clone from the cardia population of patient B221 did possess this deletion, whereas the remaining twenty-nine clones of B221 did not. Thus, *arsS* was determined to possess a length polymorphism outside of the polycytosine tract. We hypothesize that this deletion adds to the *arsS* variation in *H. pylori* strains and may add to the functional complexity of the encoded sensory histidine kinase ArsS.

Presence of a Deletion in *arsS* among Cloned Poly C Region Sequences from Gastric Populations of 12 Patients

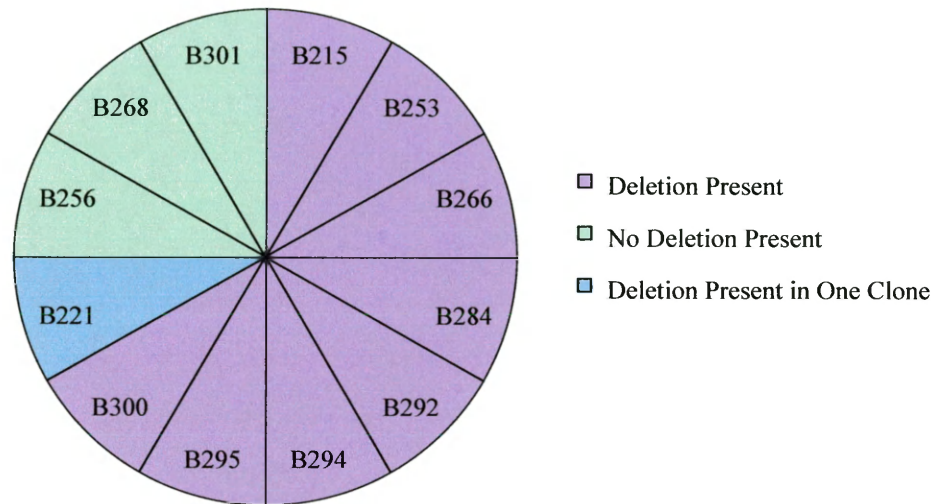


Figure 3.11: Presence of a deletion among *arsS* poly C region clones from each gastric population of twelve patients. This pie chart is to reference the patients from which clones were generated that had a polymorphic deletion (purple) or did not (green). A clone that was generated from patient B221 (blue) had a deletion when the remaining 29 clones from this patient did not.

Different *arsS* Open Reading Frames Arise Due to Changes in Polycytosine Tract

Length and the Polymorphic Deletion

With *H. pylori arsS* polycytosine tract length polymorphisms now well documented among and within *H. pylori* populations, *in silico* translation of the cloned poly C regions demonstrated that three different open reading frames arise. Thus, differences in polycytosine tract lengths allow for the generation of three ArsS histidine kinase isoforms. For simplicity, these isoforms are denoted by the last three amino acids of their peptide sequence prior to their respective stop codons (*). For example, the 0 open reading frame of *arsS* gave rise to the PKI* ArsS isoform while the -1 open reading frame of *arsS* allowed for the generation of the LWG* isoform (Figure 3.12). However, alternative isoforms were translated from the +1 frame of *arsS* among the clinical populations. These variations in +1 *arsS* open reading frame were due to the previously

discussed polymorphic thymine deletion downstream of the polycytosine tract, which specifically occurs in one of the *ArsS* stop codons (Figure 3.12). Thus, if *arsS* possesses this functional stop codon and was transcribed in the +1 open reading frame, then the EKQ* *ArsS* isoform could be expressed (Figure 3.12). However, if the thymine deletion in that stop codon was present and *arsS* was transcribed in the +1 open reading frame, then the SND* isoform could be achieved (Figure 3.12). Interestingly, the stop codon for the SND* isoform is past the start codon of *hemB* (Figure 3.12).

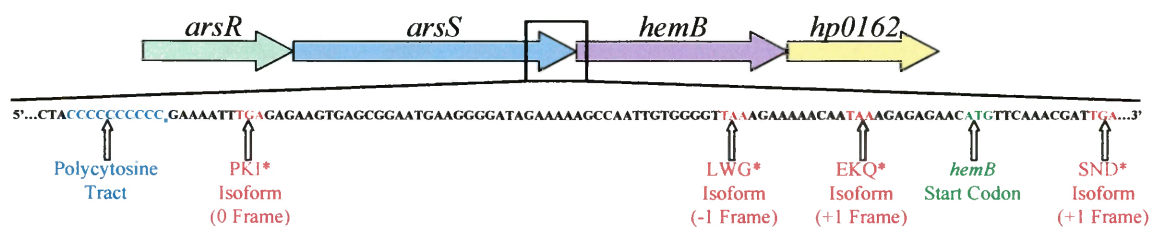


Figure 3.12: Schematic of the *arsRS* operon. The *arsRS* operon consists of *arsR*, *arsS*, *hemB*, and a gene encoding for a hypothetical protein. Specifically, this diagram focuses on the 3' terminus of *arsS*. Here, the *arsS* homopolymeric cytosine tract (blue) and associated stop codons of the different *arsS* open reading frames (red) are shown. As the poly C tract (blue) varies in length (C_n), different isoforms of *ArsS* arise because of changes in the open reading frame. Interestingly, most of the clinical *H. pylori* populations were found to have a deletion of the thymine located in the third stop codon, which allows for an alternate +1 open reading frame and isoform.

Different *arsS* Open Reading Frames are Present Within and Among *Helicobacter pylori* Populations Allowing for Different *ArsS* Isoforms

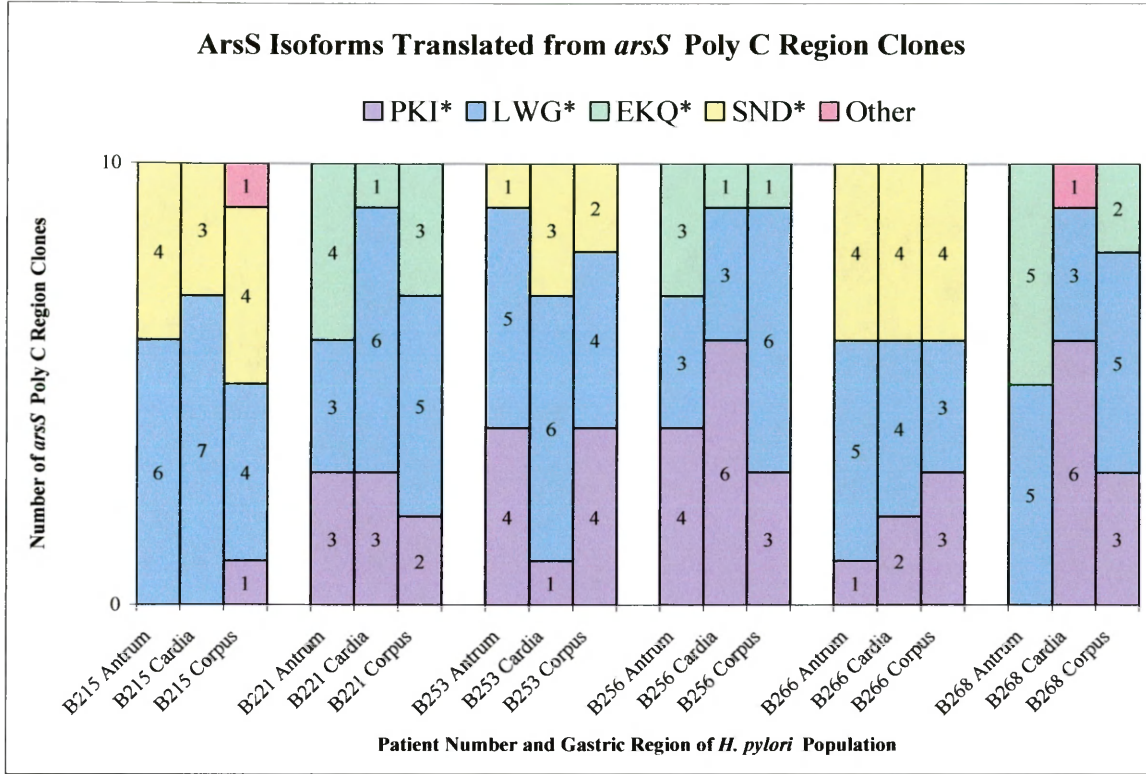
Cloned *arsS* poly C region sequences indicate different *arsS* reading frames and therefore different *ArsS* isoforms can be predicted within and among the clinical *H. pylori* populations. The 0 frame of *arsS*, encoding for the PKI* isoform (Figure 3.12), was observed in thirty-one (86.1%) of the thirty-six *H. pylori* populations when *arsS* poly C region clones were translated *in silico* (Figure 3.13). Thus, a poly C region sequence

capable of translating the PKI* isoform was not observed in five (13.9%) of the clinical populations (Figure 3.13). Conversely, the -1 frame of *arsS*, encoding the LWG* isoform (Figure 3.12), was detected in all thirty-eight of the clinical populations of *H. pylori* (Figure 3.13). The alternative +1 frames of *arsS*, encoding for the EKQ* and the SND* isoforms (Figure 3.12), were present in eleven (30.6%) and twenty-four (66.7%) populations, respectively (Figure 3.13). These low percentage values can be attributed to the observation that EKQ* or SND* isoforms were typically specific to populations infecting particular patients. Thus, the EKQ* isoform was present in eleven (91.6%) of the twelve populations that appeared to consist of strains that could encode the *arsS* alleles to produce this isoform. Furthermore, *arsS* alleles predicted to express the SND* isoform were detected in all of the *H. pylori* populations that had the thymine deletion in the third stop codon of *arsS* and thus had the ability to generate this isoform.

Interestingly, the ArsS LWG* isoform was predicted to be expressed by 158 (43.9%) of the total 360 poly C region cloned sequences (Figure 3.13). The PKI* isoform was predicted to be expressed by 93 (25.8%) of the clones (Figure 3.13). Collectively, the alternative +1 frame isoforms represented 106 (29.4%) of the cloned *arsS* poly C region sequences. Separately, the EKQ* and SND* isoforms comprised 30 (8.3%) and 76 (21.1%) of all clones. Three (0.8%) poly C region cloned sequences possessed deletions within the coding sequence that appeared to either truncate the predicted ArsS protein prior to the polycytosine tract or generate a variable C-terminal region (data not shown). Therefore, our data suggest that a number of *arsS* open reading frames are present among most of the clinical populations.

In fact, the seven projected monoclonal *H. pylori* populations yielded *arsS* poly C region cloned sequences that were predicted to encode three of the four major ArsS isoforms (PKI*, LWG*, and EKQ* or SND*) (Figures 3.6 and 3.13). In all, thirty populations had at least one strain that could produce three of the main ArsS isoforms. However, cloned poly C region sequence data suggest that only two primary ArsS isoforms are translated within six of the *H. pylori* populations (Figure 3.13).

A



B

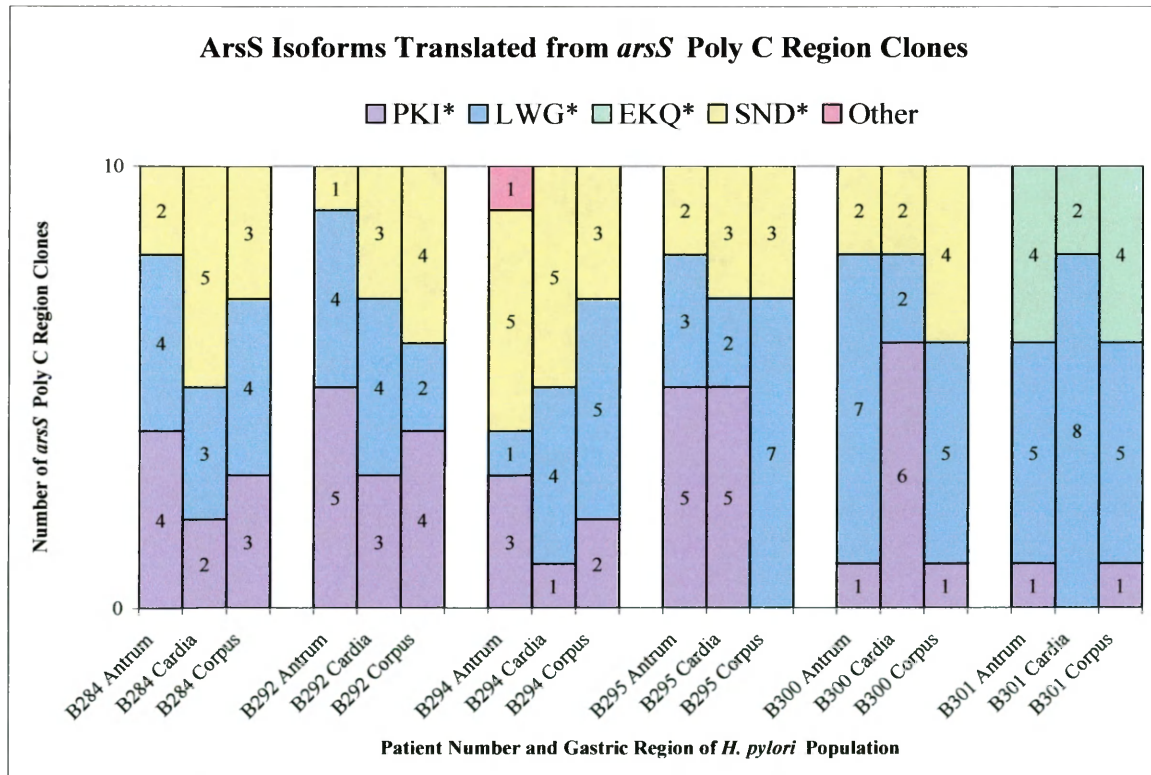


Figure 3.13: Differences in *ArsS* isoform distribution among all *H. pylori* populations.

This stacked bar graph represents the number of clones (stacked columns) within each *H. pylori* population that exhibited differences in open reading frame to allow for the generation of different ArsS isoforms (different bar colors). For A and B, the x-axes represent the patient number and gastric regions from which the *H. pylori* populations were derived. The y-axes represent the differing number of clones capable of translating different ArsS isoforms. Among populations, four main ArsS isoforms were observed and these are denoted by the last three amino acids of the peptide sequence. Within each population, sequence data indicate that at least two ArsS isoforms were present. (A) Distribution of ArsS isoforms when translated *in silico* with cloned *arsS* poly C region sequences generated from the gastric populations of patients B215, B221, B253, B256, B266, and B268. (B) Distribution of ArsS isoforms when translated *in silico* with cloned *arsS* poly C region sequences generated from the gastric populations of patients B284, B292, B294, B295, B300, and B301.

The C-terminal Peptide Sequence of ArsS Isoforms is Variable

Changes in *arsS* open reading frame allow for the production of different ArsS isoforms with C-terminal regions of varying lengths and amino acid sequences (Figure 3.12). The peptide sequences of the four main ArsS isoforms (Figure 3.13) can vary in length by two to as many as thirty-one amino acids past the final proline (P) encoded by the polycytosine tract (Figure 3.14). Thus, the shortest ArsS isoform, PKI*, only has two amino acids past this proline. Furthermore, the longest ArsS isoform, SND*, has thirty-one amino acids past the final proline. Obviously, changes in open reading frame dramatically alter the amino acid sequence among the different ArsS isoforms (Figure 3.14). Interestingly, the EKQ* and SND* isoforms are very similar until the sequences diverge due to a deletion in the stop codon observed in the *arsS* encoding sequence of particular strains (Figures 3.12 and 3.14). Furthermore, there appears to be conservation within the peptide sequence for each of these isoforms, as their amino acid residues do not appear to vary greatly (Figure 3.14).

Figure 3.14 shows the probability (y-axes) of particular amino acid residues (x-axes) in the C-terminal region of the ArsS isoforms. These sequence logos indicate the

hydrophobicity of the amino acid residues with green (neutral), black (hydrophobic), and blue (hydrophilic). Interestingly, hydrophobic and hydrophilic residues occur in relatively similar regions among the C-terminus of the ArsS isoforms (Figure 3.14).

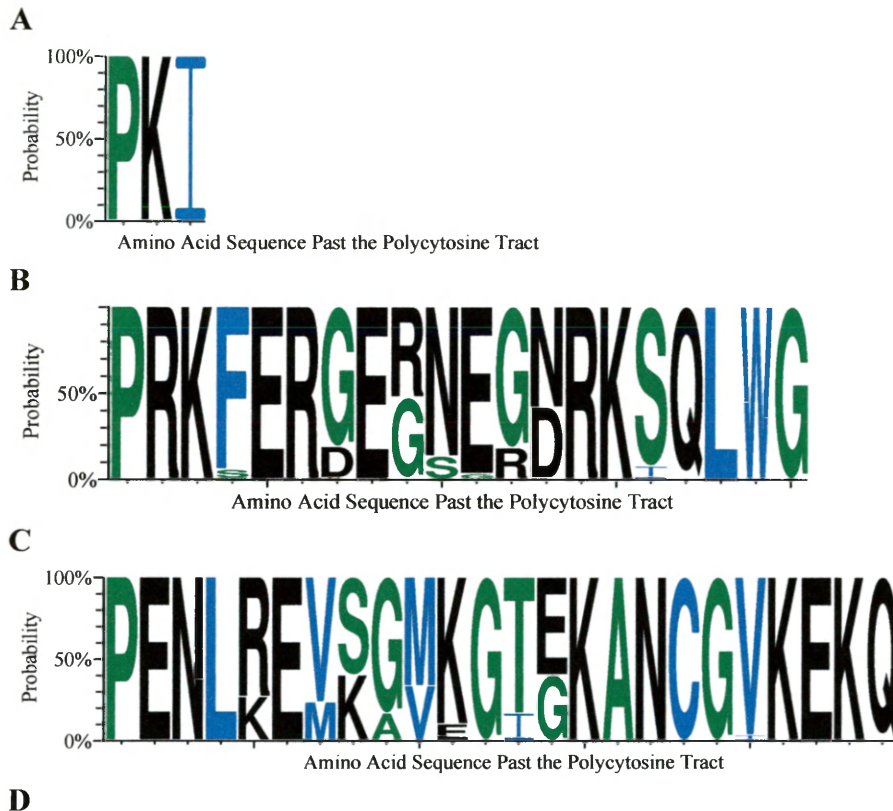


Figure 3.14: Variable C-terminal amino acid sequence of the ArsS isoforms. These sequence logos show the C-terminal amino acid sequences past the polycytosine tract for PKI*, LWG*, EKQ*, and SND* ArsS isoforms. These were intended to show peptide sequence differences and conservation among and within the isoforms. The x-axes represent the amino acid sequences past the final proline (P) encoded by the polycytosine tract. The y-axes represent the probability of an amino acid occurring at a specific site. Amino acids are also colored due to their hydrophathy index values, where hydrophobic amino acids are black, hydrophilic are blue, and neutral are green. (A) Sequence logo for the PKI* isoform, which was generated with 93 cloned poly C region sequences from 31 clinical populations. (B) Sequence logo for the LWG* isoform derived from 158 cloned sequences from 36 clinical populations. (C) Sequence logo for the EKQ* isoform,

which was generated with 30 cloned sequences from 11 populations. (D) Sequence logo for the SND isoform derived from 76 cloned poly C region sequences from 24 populations. (A – D) Variation is prevalent amongst isoforms, however it appears that peptide sequence differences are relatively similar when regarding the EKQ* and SND* isoforms. Within the C-terminal region of these ArsS isoforms, sequence conservation exists in positions that vary in amino acid composition as all locations do not vary by more than a few amino acids.*

Chapter 4

Discussion

The published genome sequences of *Helicobacter pylori* clinical isolates demonstrate that each strain may possess a polycytosine tract of specific length in the 3' terminus of *arsS* (3, 4, 19, 26, 27, 28, 40, 64, 66). Furthermore, one study claimed that these different cytosine tract lengths cause frameshifts allowing for individual strains to encode for single *arsS* alleles, which are subsequently translated into particular ArsS isoforms (5). However, AFLP and sequencing data presented in this thesis research suggest that a single *H. pylori* strain is capable of encoding for multiple *arsS* alleles and therefore various ArsS isoforms. In concordance with this finding, the research group that sequenced *H. pylori* strain J99 observed that different *arsS* polycytosine tract lengths were present among sequencing reads (3). However, they suggested that these multiple tract lengths may have been generated while the organism was cultured *in vitro*. While we have no evidence that disputes this assertion, we do have data suggesting that from six to nine different *arsS* polycytosine tract lengths could be detected within each of the thirty-six populations investigated.

The study describing the whole genome sequence of strain J99 states that only three different polycytosine tract lengths could be detected with their methods, which included minimal subculturing of the organism (3). We also used minimal subculturing with our *H. pylori* populations, but were still able to frequently detect a larger number of polycytosine tract lengths through AFLP and sequencing analyses. However, one may argue that the larger quantity of tract lengths identified in this study may be due to the fact that we investigated entire *H. pylori* populations and not individual isolates as

previous studies have (3, 4, 19, 26, 27, 28, 40, 64, 66). This would be a valid argument, but considering we identified seven monoclonal *H. pylori* populations that appeared to possess *arsS* alleles with various polycytosine tract lengths, *arsS* of individual *H. pylori* strains ought to be considered polymorphic in polycytosine tract length.

The study describing the genome of *H. pylori* strain J99 indicated that there were nucleotide substitutions among the sequences they utilized to generate the entire genomic sequence (3). In the current study, we also detected nucleotide substitutions within regions of *arsS*, but considered each substitution outside of the polycytosine tract to be evidence of distinct strains. Thus, if we had considered such substitutions to be artifacts of the cloning or sequencing processes, as other researchers seemed to (3), more of our populations would likely have been considered to be monoclonal and still capable of generating a number of *arsS* alleles and ArsS isoforms. However, our approach did not ignore nucleotide substitutions, as NCBI-recognized sequenced strains indicate that at least one nucleotide difference can be detected when comparative genomic analyses are performed with the *arsS* regions pertinent to this project (3, 4, 19, 26, 27, 28, 40, 64, 66).

The finding that multiple fragment lengths could be detected from our control *arsS* region via AFLP is another way the validity of our claims could be disputed. Here, we contend that multiple fragment lengths of this region could also be detected through sequencing analyses due to deletions in the primer binding regions of the *arsS* poly C and control region amplicons. With a single nucleotide deletion in these regions, fragment length sizing data would obviously be affected by one base pair. Thus, we acknowledge that some poly C region and control fragment lengths may have been generated due to the misconstruction of oligonucleotide primers. Errors occurring in the primer synthesis

process are likely inevitable, but the quantity of flawed primers used in this research were likely minimal as only seven (1.7%) out of four-hundred and twenty clones encoded deletions in such regions. Yet, these data still lead us to believe that these errors may have at least partially affected AFLP fragment length data.

Another way erroneous fragment sizes may have been generated is through the nonspecific binding of oligonucleotide primers due to the genomes of *H. pylori* being so remarkably plastic (1, 24). The primers used to amplify poly C and control region sequences in this research were developed using consensus sequences of strains 26695 and J99 (3, 66). However, not all of the NCBI-recognized strains were always homologous in any of these short regions (3, 4, 19, 26, 27, 28, 40, 64, 66). Thus, not all of the strains within the populations utilized in this research were likely homologous in these regions either. Mismatches between oligonucleotide primers and the primer binding regions of the genomes may have promoted the mutagenesis of amplicons causing the insertion, deletion, or substitution of bases. If insertions or deletions were to occur, then AFLP fragment length data would be affected. However, the fact that sequencing analyses suggested that two-hundred and ninety (80.5%) of the total three-hundred and sixty clones encoded for three different, yet consecutive polycytosine tract lengths leads us to believe that at least three of the AFLP fragments detected within each population were real. Thus, *arsS* alleles encoding for three open reading frames of ArsS would also be existent within each population or, in some cases, each strain.

Conversely, an allegation such as this is disputed by the claim that individual *H. pylori* strains encode for a particular *arsS* allele and ArsS isoform (5). Beier *et al.* also claim that they have isolated the transmitter domain of individual ArsS isoforms and

subsequently determined that these isoforms are not hindered in their kinase activities (5). They acknowledge that previous studies have shown that if the transmitter domain of a histidine kinase, such as ArsS, is expressed *in vitro*, then it will typically be constitutively “on” or able to autophosphorylate at all times (5). Therefore, these researchers expressed a “single” isoform of ArsS from three strains and demonstrate that these transmitter domains are able to autophosphorylate in the presence of radiolabeled ATP (5). However, our data indicate that they may have expressed any of the ArsS isoforms from individual strains unless careful considerations were taken. This group states that each expression vector was sequenced to verify polycytosine tract length and thus the ArsS isoform expressed (5). However, they did not perform AFLP analyses on their plasmid preparations to confirm that only vectors encoding for one polycytosine tract length were present (5). Therefore, the ArsS protein that was able to autophosphorylate in each assay may have always been the same isoform. We make this claim because a control AFLP study was not utilized to authenticate that only one isoform was actually being utilized in their phosphorylation assays.

In spite of this, we do not hypothesize that only one isoform of ArsS is functional. Again, considering the plasticity of *H. pylori* genomes, we typically observed little variation within the C-terminal peptide sequences of particular ArsS isoforms among different bacterial populations (1, 24). For instance, the LWG* isoform, encoded by one-hundred and fifty-eight clones from all thirty-six populations, exhibited conservation in twelve (63.2%) of nineteen amino acid residues following the final proline encoded by the polycytosine tract. Furthermore, the remaining seven (36.8%) amino acids that were not conserved only varied by two amino acids at any of these locations. Similar

conservation is also observed among the remaining isoforms, most notably the PKI* isoform. The two amino acids between the final proline and the terminal end of this protein were always conserved in ninety-three clones from thirty-one different populations. Interestingly, the C-terminal peptide sequence of the EKQ* and SND* isoforms are similarly conserved until they diverge in length due to a previously unrecognized polymorphic deletion in the *arsS* encoding sequence.

When Beier *et al.* performed ArsS phosphorylation assays, the group only expressed the PKI*, LWG*, and EKQ* isoforms of the protein (5). However, they did not even recognize the SND* isoform as being existent, which could be due to two possibilities. One, they may have not utilized strains that possessed the thymine deletion in the third *arsS* stop codon and therefore only had strains capable of expressing EKQ*. The second possibility is they may have had strains with the deletion, but their sequencing analyses showed that the polycytosine tract was not of correct length to encode the necessary *arsS* open reading frame for SND* translation. In fact, this deletion is only observed in two (22.2%) of the nine NCBI-recognized *H. pylori* sequenced strains (3, 4, 19, 26, 27, 28, 40, 64, 66). However, the polycytosine tract lengths of these particular strains, J99 and B38, do not suggest that their 3' terminal regions of *arsS* are in frame to allow for the generation of the SND* isoform (3, 64). Thus, the possibility exists that the ArsS histidine kinase SND* isoform has not been recognized until described in this current study.

Our data indicate that a large number of the populations investigated here have strains that encode for the polymorphic deletion. In fact, the presence of the deletion tended to be patient specific, rather than population or even strain specific. Thus, we

detected the deletion in each of the thirty clones from the gastric populations of eight patients. Furthermore, we also observed the deletion in only a single cloned *arsS* poly C region amplicon derived from an *H. pylori* population that seemed to mostly consist of strains that did not possess this stop codon deletion. In total, we observed that 241 (66.9%) of 360 cloned *arsS* poly C region sequences possessed the thymine deletion. Of these, seventy-six (31.5%) encoded polycytosine tracts of specific lengths allowing for SND* translation.

The fact that we have observed eight (66.7%) of our twelve patients to be infected with strains that only seem to encode for the thymine deletion, may be due to a bias with patient locality as all patients were treated in Nashville, Tennessee. Interestingly, *H. pylori* J99, which encodes the deletion, was also isolated from a patient in the middle Tennessee region (3). However, to counteract the notion that the deletion may be restricted to strains observed in the United States, strain B38 was isolated from a French patient (64). Regardless, both NCBI-recognized strains and the clinical populations investigated here, have the collective ability to encode for open reading frames allowing for the generation of PKI*, LWG*, EKQ* and SND* ArsS isoforms and these isoforms differ not only in amino acid sequence, but in overall length of the primary structures (3, 4, 19, 26, 27, 28, 40, 64, 66). Thus, these findings should prompt future research with these isoforms of ArsS.

Conclusion

Here, we conclude that entire *H. pylori* populations harvested from the gastric antrum, cardia, and corpus of twelve patients were observed to be polymorphic in regards to the 3' terminus of the *arsS* gene, which encodes for the histidine kinase ArsS. We

determined that polymorphic differences were due to variable *arsS* polycytosine tract lengths that promoted the generation of multiple alleles and reading frames of *arsS*. A number of *arsS* alleles could be detected from each population as well as a number of *arsS* open reading frames. Thus, each *H. pylori* population appears to be capable of expressing multiple isoforms of the ArsS protein as well, including an isoform that had not been previously described. The ability of this isoform to be expressed was due to another polymorphism observed in populations mainly from specific patients in mostly an “all or none” fashion, as virtually all (99.7%) cloned sequences from the gastric populations of patients either possessed the thymine deletion or encoded a functional stop codon.

Interestingly, the clinical *H. pylori* populations seemed to possess predominant *arsS* alleles and thus are predicted to express predominant ArsS isoforms. While we still need to elucidate the advantage of this variation in the histidine kinase C-terminal domain, we did find that *arsS* and ArsS predominance was variable among all populations and, in some cases, within populations from a single patient. We speculate that this observation could be due to a number of factors, but we hypothesize that these alleles or isoforms may confer adaptive advantages on *H. pylori* and contribute to its ability to persist in the human host for decades (48). However, a great deal of future research must be performed to answer questions such as how various isoforms of ArsS differ in function.

The findings presented in this thesis research are important because they challenge the assertions of previous studies that indicated *arsS* is only polymorphic among *H. pylori* strains and not within them (3, 5). This research is also significant

because we have acknowledged a previously unmentioned isoform of ArsS due to yet another polymorphism in the coding sequence of *arsS*. Our data also indicate that similarities and differences among and within the C-terminal regions of various ArsS isoforms may be important. Thus, in totality, this research adds to the limited knowledge concerning the ArsRS two-component signal transduction system of the carcinogenic human pathogen, *Helicobacter pylori*.

Future Directions

Generating More Accurate AFLP Data

We previously mentioned that fragment length polymorphisms for our control and poly C region amplicons were detected. We contended that the differences in control region amplicons and some of the variations in poly C region amplicons lengths may have been partially due to oligonucleotide synthesis errors. Thus, future research should consider performing AFLP analyses with primers that have been purified through high-pressure liquid chromatography (HPLC), as this process should eliminate primers that are polymorphic in length. If possible, primers should also be developed to be strain specific because *H. pylori* genomes are recognized to be very plastic among unrelated strains (1, 24). This consideration is important because nonspecific binding could be attributing to different amplicons lengths detected via AFLP.

Another way AFLP analyses can be improved is with the use of a better sizing standard and fluorescent label. Most of our data indicate that fractions of base pairs were detected, but we know that these are not real and must be artifacts of the AFLP process. Thus, utilizing a sizing standard and fluorescent label that could assist the AFLP

machinery and software more accurately recognize distinct amplicon sizes would be beneficial to the entire process.

Even though we observed differences in control region amplicon lengths, we do not consider our AFLP data to be invalid. This is because the frequency distributions at which the control region amplicons were detected were much different than those of the poly C region. Thus, we do acknowledge that some peaks were generated as artifacts of the entire process, but most of these lengths are likely not. In any case, making all of the listed changes to the AFLP process would likely strengthen future data because the accuracy and precision of the results should be considerably improved.

Selection of ArsS Isoforms

AFLP and sequencing data did not only indicate the mere presence of *arsS* alleles and ArsS isoforms, as these data also revealed their relative frequencies and predominance within each population. Here, we hypothesize that the relative frequency and predominance of alleles and, therefore, isoforms within the populations may be due to the adaptive advantages that each isoform may confer. Considering that ArsS is a recognized sensor of acidity, one may speculate that different ArsS isoforms may deviate in their response to different pH levels (47, 61). Furthermore, since the high mutation rates of *H. pylori* have been attributed to oxidative stress and acidity, possibility exists that either of these stressors, or their combined efforts, could allow for the propagation of strains encoding for particular isoforms (7, 30, 69). Our laboratory tried to develop methods to select for strains encoding specific *arsS* alleles in the presence of variable pH, but controlling the acidity of the media proved to be an arduous task as problems frequently arose. However, future studies should focus on developing better methods to

ascertain the influence of environmental stressors on the selection of *H. pylori* strains expressing particular ArsS isoforms.

ArsS C-terminal Protein Structure

Given that the different isoforms of ArsS varied in peptide length past the final proline, translated from the polycytosine tract, by two to as many as thirty-one different amino acids, a study should be conducted to determine if these variations in sequence affect the protein structure of ArsS. Recently, a group determined that the C-terminal peptide sequences of particular proteins in *H. canadensis* were affected by frameshifts due to differences in polymeric tract lengths of their 3' gene encoding sequences (59). Furthermore, the group analyzed the variable protein structures with the NCBI Conserved Domain Database (CDD) and found that these structures were likely affected (59). While the CDD does not infer any major structural differences for our ArsS isoforms, possibility exists that the isoforms may fold differently. We hypothesize that ArsS isoforms may have alternate folding patterns because of their variable C-terminal peptide sequences. Interestingly, the C-terminal sequence within each ArsS isoform appears to be conserved, suggesting that this region of ArsS may be important. Considering that hydrophobic and hydrophilic amino acids appeared to be similarly located among the different ArsS isoforms, we believe that this region may be involved in protein folding. Thus, approaches could be taken to isolate each isoform of ArsS and determine if their protein structures vary.

Phosphorylation Abilities of ArsS

Even if we determined that strains expressing particular isoforms of ArsS can be selected or if different ArsS isoforms have alternate folding patterns, we would still want

to perceive the functional differences among the isoforms, if any. Beier *et al.* suggested that they had individually expressed the PKI*, LWG*, and EKQ* ArsS isoforms and determined that none of these were defunct in their phosphorylation or phosphotransfer capabilities (5). Due to discrepancies with their methods, possibility exists that they did not successfully develop individual expression plasmid preparations that were void of alternate *arsS* open reading frames. Thus, we suggest that when such expression vectors are produced, sequencing and AFLP analyses be performed to verify that each preparation is only able to express one ArsS isoform. Once these expression vector preparations are achieved, similar phosphorylation assays can be performed as described by Beier *et al.* (5). We would also want to include the SND* isoform in our investigation as this group did not recognize the existence of this particular ArsS isoform (5). Furthermore, the assay should be executed to determine if differences exist regarding the phosphorylation rates and efficiencies for each isoform.

Differences in ArsRS Operons

Previous transcriptome analyses have shown that there are differences in the number of genes that the ArsRS TCST is able to regulate (20, 31, 32, 47, 71). If different ArsS isoforms are involved in the generation of variant transcriptomes, then our data suggest that this may be due to predominant ArsS isoforms. We observed differences in predominant *arsS* alleles and ArsS isoforms among and within the clinical populations. If the predominance of particular isoforms is linked to bestowing adaptive advantages, then the transcriptome of the organism could be affected. Thus, we hypothesize that predominant isoforms are affecting the transcriptome of the clinical populations. In order to determine variations among transcriptomes, we would prefer to develop strains that are

individually capable of expressing one ArsS isoform. Once we generate and verify these mutant strains, we could perform analyses with the transcriptomes to determine if ArsS isoform type is playing a role in the variations previously observed (20, 31, 32, 47, 71).

Strain-specific ArsS Predominance

We have acknowledged that different predominant strains can be observed among the gastric populations of four (33.3%) patients. In three of these cases, predominant ArsS isoform also varied accordingly. Thus, we must recognize that ArsS isoform predominance could be a strain specific phenomenon, especially considering that prior studies have concluded that individual *H. pylori* strains have single 3' terminus *arsS* polycytosine tract lengths (3, 5). We do not mean to contradict our data here because we certainly have evidence that counteract these claims, but considering all routes is necessary to accurately determine the answer to the question at hand. Given that Alm *et al.* stated that the generation of various polycytosine tract lengths is an artifact of the subculturing process, future studies need to isolate *H. pylori* cells promptly for AFLP analyses. We know that one cell of *H. pylori* encodes for a single *arsS* polycytosine tract length, but do not know if this tract length mutates explicitly as the cell replicates in the absence of a selective pressure. Thus, we would want to isolate single colonies from the populations to perform our AFLP analyses. Theoretically, our data suggest that we should be able to detect single colonies that are of the same strain, but encode for variable polycytosine tract lengths and thus express different ArsS isoforms. This analysis would be essential in debunking the previous claims that particular strains of *H. pylori* encode for specific polycytosine tract lengths (3, 5). Furthermore, we would be able to

accurately determine if *arsS* allele and ArsS isoform predominance is linked to specific strains or an unknown factor.

Appendices

Appendix 1: General Broth and Agar Media

Luria-Bertani (LB) Broth

LB broth was produced with 1% tryptone, 0.5% yeast extract, and 85mM NaCl dissolved in dH₂O. Media was autoclaved and allowed to cool to 22°C before the addition of antibiotics.

Luria-Bertani (LB) Agar

LB agar was produced with 1% tryptone, 0.5% yeast extract, 85mM NaCl, and 1.5% agar dissolved in dH₂O. Media was autoclaved and allowed to cool to 55°C before the addition of antibiotics and plate pouring.

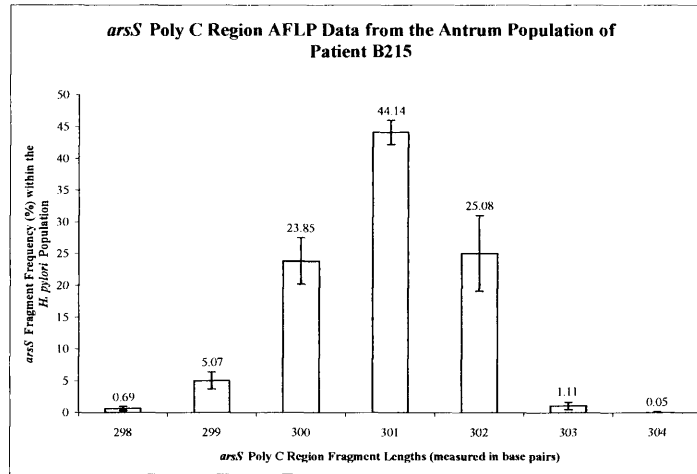
Sulfite-Free Brucella Broth (SFBB)

SFBB was produced with 1% tryptone, 1% peptone, 0.2% yeast extract, 5.5mM glucose, and 85mM NaCl dissolved in dH₂O. Media was autoclaved and allowed to cool to 22°C before the addition of 5% (v/v) Newborn Calf Serum (NCS) and antibiotics.

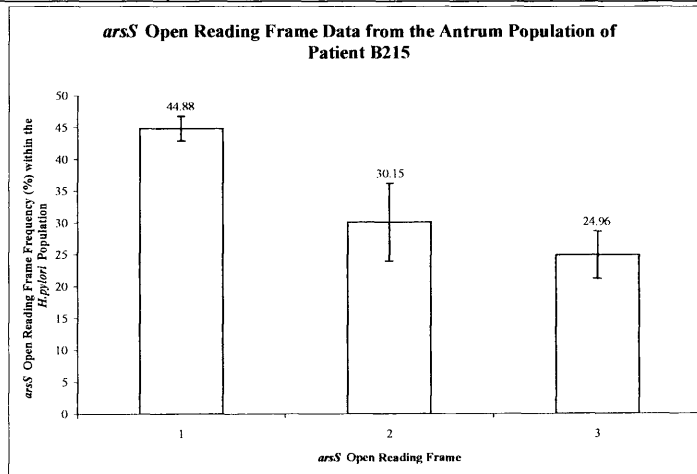
Super Optimal Broth with Catabolite Repression (SOC)

SOC broth was produced with 2% tryptone, 0.5% yeast extract, 10mM NaCl, 2.5mM KCl, 10mM MgCl₂, 10mM MgSO₄, and 20mM glucose dissolved in dH₂O. Media was autoclaved and allowed to cool to 22°C before the addition of antibiotics.

Appendix 2A: AFLP Data from Poly C Regions of *H. pylori* Populations
Antrum Population of Patient B215



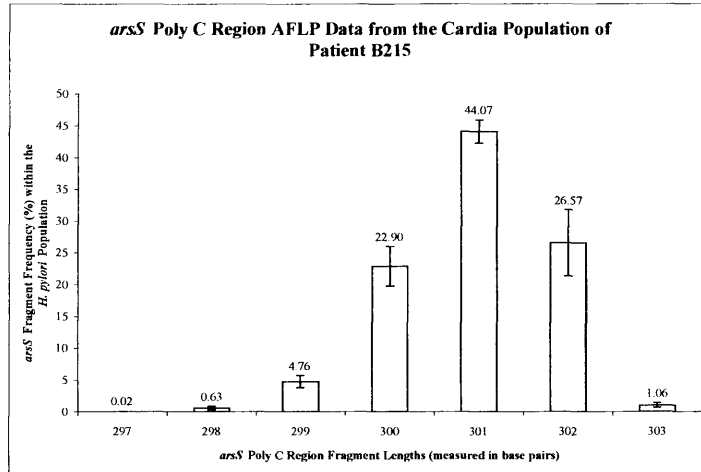
Fragment Length (bp)	Frequency	Standard Deviation
298	0.69	0.33
299	5.07	1.30
300	23.85	3.66
301	44.14	1.92
302	25.08	5.97
303	1.11	0.61
304	0.05	0.11



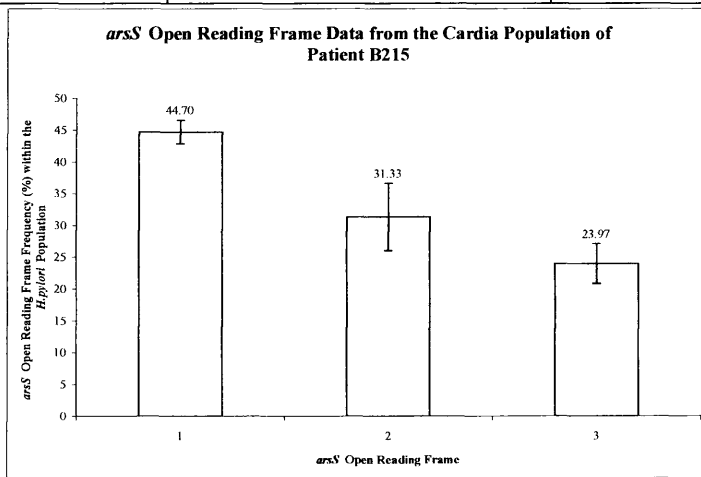
Open Reading Frame	Frequency	Standard Deviation
1	44.88	1.95
2	30.15	6.11
3	24.96	3.71

T-test	T-value	P-value
ORF1 vs. ORF2	6.90	1.80×10^{-6}
ORF2 vs. ORF3	2.18	2.23×10^{-2}
ORF1 vs. ORF3	14.27	8.04×10^{-11}

Cardia Population of Patient B215



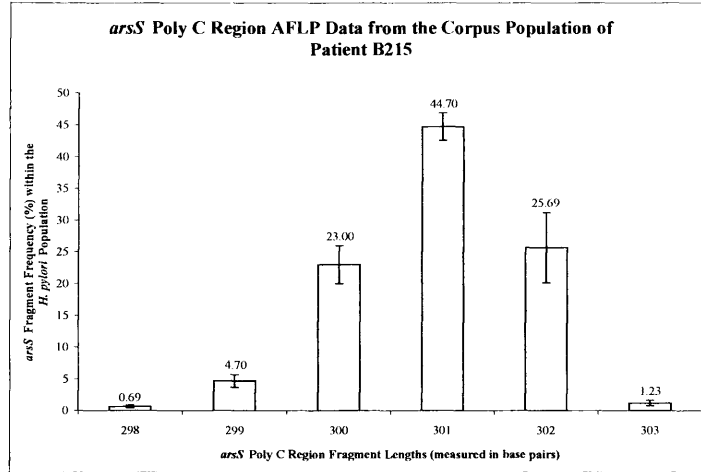
Fragment Length (bp)	Frequency	Standard Deviation
297	0.02	0.05
298	0.63	0.28
299	4.76	0.95
300	22.90	3.13
301	44.07	1.83
302	26.57	5.21
303	1.06	0.36



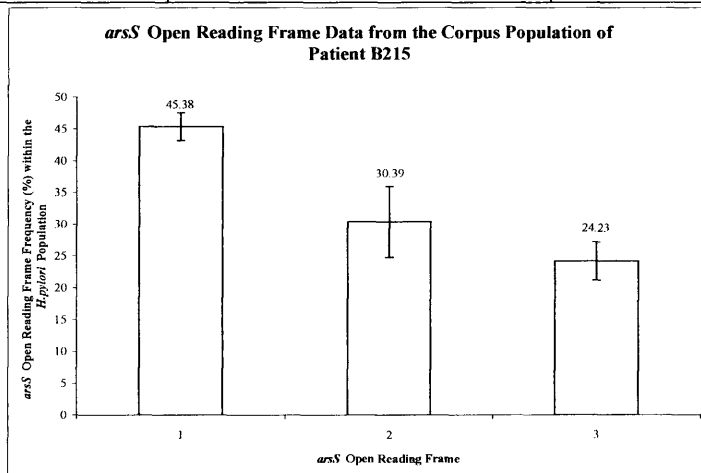
Open Reading Frame	Frequency	Standard Deviation
1	44.70	1.85
2	31.33	5.30
3	23.97	3.15

T-test	T-value	P-value
ORF1 vs. ORF2	7.14	1.16×10^{-6}
ORF2 vs. ORF3	3.58	1.25×10^{-3}
ORF1 vs. ORF3	17.02	5.66×10^{-12}

Corpus Population of Patient B215



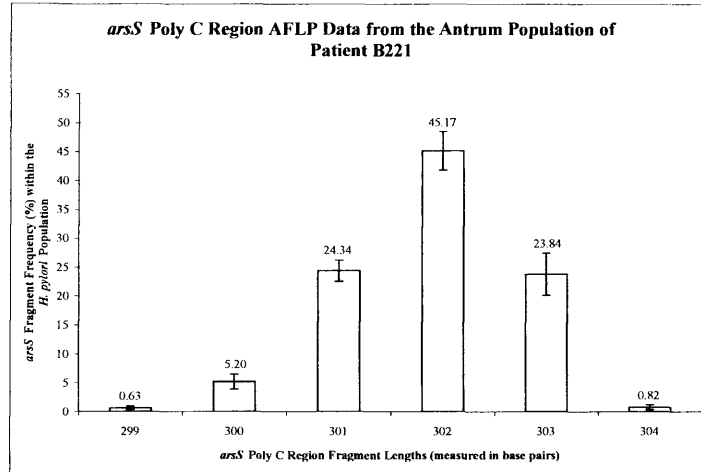
Fragment Length (bp)	Frequency	Standard Deviation
298	0.69	0.23
299	4.70	0.98
300	23.00	2.98
301	44.70	2.17
302	25.69	5.51
303	1.23	0.43



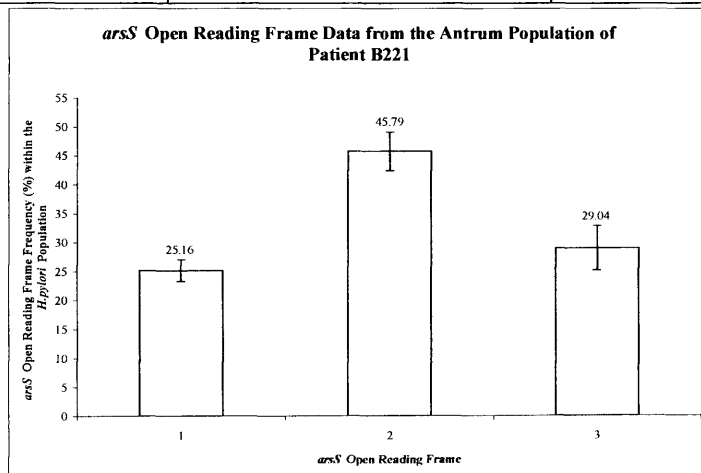
Open Reading Frame	Frequency	Standard Deviation
1	45.38	2.18
2	30.39	5.60
3	24.23	3.01

T-test	T-value	P-value
ORF1 vs. ORF2	7.48	6.51×10^{-7}
ORF2 vs. ORF3	2.91	5.12×10^{-3}
ORF1 vs. ORF3	17.07	5.40×10^{-12}

Antrum Population of Patient B221



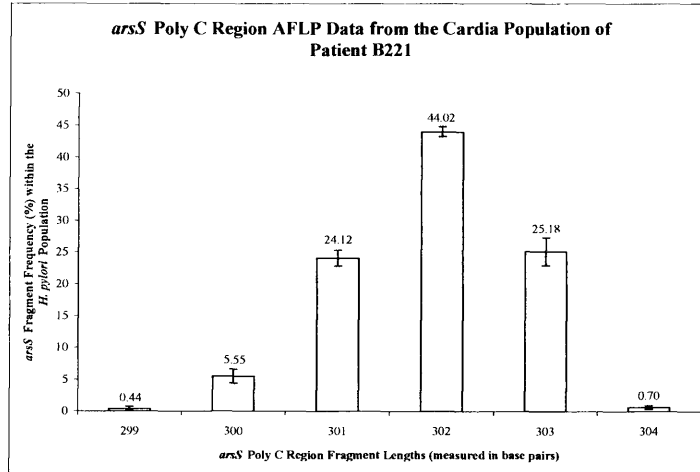
Fragment Length (bp)	Frequency	Standard Deviation
299	0.63	0.30
300	5.20	1.28
301	24.34	1.82
302	45.17	3.35
303	23.84	3.63
304	0.82	0.42



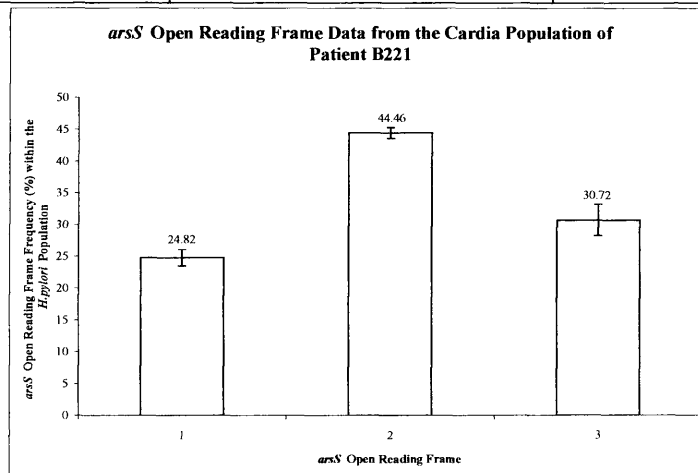
Open Reading Frame	Frequency	Standard Deviation
1	25.16	1.87
2	45.79	3.36
3	29.04	3.85

T-test	T-value	P-value
ORF1 vs. ORF2	16.10	1.31×10^{-11}
ORF2 vs. ORF3	9.83	1.75×10^{-8}
ORF1 vs. ORF3	2.72	7.53×10^{-3}

Cardia Population of Patient B221



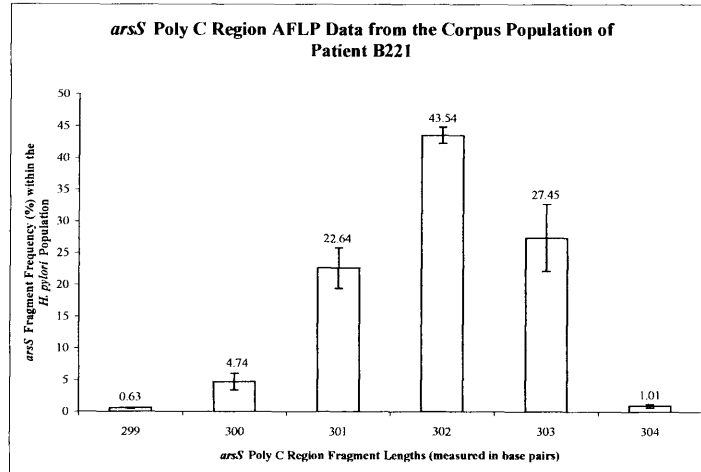
Fragment Length (bp)	Frequency	Standard Deviation
299	0.44	0.34
300	5.55	1.12
301	24.12	1.25
302	44.02	0.79
303	25.18	2.19
304	0.70	0.27



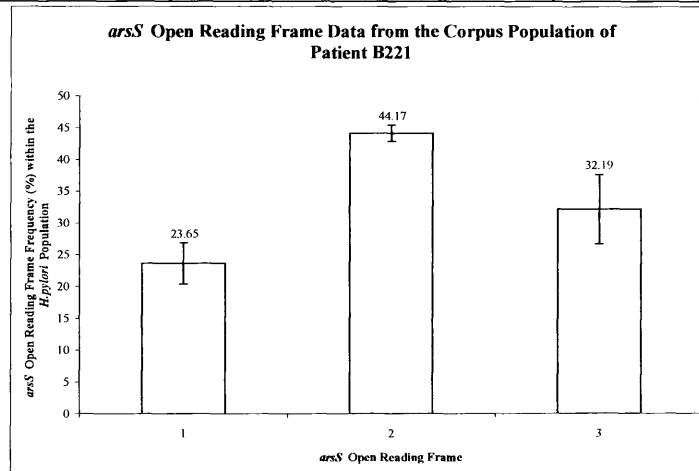
Open Reading Frame	Frequency	Standard Deviation
1	24.82	1.28
2	44.46	0.86
3	30.72	2.46

T-test	T-value	P-value
ORF1 vs. ORF2	38.25	1.85×10^{-17}
ORF2 vs. ORF3	15.82	1.71×10^{-11}
ORF1 vs. ORF3	6.39	4.44×10^{-6}

Corpus Population of Patient B221



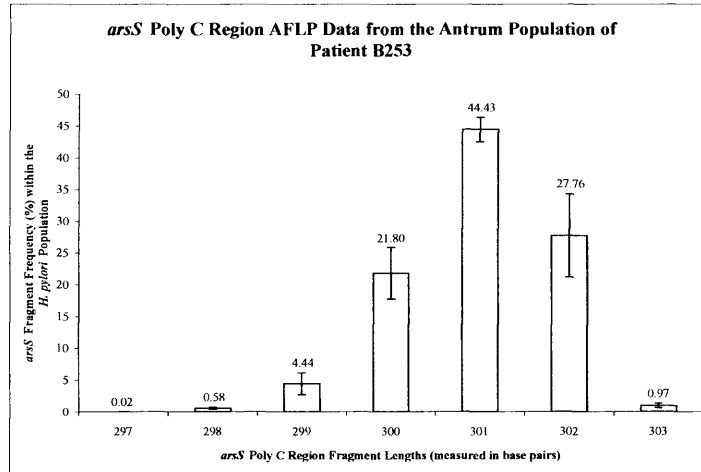
Fragment Length (bp)	Frequency	Standard Deviation
299	0.63	0.07
300	4.74	1.30
301	22.64	3.22
302	43.54	1.26
303	27.45	5.28
304	1.01	0.27



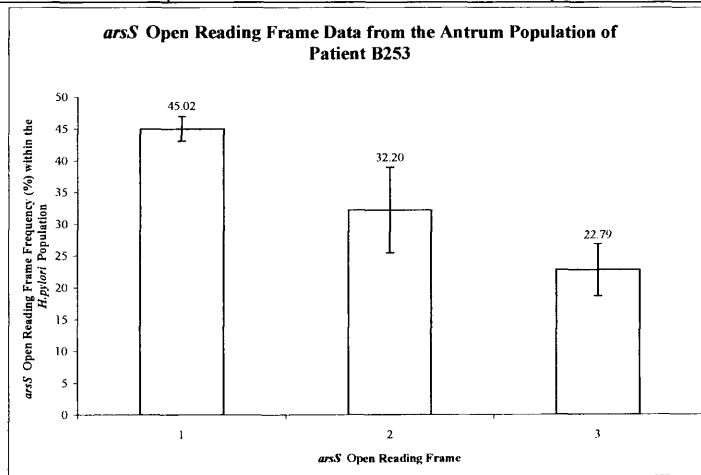
Open Reading Frame	Frequency	Standard Deviation
1	23.65	3.23
2	44.17	1.27
3	32.19	5.43

T-test	T-value	P-value
ORF1 vs. ORF2	17.74	3.02×10^{-12}
ORF2 vs. ORF3	6.44	4.09×10^{-6}
ORF1 vs. ORF3	4.05	4.61×10^{-4}

Antrum Population of Patient B253



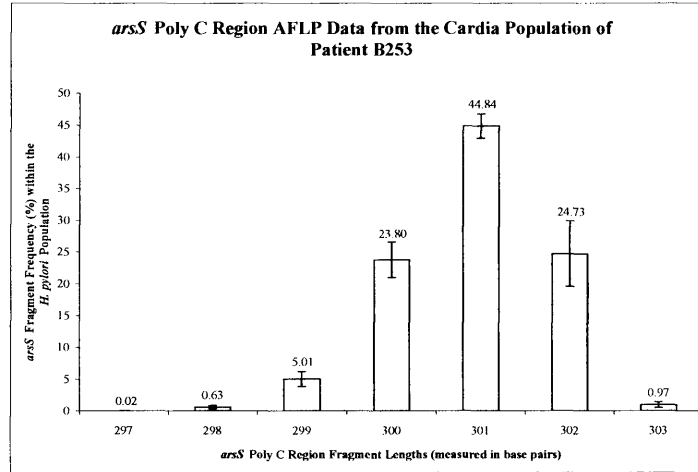
Fragment Length (bp)	Frequency	Standard Deviation
297	0.02	0.06
298	0.58	0.16
299	4.44	1.70
300	21.80	4.07
301	44.43	1.93
302	27.76	6.53
303	0.97	0.34



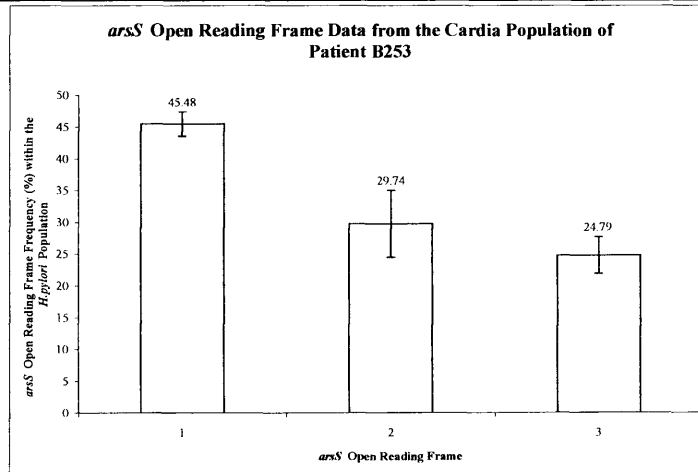
Open Reading Frame	Frequency	Standard Deviation
1	45.02	1.94
2	32.20	6.74
3	22.79	4.09

T-test	T-value	P-value
ORF1 vs. ORF2	5.48	2.52×10^{-5}
ORF2 vs. ORF3	3.58	1.25×10^{-3}
ORF1 vs. ORF3	14.74	4.97×10^{-11}

Cardia Population of Patient B253



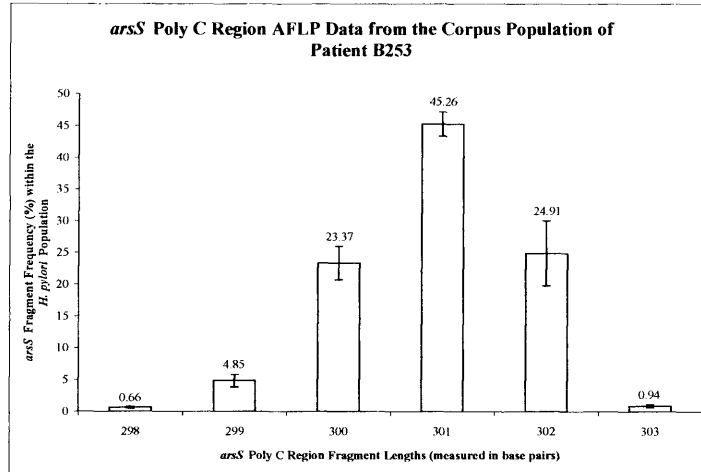
Fragment Length (bp)	Frequency	Standard Deviation
297	0.02	0.05
298	0.63	0.27
299	5.01	1.18
300	23.80	2.82
301	44.84	1.92
302	24.73	5.13
303	0.97	0.47



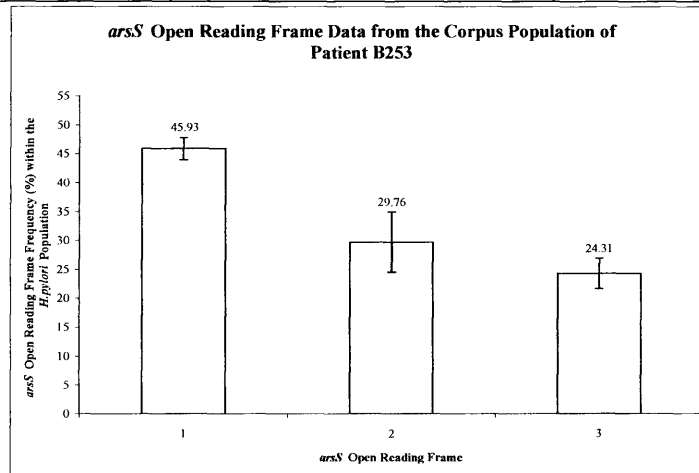
Open Reading Frame	Frequency	Standard Deviation
1	45.48	1.94
2	29.74	5.27
3	24.79	2.86

T-test	T-value	P-value
ORF1 vs. ORF2	8.41	1.44×10^{-7}
ORF2 vs. ORF3	2.48	1.24×10^{-2}
ORF1 vs. ORF3	17.94	2.55×10^{-12}

Corpus Population of Patient B253



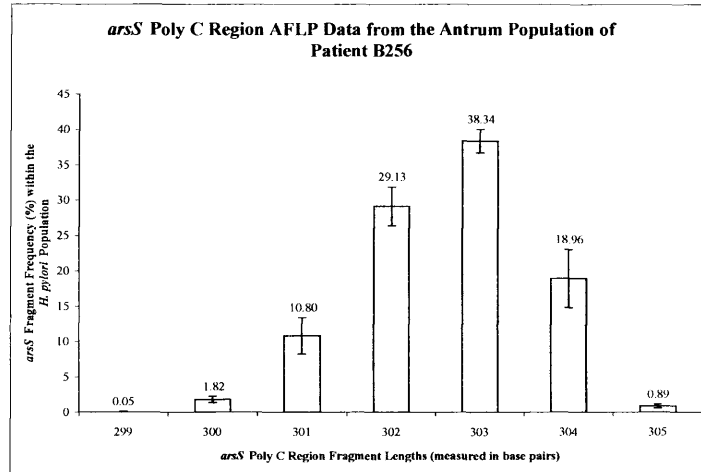
Fragment Length (bp)	Frequency	Standard Deviation
298	0.66	0.15
299	4.85	0.98
300	23.37	2.62
301	45.26	1.92
302	24.91	5.13
303	0.94	0.25



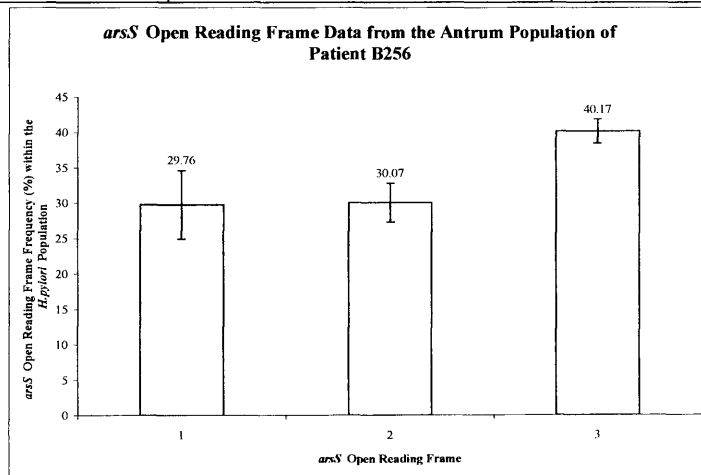
Open Reading Frame	Frequency	Standard Deviation
1	45.93	1.92
2	29.76	5.22
3	24.31	2.63

T-test	T-value	P-value
ORF1 vs. ORF2	8.71	8.97×10^{-8}
ORF2 vs. ORF3	2.80	6.48×10^{-3}
ORF1 vs. ORF3	19.89	5.20×10^{-13}

Antrum Population of Patient B256



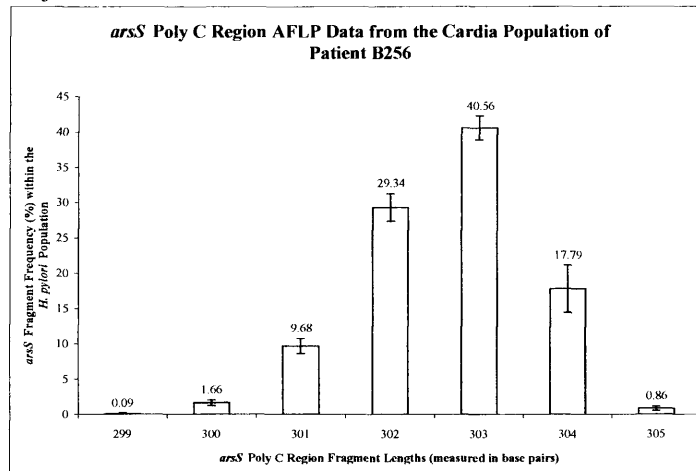
Fragment Length (bp)	Frequency	Standard Deviation
299	0.05	0.10
300	1.82	0.42
301	10.80	2.58
302	29.13	2.74
303	38.34	1.67
304	18.96	4.10
305	0.89	0.27



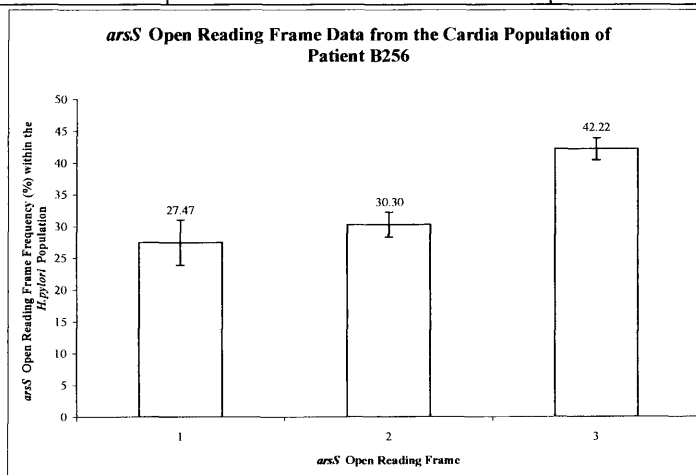
Open Reading Frame	Frequency	Standard Deviation
1	29.76	4.84
2	30.07	2.76
3	40.17	1.72

T-test	T-value	P-value
ORF1 vs. ORF2	0.17	4.34×10^{-1}
ORF2 vs. ORF3	9.32	3.63×10^{-8}
ORF1 vs. ORF3	6.07	8.05×10^{-6}

Cardia Population of Patient B256



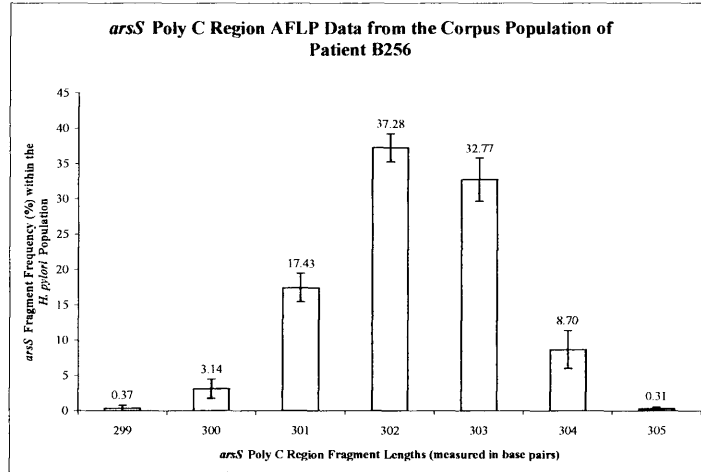
Fragment Length (bp)	Frequency	Standard Deviation
299	0.09	0.16
300	1.66	0.40
301	9.68	1.10
302	29.34	1.94
303	40.56	1.67
304	17.79	3.37
305	0.86	0.30



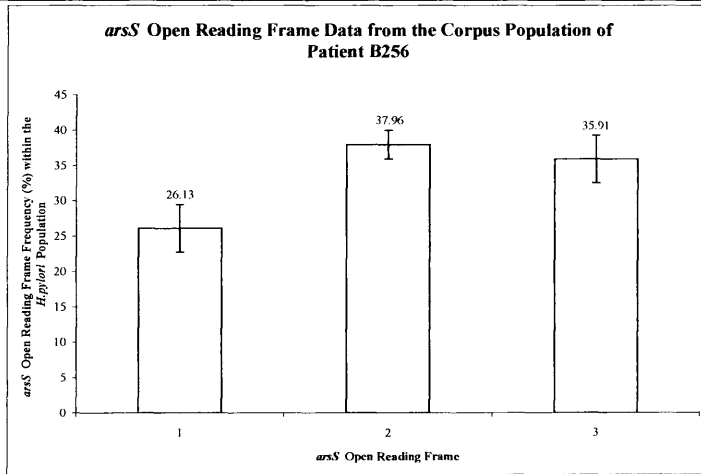
Open Reading Frame	Frequency	Standard Deviation
1	27.47	3.55
2	30.30	1.97
3	42.22	1.72

T-test	T-value	P-value
ORF1 vs. ORF2	2.09	2.65×10^{-2}
ORF2 vs. ORF3	13.67	1.52×10^{-10}
ORF1 vs. ORF3	11.23	2.68×10^{-9}

Corpus Population of Patient B256



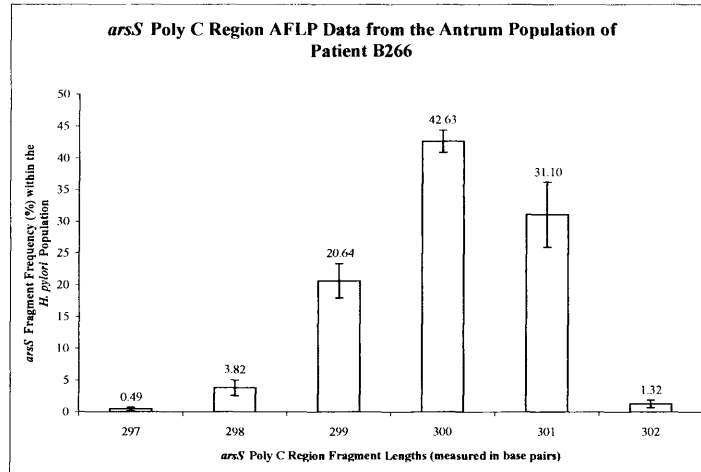
Fragment Length (bp)	Frequency	Standard Deviation
299	0.37	0.42
300	3.14	1.36
301	17.43	2.03
302	37.28	1.98
303	32.77	3.07
304	8.70	2.70
305	0.31	0.24



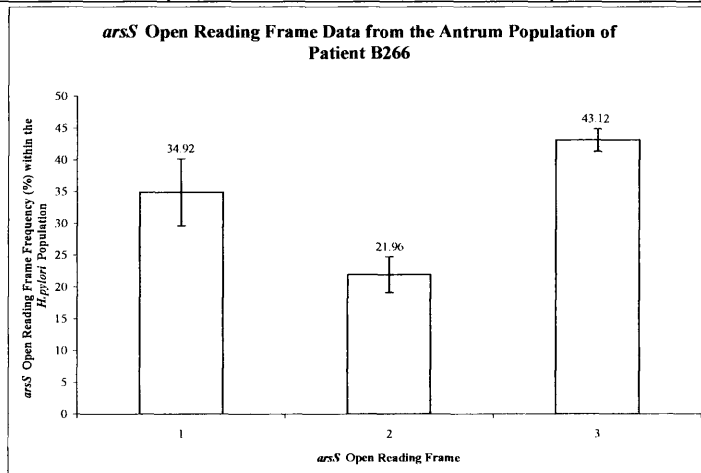
Open Reading Frame	Frequency	Standard Deviation
1	26.13	3.37
2	37.96	2.04
3	35.91	3.36

T-test	T-value	P-value
ORF1 vs. ORF2	9.00	5.78×10^{-8}
ORF2 vs. ORF3	1.56	6.86×10^{-2}
ORF1 vs. ORF3	6.17	6.79×10^{-6}

Antrum Population of Patient B266



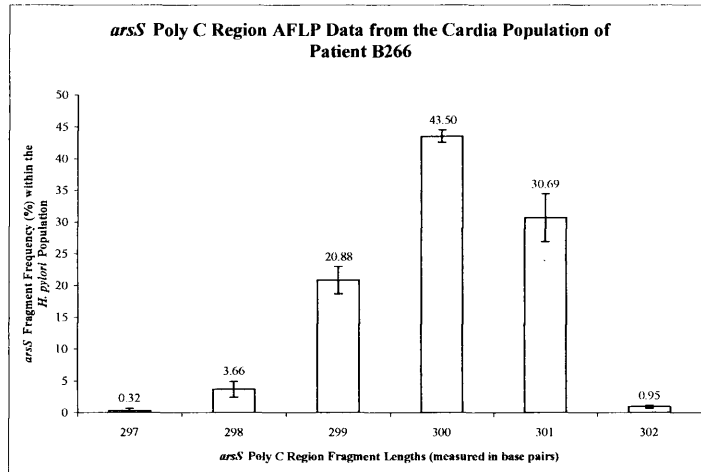
Fragment Length (bp)	Frequency	Standard Deviation
297	0.49	0.28
298	3.82	1.24
299	20.64	2.73
300	42.63	1.76
301	31.10	5.13
302	1.32	0.60



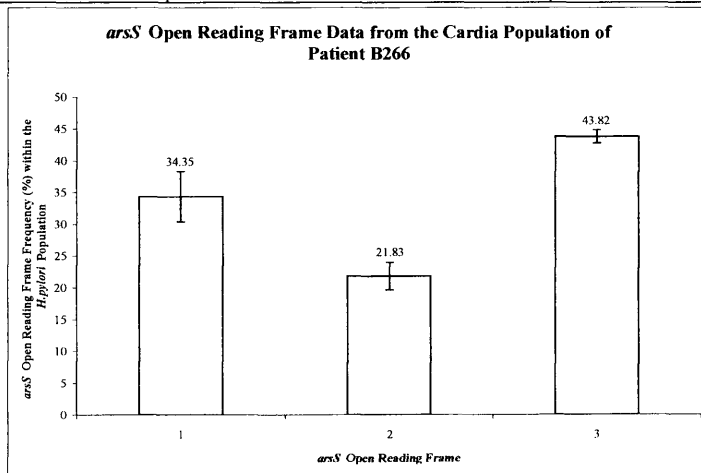
Open Reading Frame	Frequency	Standard Deviation
1	34.92	5.27
2	21.96	2.79
3	43.12	1.78

T-test	T-value	P-value
ORF1 vs. ORF2	6.51	3.57×10^{-6}
ORF2 vs. ORF3	19.17	9.18×10^{-13}
ORF1 vs. ORF3	4.42	2.14×10^{-4}

Cardia Population of Patient B266



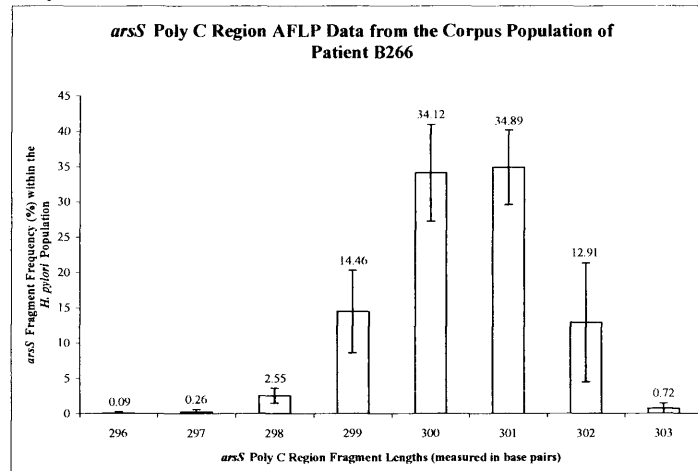
Fragment Length (bp)	Frequency	Standard Deviation
297	0.32	0.36
298	3.66	1.23
299	20.88	2.17
300	43.50	0.96
301	30.69	3.77
302	0.95	0.22



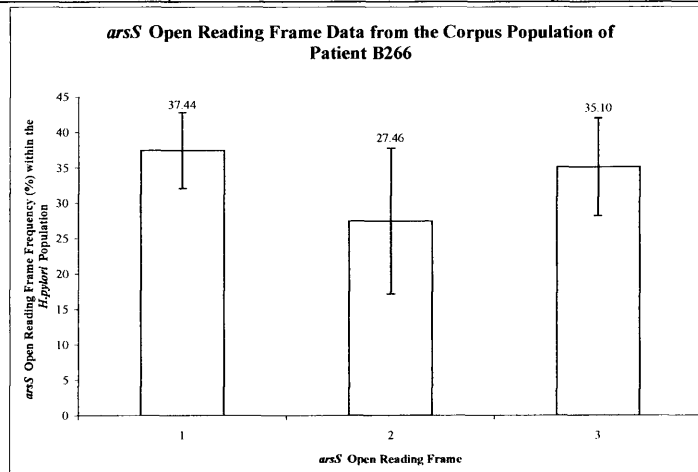
Open Reading Frame	Frequency	Standard Deviation
1	34.35	3.97
2	21.83	2.18
3	43.82	1.03

T-test	T-value	P-value
ORF1 vs. ORF2	8.30	1.72×10^{-7}
ORF2 vs. ORF3	27.38	3.60×10^{-15}
ORF1 vs. ORF3	6.93	1.68×10^{-6}

Corpus Population of Patient B266



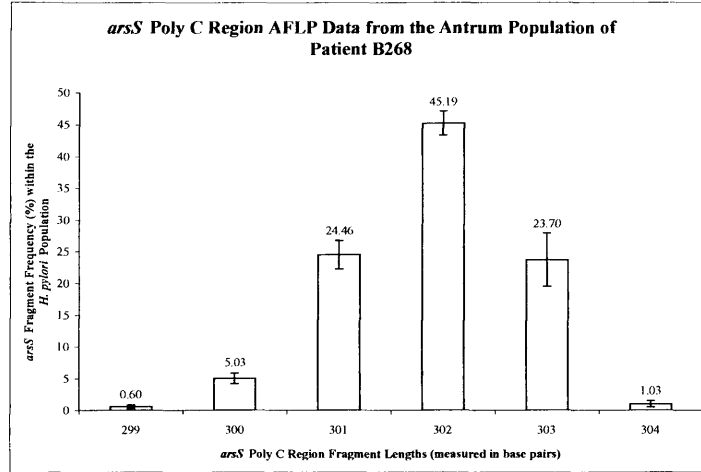
Fragment Length (bp)	Frequency	Standard Deviation
296	0.09	0.18
297	0.26	0.33
298	2.55	1.06
299	14.46	5.86
300	34.12	6.86
301	34.89	5.26
302	12.91	8.43
303	0.72	0.74



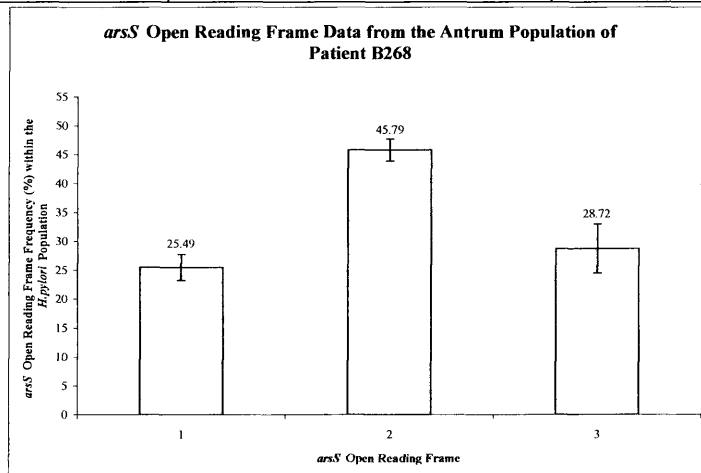
Open Reading Frame	Frequency	Standard Deviation
1	37.44	5.37
2	27.46	10.27
3	35.10	6.90

T-test	T-value	P-value
ORF1 vs. ORF2	2.58	9.98×10^{-3}
ORF2 vs. ORF3	1.85	4.12×10^{-2}
ORF1 vs. ORF3	0.80	2.17×10^{-1}

Antrum Population of Patient B268



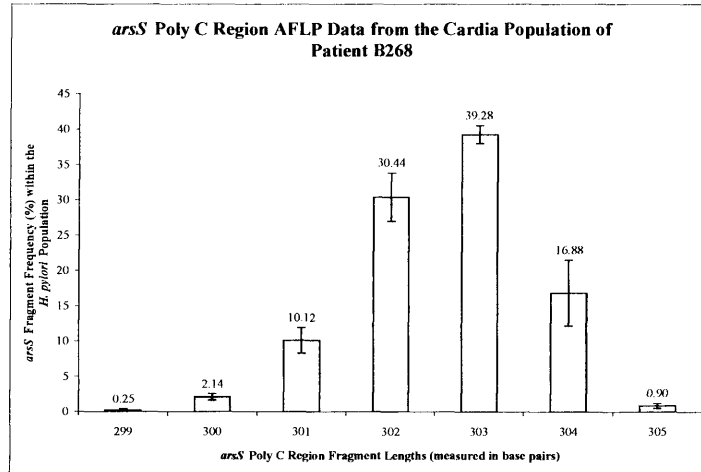
Fragment Length (bp)	Frequency	Standard Deviation
299	0.60	0.24
300	5.03	0.82
301	24.46	2.23
302	45.19	1.89
303	23.70	4.17
304	1.03	0.52



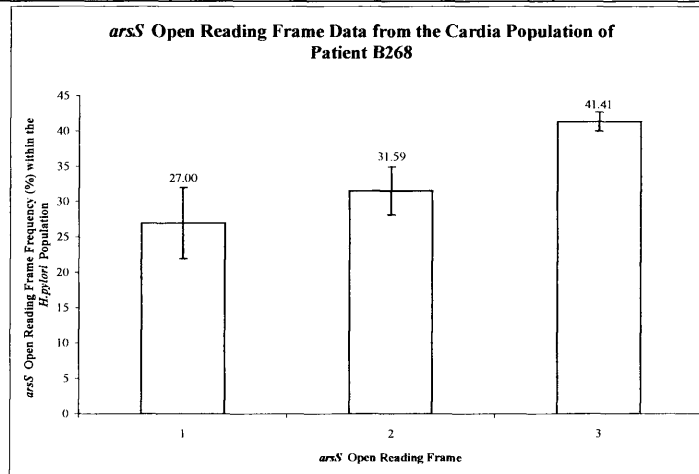
Open Reading Frame	Frequency	Standard Deviation
1	25.49	2.29
2	45.79	1.91
3	28.72	4.25

T-test	T-value	P-value
ORF1 vs. ORF2	20.44	3.41×10^{-13}
ORF2 vs. ORF3	10.98	3.67×10^{-9}
ORF1 vs. ORF3	2.01	3.09×10^{-2}

Cardia Population of Patient B268



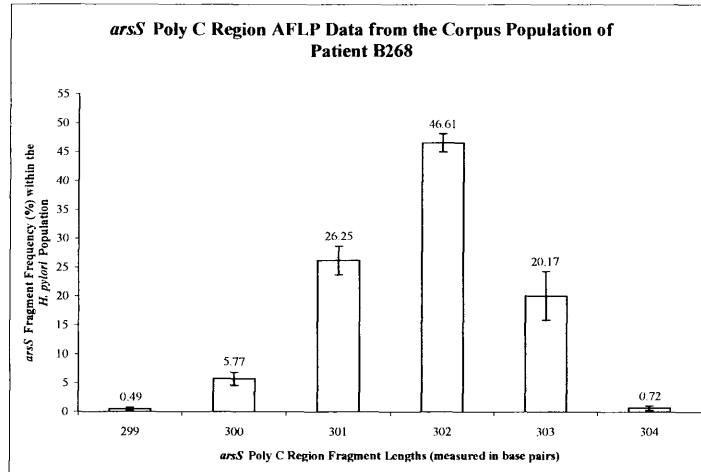
Fragment Length (bp)	Frequency	Standard Deviation
299	0.25	0.19
300	2.14	0.47
301	10.12	1.85
302	30.44	3.38
303	39.28	1.28
304	16.88	4.68
305	0.90	0.35



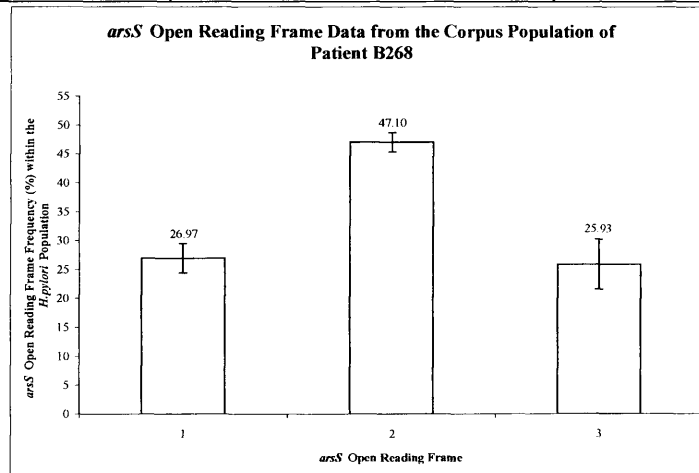
Open Reading Frame	Frequency	Standard Deviation
1	27.00	5.03
2	31.59	3.41
3	41.41	1.36

T-test	T-value	P-value
ORF1 vs. ORF2	2.27	1.88×10^{-2}
ORF2 vs. ORF3	8.03	2.66×10^{-7}
ORF1 vs. ORF3	8.30	1.71×10^{-7}

Corpus Population of Patient B268



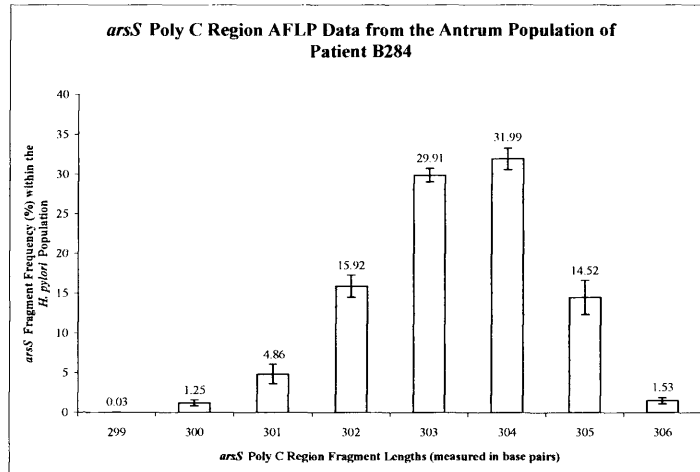
Fragment Length (bp)	Frequency	Standard Deviation
299	0.49	0.33
300	5.77	1.12
301	26.25	2.48
302	46.61	1.61
303	20.17	4.21
304	0.72	0.39



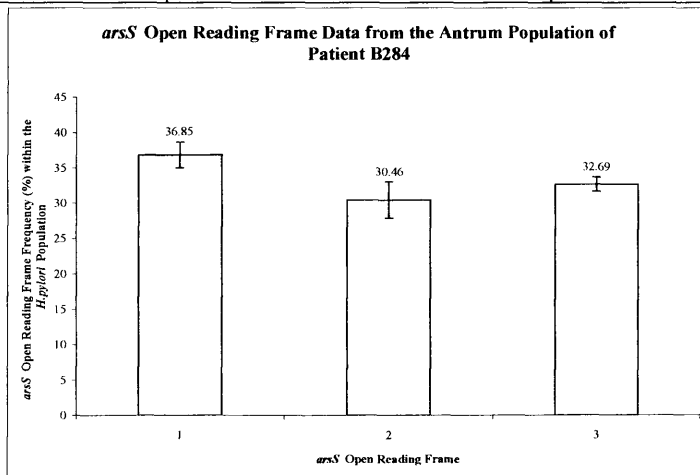
Open Reading Frame	Frequency	Standard Deviation
1	26.97	2.51
2	47.10	1.64
3	25.93	4.36

T-test	T-value	P-value
ORF1 vs. ORF2	20.14	4.30×10^{-13}
ORF2 vs. ORF3	13.63	1.59×10^{-10}
ORF1 vs. ORF3	0.62	2.73×10^{-1}

Antrum Population of Patient B284



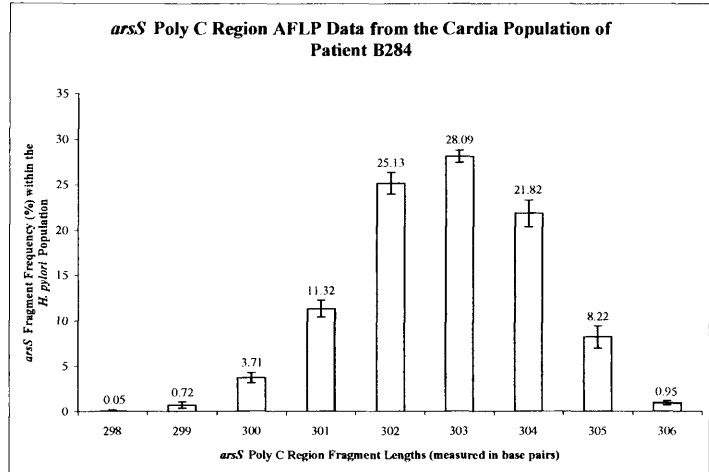
Fragment Length (bp)	Frequency	Standard Deviation
299	0.03	0.08
300	1.25	0.38
301	4.86	1.24
302	15.92	1.38
303	29.91	0.85
304	31.99	1.35
305	14.52	2.16
306	1.53	0.39



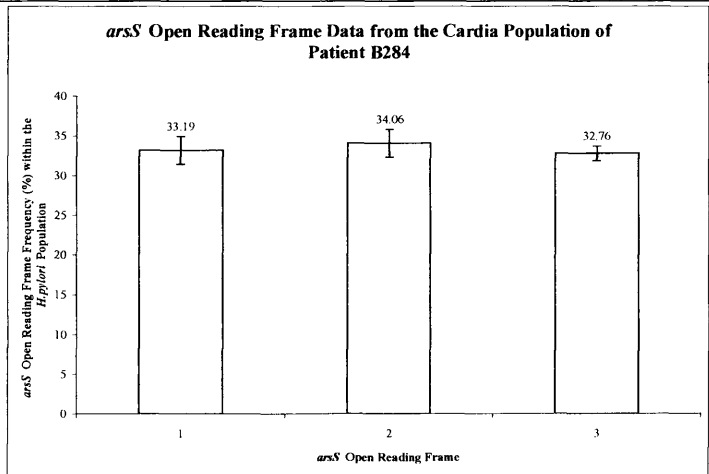
Open Reading Frame	Frequency	Standard Deviation
1	36.85	1.83
2	30.46	2.56
3	32.69	1.01

T-test	T-value	P-value
ORF1 vs. ORF2	6.09	7.89×10^{-6}
ORF2 vs. ORF3	2.42	1.38×10^{-2}
ORF1 vs. ORF3	5.99	9.48×10^{-6}

Cardia Population of Patient B284



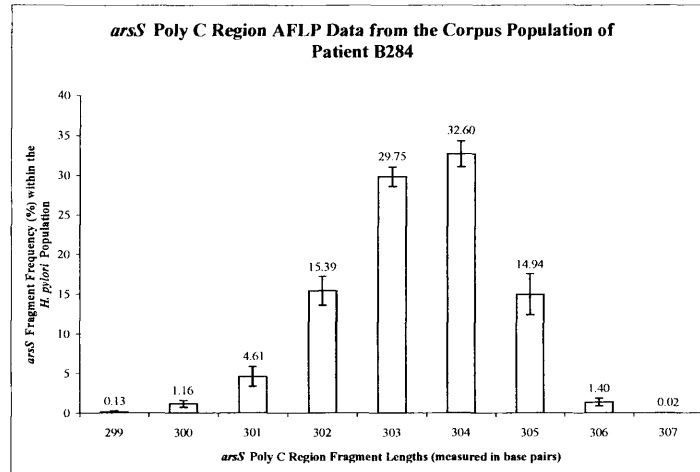
Fragment Length (bp)	Frequency	Standard Deviation
298	0.05	0.15
299	0.72	0.34
300	3.71	0.57
301	11.32	0.92
302	25.13	1.19
303	28.09	0.67
304	21.82	1.47
305	8.22	1.23
306	0.95	0.22



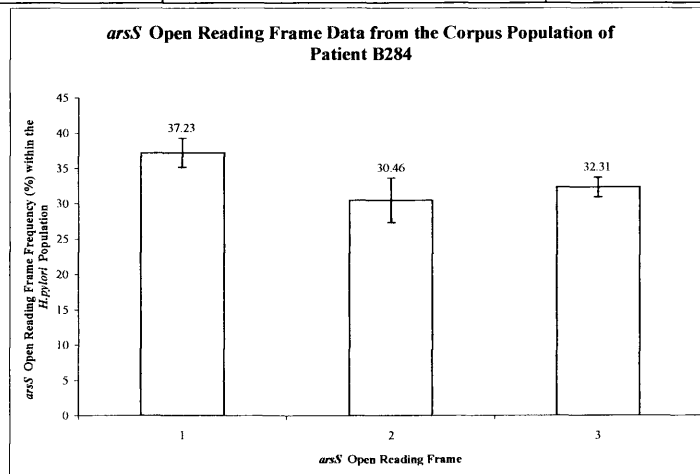
Open Reading Frame	Frequency	Standard Deviation
1	33.19	1.75
2	34.06	1.75
3	32.76	0.90

T-test	T-value	P-value
ORF1 vs. ORF2	1.06	0.15
ORF2 vs. ORF3	1.99	0.03
ORF1 vs. ORF3	0.66	0.26

Corpus Population of Patient B284



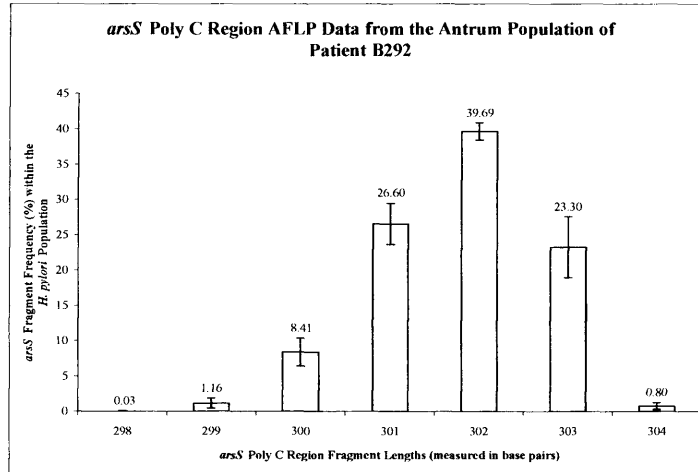
Fragment Length (bp)	Frequency	Standard Deviation
299	0.13	0.19
300	1.16	0.42
301	4.61	1.25
302	15.39	1.81
303	29.75	1.23
304	32.60	1.63
305	14.94	2.59
306	1.40	0.49
307	0.02	0.05



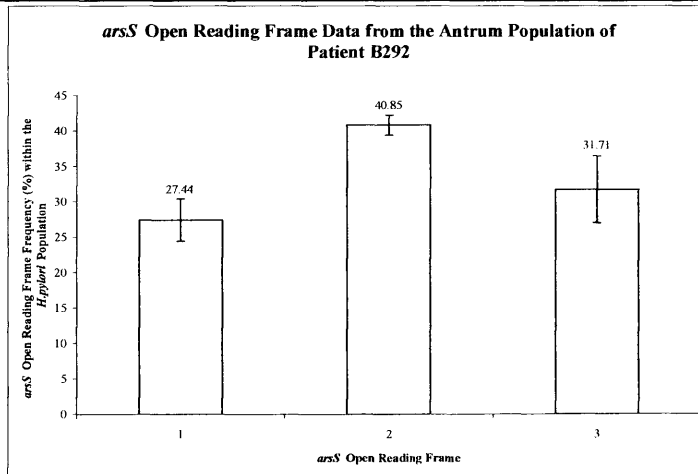
Open Reading Frame	Frequency	Standard Deviation
1	37.23	2.06
2	30.46	3.16
3	32.31	1.39

T-test	T-value	P-value
ORF1 vs. ORF2	5.38	3.07×10^{-5}
ORF2 vs. ORF3	1.61	6.37×10^{-2}
ORF1 vs. ORF3	5.94	1.04×10^{-5}

Antrum Population of Patient B292



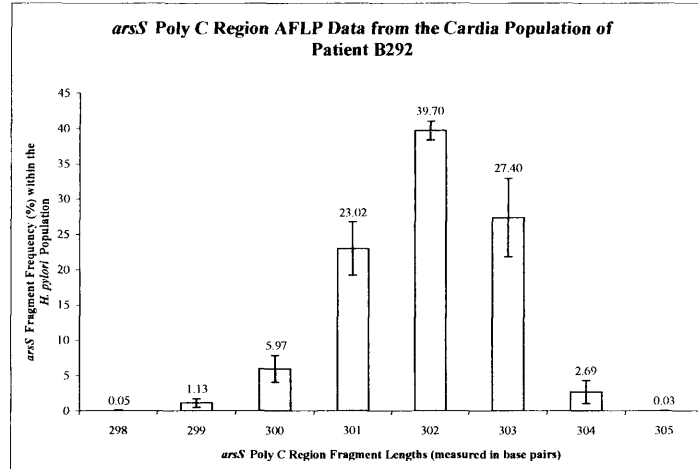
Fragment Length (bp)	Frequency	Standard Deviation
298	0.03	0.10
299	1.16	0.72
300	8.41	1.97
301	26.60	2.93
302	39.69	1.21
303	23.30	4.30
304	0.80	0.49



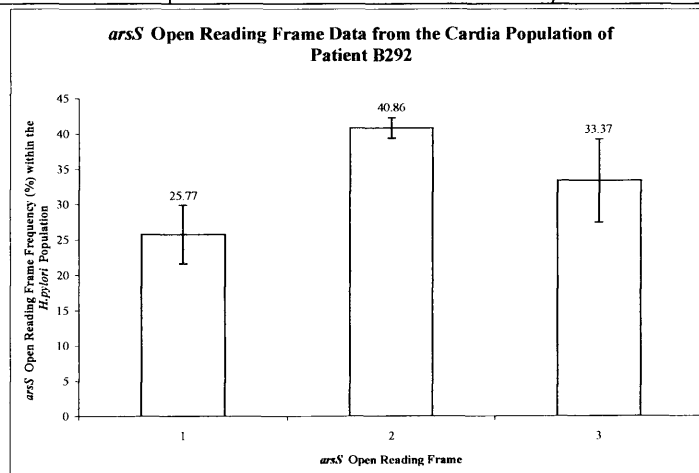
Open Reading Frame	Frequency	Standard Deviation
1	27.44	2.97
2	40.85	1.41
3	31.71	4.73

T-test	T-value	P-value
ORF1 vs. ORF2	12.22	7.88×10^{-10}
ORF2 vs. ORF3	5.56	2.16×10^{-5}
ORF1 vs. ORF3	2.29	1.78×10^{-2}

Cardia Population of Patient B292



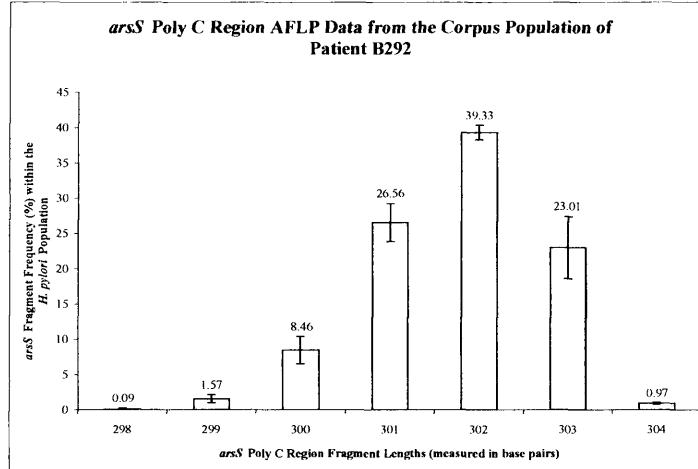
Fragment Length (bp)	Frequency	Standard Deviation
298	0.05	0.10
299	1.13	0.61
300	5.97	1.88
301	23.02	3.79
302	39.70	1.32
303	27.40	5.57
304	2.69	1.62
305	0.03	0.08



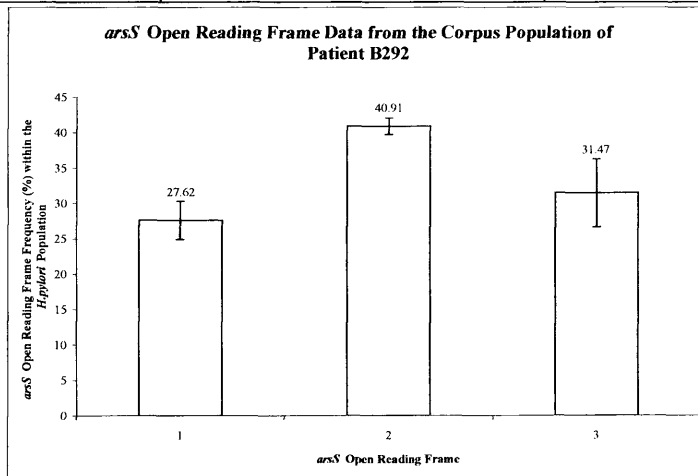
Open Reading Frame	Frequency	Standard Deviation
1	25.77	4.13
2	40.86	1.46
3	33.37	5.88

T-test	T-value	P-value
ORF1 vs. ORF2	10.35	8.52×10^{-9}
ORF2 vs. ORF3	3.71	9.55×10^{-4}
ORF1 vs. ORF3	3.17	2.94×10^{-3}

Corpus Population of Patient B292



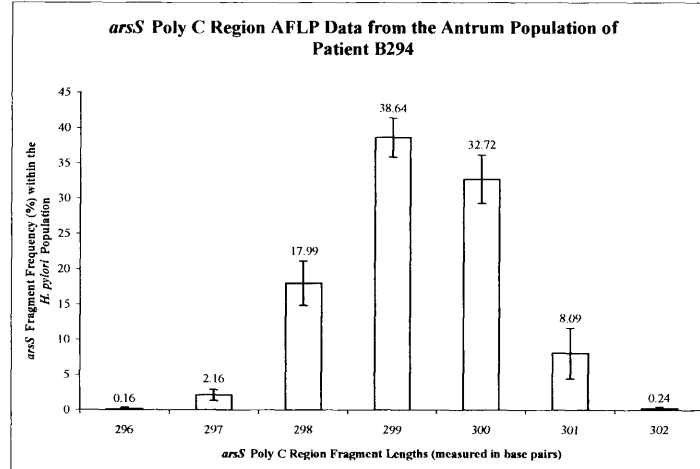
Fragment Length (bp)	Frequency	Standard Deviation
298	0.09	0.19
299	1.57	0.57
300	8.46	1.94
301	26.56	2.70
302	39.33	1.03
303	23.01	4.40
304	0.97	0.13



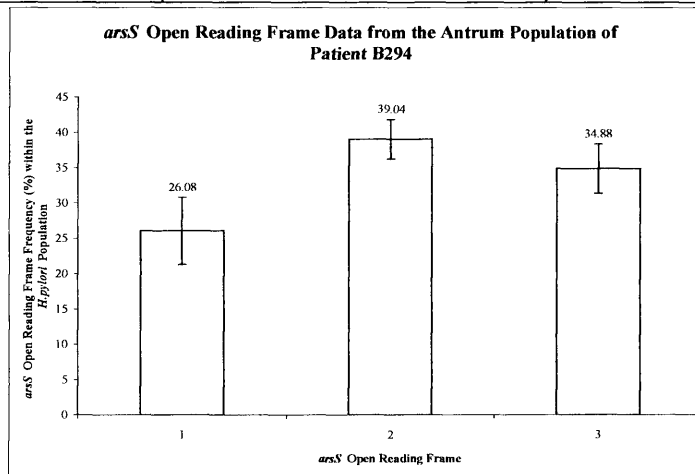
Open Reading Frame	Frequency	Standard Deviation
1	27.62	2.71
2	40.91	1.18
3	31.47	4.81

T-test	T-value	P-value
ORF1 vs. ORF2	13.50	1.84×10^{-10}
ORF2 vs. ORF3	5.72	1.59×10^{-5}
ORF1 vs. ORF3	2.09	2.63×10^{-2}

Antrum Population of Patient B294



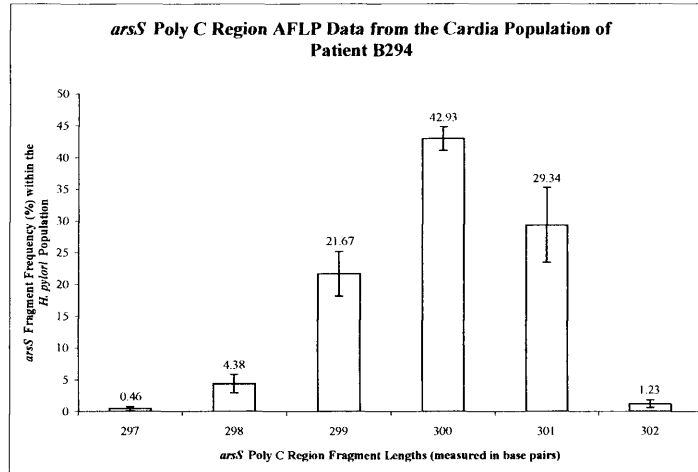
Fragment Length (bp)	Frequency	Standard Deviation
296	0.16	0.20
297	2.16	0.77
298	17.99	3.13
299	38.64	2.79
300	32.72	3.44
301	8.09	3.60
302	0.24	0.23



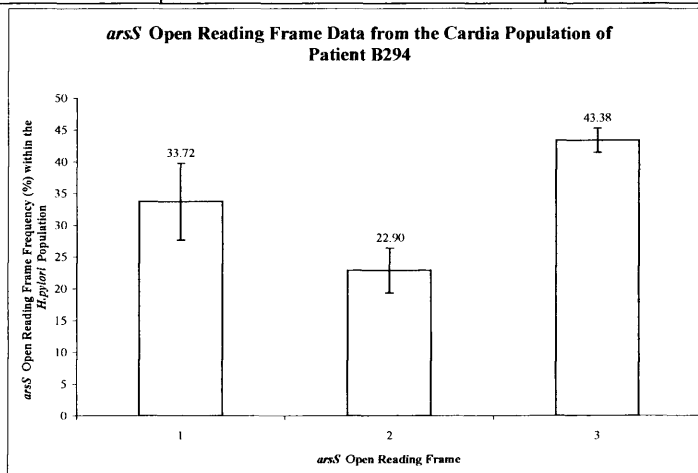
Open Reading Frame	Frequency	Standard Deviation
1	26.08	4.77
2	39.04	2.81
3	34.88	3.52

T-test	T-value	P-value
ORF1 vs. ORF2	7.02	1.43×10^{-6}
ORF2 vs. ORF3	2.78	6.74×10^{-3}
ORF1 vs. ORF3	4.45	2.02×10^{-4}

Cardia Population of Patient B294



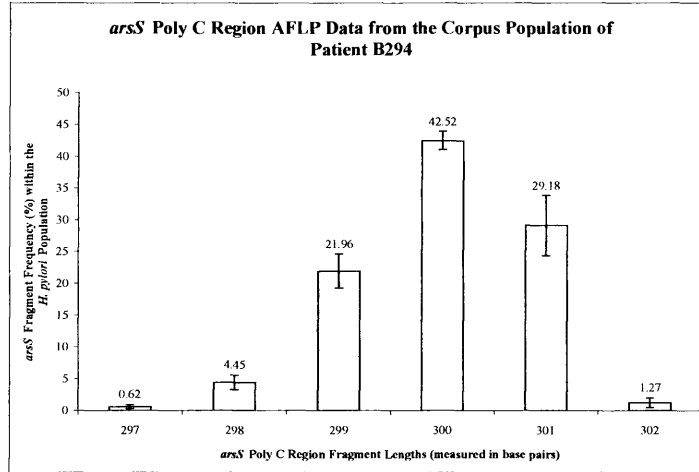
Fragment Length (bp)	Frequency	Standard Deviation
297	0.46	0.33
298	4.38	1.45
299	21.67	3.50
300	42.93	1.87
301	29.34	5.88
302	1.23	0.60



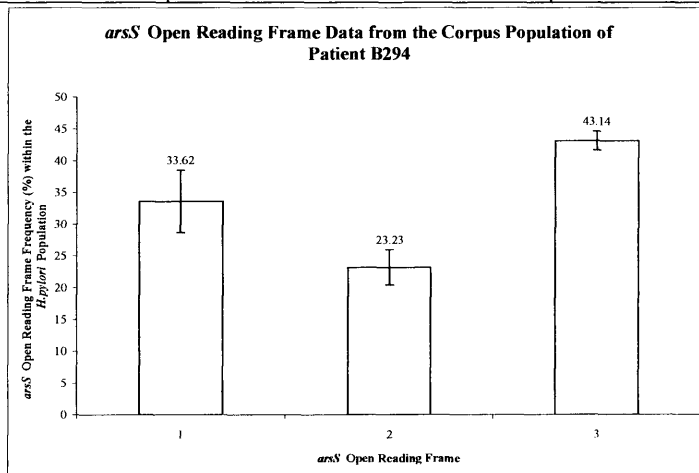
Open Reading Frame	Frequency	Standard Deviation
1	33.72	6.06
2	22.90	3.56
3	43.38	1.90

T-test	T-value	P-value
ORF1 vs. ORF2	4.63	1.40×10^{-4}
ORF2 vs. ORF3	15.25	2.97×10^{-11}
ORF1 vs. ORF3	4.56	1.59×10^{-4}

Corpus Population of Patient B294



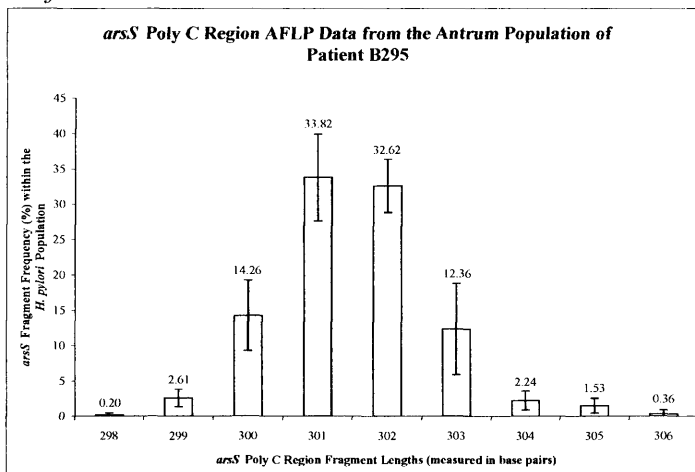
Fragment Length (bp)	Frequency	Standard Deviation
297	0.62	0.32
298	4.45	1.16
299	21.96	2.69
300	42.52	1.43
301	29.18	4.76
302	1.27	0.76



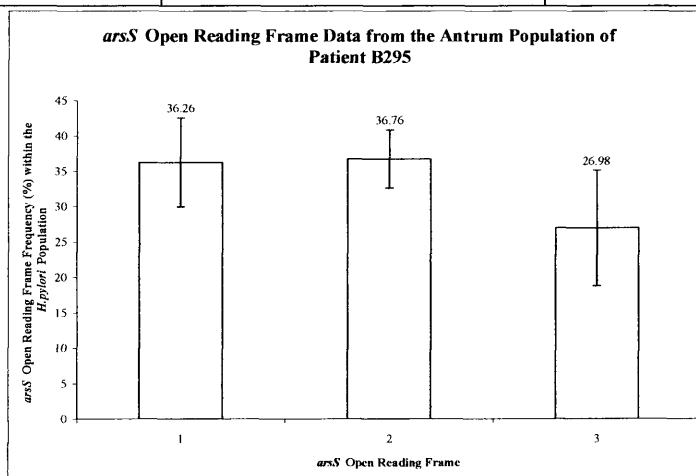
Open Reading Frame	Frequency	Standard Deviation
1	33.62	4.90
2	23.23	2.80
3	43.14	1.47

T-test	T-value	P-value
ORF1 vs. ORF2	5.53	2.29×10^{-5}
ORF2 vs. ORF3	18.90	1.14×10^{-12}
ORF1 vs. ORF3	5.59	2.05×10^{-5}

Antrum Population of Patient B295



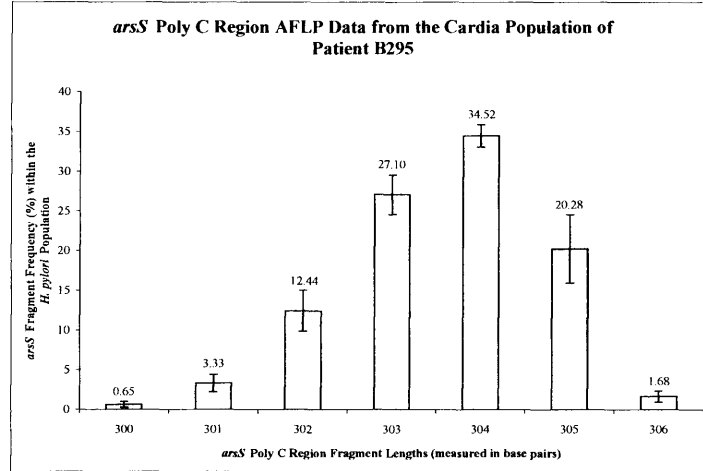
Fragment Length (bp)	Frequency	Standard Deviation
298	0.20	0.25
299	2.61	1.22
300	14.26	4.99
301	33.82	6.14
302	32.62	3.76
303	12.36	6.45
304	2.24	1.33
305	1.53	1.04
306	0.36	0.55



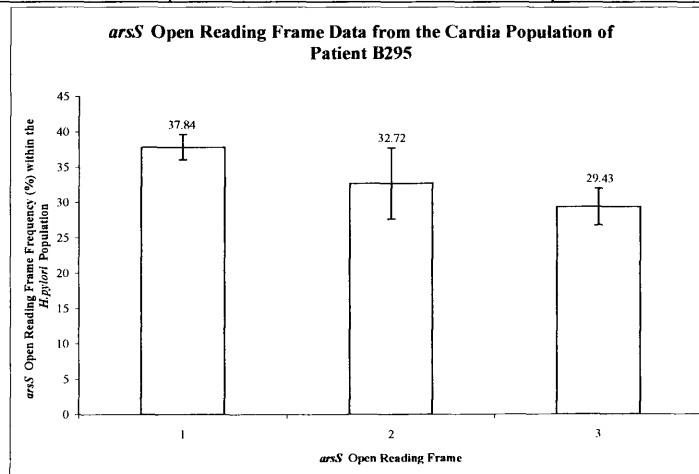
Open Reading Frame	Frequency	Standard Deviation
1	36.26	6.29
2	36.76	4.09
3	26.98	8.17

T-test	T-value	P-value
ORF1 vs. ORF2	0.20	4.22×10^{-1}
ORF2 vs. ORF3	3.21082741	2.73×10^{-3}
ORF1 vs. ORF3	2.69959826	7.89×10^{-3}

Cardia Population of Patient B295



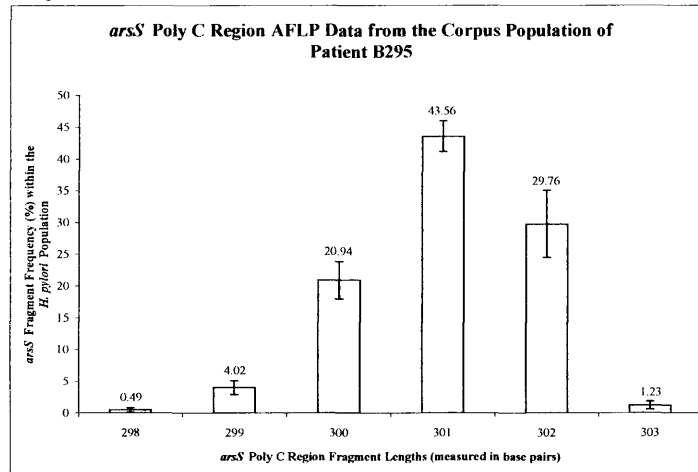
Fragment Length (bp)	Frequency	Standard Deviation
300	0.65	0.38
301	3.33	1.11
302	12.44	2.58
303	27.10	2.49
304	34.52	1.40
305	20.28	4.32
306	1.68	0.68



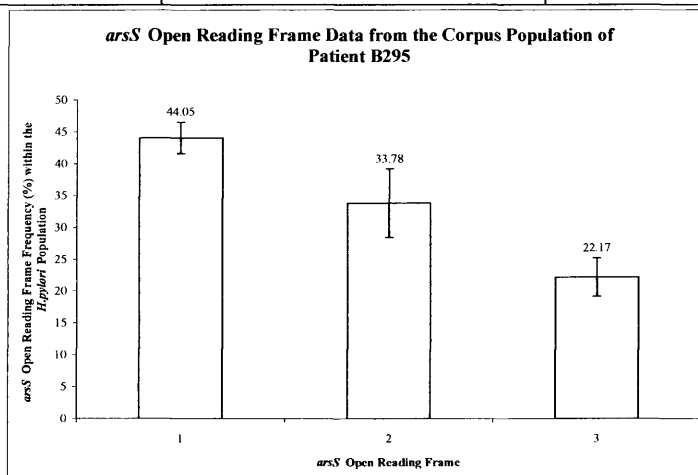
Open Reading Frame	Frequency	Standard Deviation
1	37.84	1.78
2	32.72	5.03
3	29.43	2.61

T-test	T-value	P-value
ORF1 vs. ORF2	2.88	5.46×10^{-3}
ORF2 vs. ORF3	1.74	5.05×10^{-2}
ORF1 vs. ORF3	7.98	2.88×10^{-7}

Corpus Population of Patient B295



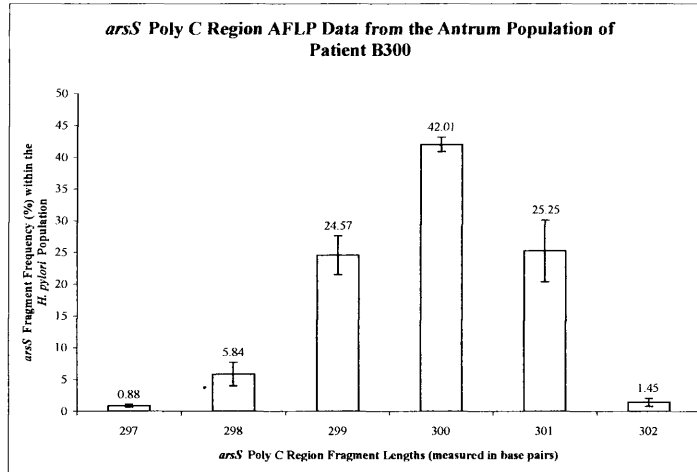
Fragment Length (bp)	Frequency	Standard Deviation
298	0.49	0.36
299	4.02	1.09
300	20.94	2.95
301	43.56	2.45
302	29.76	5.31
303	1.23	0.64



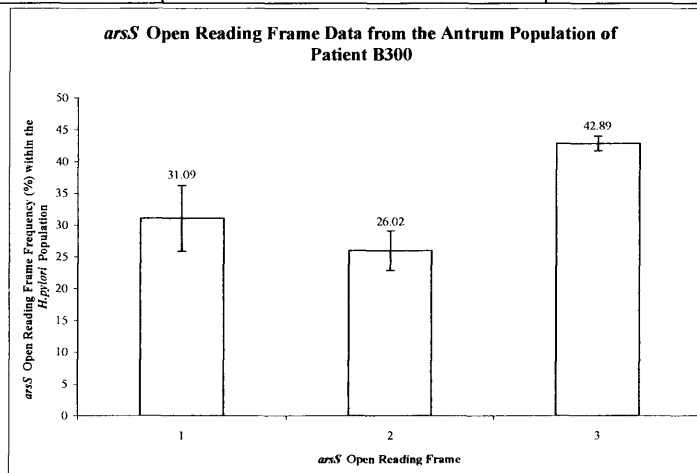
Open Reading Frame	Frequency	Standard Deviation
1	44.05	2.47
2	33.78	5.42
3	22.17	3.01

T-test	T-value	P-value
ORF1 vs. ORF2	5.17	4.61×10^{-5}
ORF2 vs. ORF3	5.62	1.93×10^{-5}
ORF1 vs. ORF3	16.84	6.68×10^{-12}

Antrum Population of Patient B300



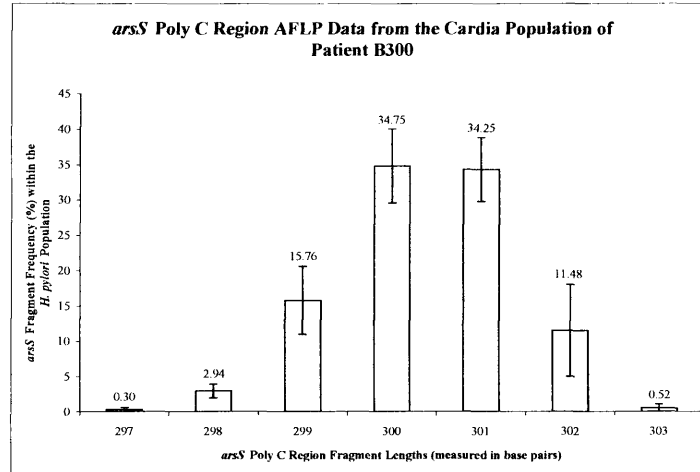
Fragment Length (bp)	Frequency	Standard Deviation
297	0.88	0.24
298	5.84	1.85
299	24.57	3.05
300	42.015	1.14
301	25.25	4.84
302	1.45	0.64



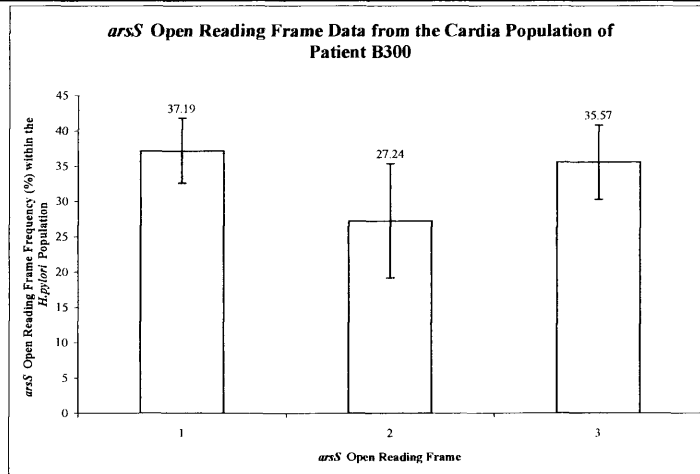
Open Reading Frame	Frequency	Standard Deviation
1	31.09	5.18
2	26.02	3.12
3	42.89	1.17

T-test	T-value	P-value
ORF1 vs. ORF2	2.51	1.15×10^{-2}
ORF2 vs. ORF3	15.20	3.18×10^{-11}
ORF1 vs. ORF3	6.66	2.73×10^{-6}

Cardia Population of Patient B300



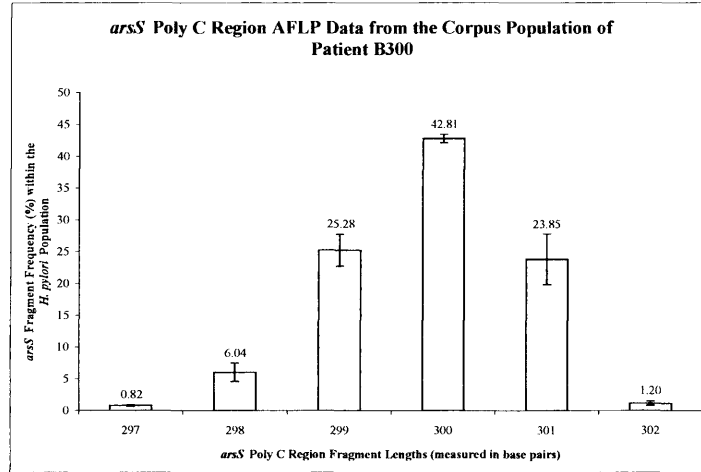
Fragment Length (bp)	Frequency	Standard Deviation
297	0.30	0.30
298	2.94	0.95
299	15.76	4.80
300	34.75	5.23
301	34.25	4.52
302	11.48	6.50
303	0.52	0.53



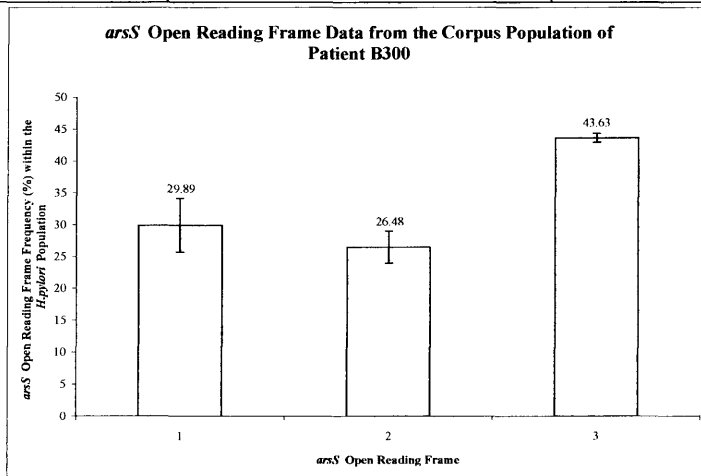
Open Reading Frame	Frequency	Standard Deviation
1	37.19	4.61
2	27.24	8.08
3	35.57	5.27

T-test	T-value	P-value
ORF1 vs. ORF2	3.21	2.74×10^{-3}
ORF2 vs. ORF3	2.59	9.87×10^{-3}
ORF1 vs. ORF3	0.70	2.48×10^{-1}

Corpus Population of Patient B300

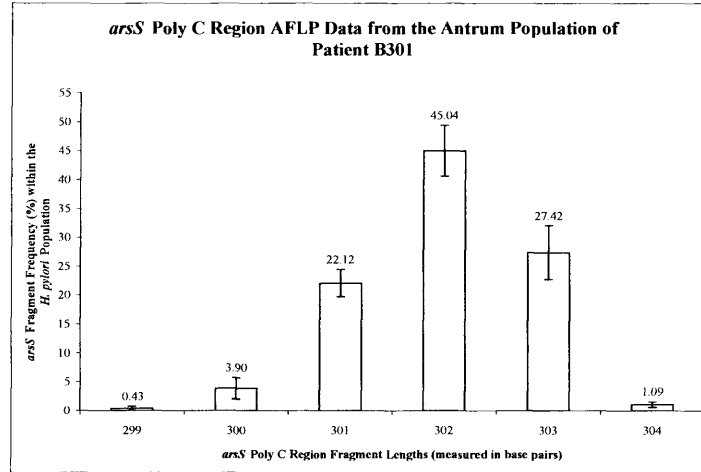


Fragment Length (bp)	Frequency	Standard Deviation
297	0.82	0.12
298	6.04	1.45
299	25.28	2.51
300	42.81	0.70
301	23.85	3.97
302	1.20	0.33

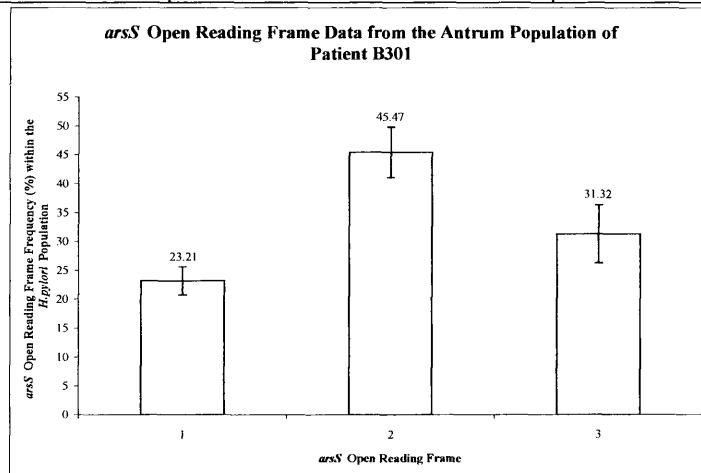


Open Reading Frame	Frequency	Standard Deviation
1	29.89	4.23
2	26.48	2.53
3	43.63	0.72
T-test	T-value	P-value
ORF1 vs. ORF2	2.07	2.75×10^{-2}
ORF2 vs. ORF3	19.57	6.69×10^{-13}
ORF1 vs. ORF3	9.61	2.37×10^{-8}

Antrum Population of Patient B301



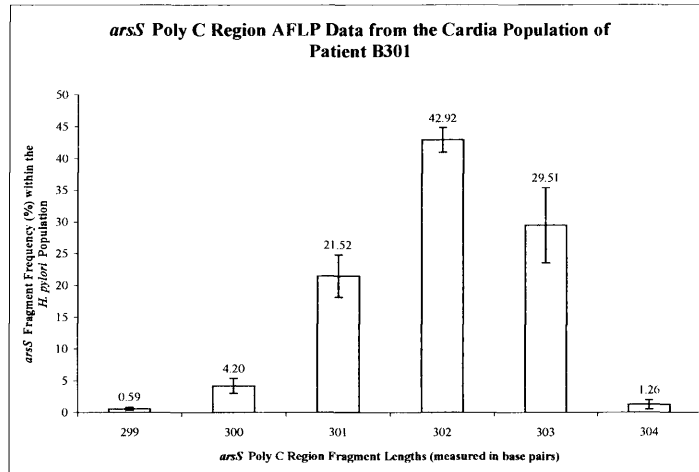
Fragment Length (bp)	Frequency	Standard Deviation
299	0.43	0.36
300	3.90	1.84
301	22.12	2.37
302	45.04	4.37
303	27.42	4.65
304	1.09	0.47



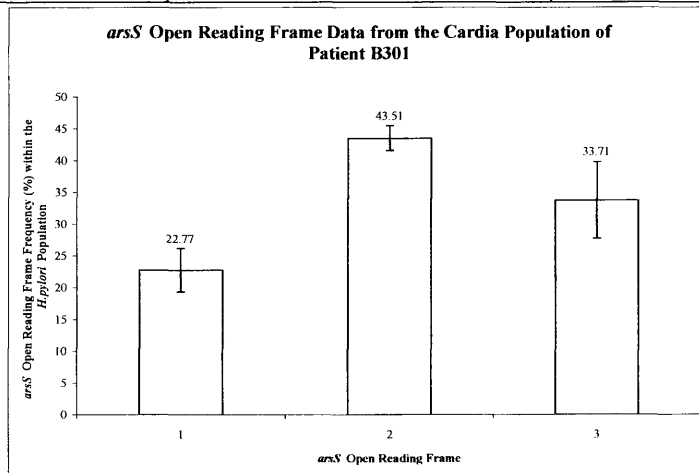
Open Reading Frame	Frequency	Standard Deviation
1	23.21	2.42
2	45.47	4.39
3	31.32	5.00

T-test	T-value	P-value
ORF1 vs. ORF2	13.33	2.20×10^{-10}
ORF2 vs. ORF3	6.38	4.59×10^{-6}
ORF1 vs. ORF3	4.38	2.32×10^{-4}

Cardia Population of Patient B301



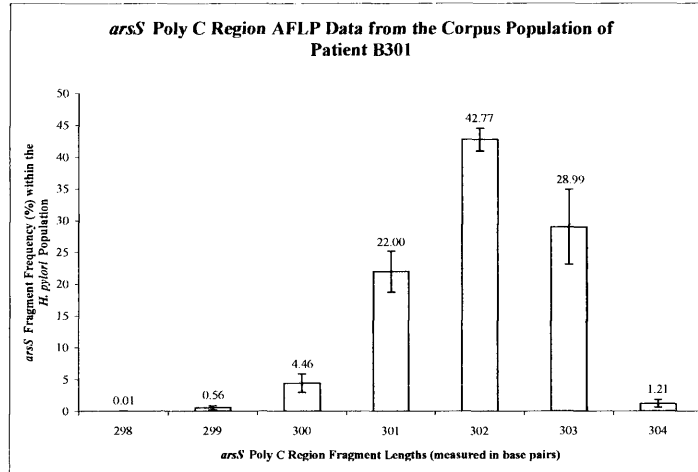
Fragment Length (bp)	Frequency	Standard Deviation
299	0.59	0.21
300	4.20	1.18
301	21.52	3.34
302	42.92	1.94
303	29.51	5.93
304	1.26	0.70



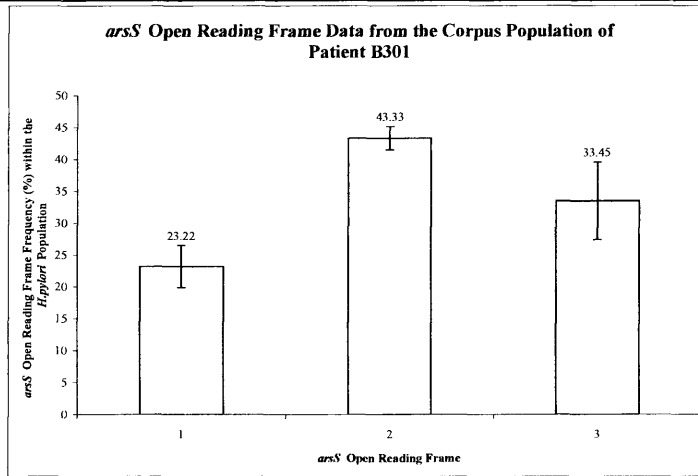
Open Reading Frame	Frequency	Standard Deviation
1	22.77	3.41
2	43.51	1.95
3	33.71	6.05

T-test	T-value	P-value
ORF1 vs. ORF2	15.83	1.71×10^{-11}
ORF2 vs. ORF3	4.63	1.40×10^{-4}
ORF1 vs. ORF3	4.72	1.15×10^{-4}

Corpus Population of Patient B301

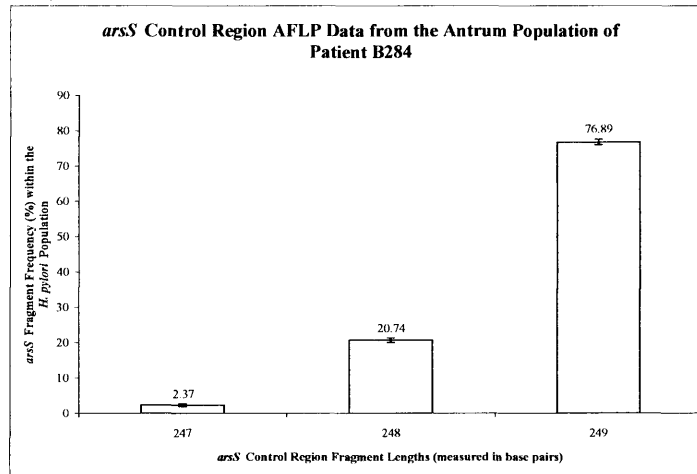


Fragment Length (bp)	Frequency	Standard Deviation
298	0.01	0.04
299	0.56	0.34
300	4.46	1.47
301	22.00	3.24
302	42.77	1.80
303	28.99	5.89
304	1.21	0.60



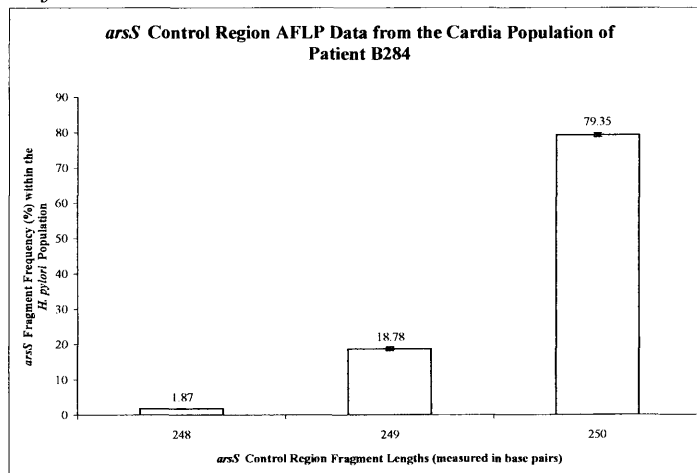
Open Reading Frame	Frequency	Standard Deviation
1	23.22	3.30
2	43.33	1.83
3	33.45	6.07
T-test	T-value	P-value
ORF1 vs. ORF2	16.01	1.43×10^{-11}
ORF2 vs. ORF3	4.68	1.25×10^{-4}
ORF1 vs. ORF3	4.45	2.04×10^{-4}

Appendix 2B: AFLP Data from Control Regions of *H. pylori* Populations
Antrum Population of Patient B284



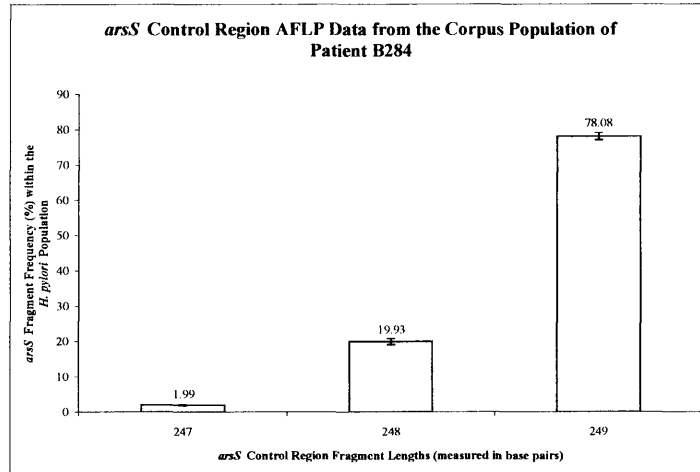
Fragment Length (bp)	Frequency	Standard Deviation
247	2.37	0.36
248	20.74	0.62
249	76.89	0.83

Cardia Population of Patient B284



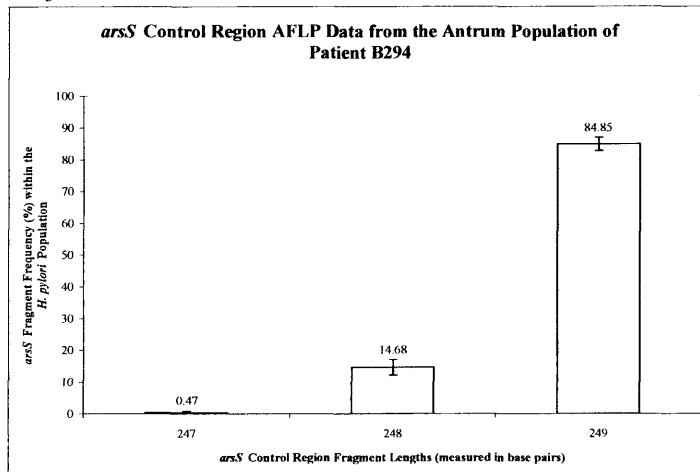
Fragment Length (bp)	Frequency	Standard Deviation
248	1.87	0.11
249	18.78	0.52
250	79.35	0.55

Corpus Population of Patient B284



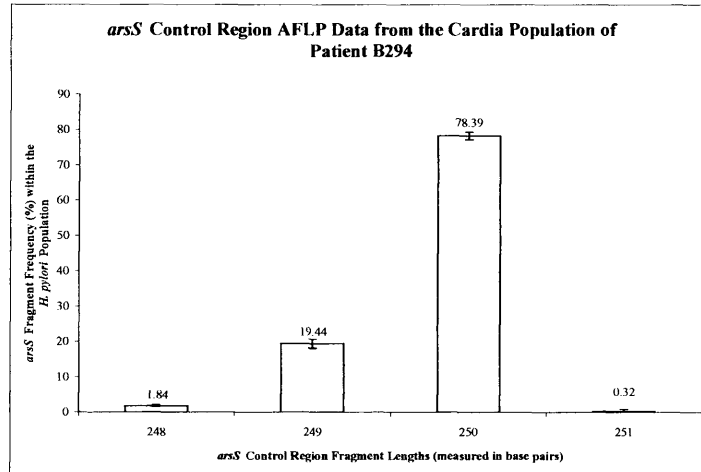
Fragment Length (bp)	Frequency	Standard Deviation
247	1.99	0.16
248	19.9316866	0.86
249	78.08	1.02

Antrum Population of Patient B294



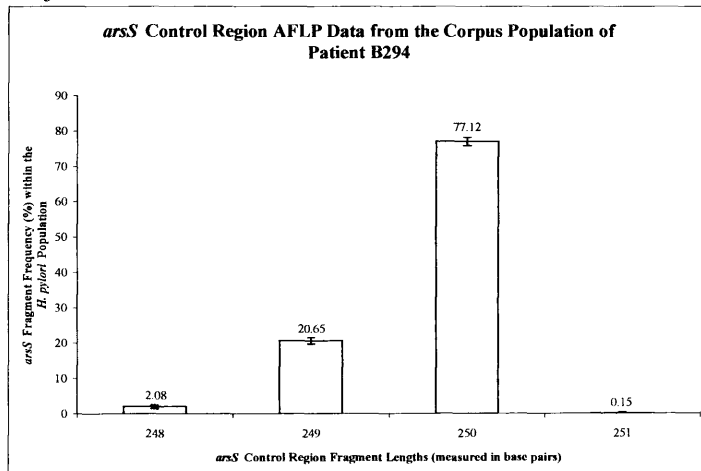
Fragment Length (bp)	Frequency	Standard Deviation
247	0.47	0.42
248	14.68	2.47
249	84.85	2.06

Cardia Population of Patient B294



Fragment Length (bp)	Frequency	Standard Deviation
248	1.84	0.26
249	19.44	1.30
250	78.39	1.10
251	0.32	0.56

Corpus Population of Patient B294



Fragment Length (bp)	Frequency	Standard Deviation
248	2.08	0.53
249	20.65	0.91
250	77.12	1.17
251	0.15	0.27

References

1. **Achtman, M., T. Azuma, D. E. Berg, Y. Ito, G. Morelli, Z. Pan, S. Suerbaum, S. A. Thompson, A. van der Ende, and L. van Doorn.** 1999. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol. Microbiol.* **32**:459 – 470.
2. **Akada, J. K., M. Shirai, H. Takeuchi, M. Tsuda, and T. Nakazawa.** 2000. Identification of the urease operon in *Helicobacter pylori* and its control by mRNA decay in response to pH. *Mol. Microbiol.* **36**:1071 – 1084.
3. **Alm, R. A., L. S. Ling, D. T. Moir, B. L. King, E. D. Brown, P. C. Doig, D. R. Smith, B. Noonan, B. C. Guild, B. L. deJonge, G. Carmel, P. J. Tummino, A. Caruso, M. Uria-Nickelsen, D. M. Mills, C. Ives, R. Gibson, D. Merberg, S. D. Mills, Q. Jiang, D. E. Taylor, G. F. Vovis, and T. J. Trust.** 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **397**:176 – 180.
4. **Baltus, D. A., M. R. Amieva, A. Covacci, T. M. Lowe, D. S. Merrell, K. M. Ottemann, M. Stein, N. R. Salama, and K. Guillemin.** 2009. The Complete Genome Sequence of *Helicobacter pylori* Strain G27. *J. Bacteriol.* **191**:447 – 448.
5. **Beier, D., and R. Frank.** 2000. Molecular characterization of two-component systems of *Helicobacter pylori*. *J. Bacteriol.* **182**:2068 – 2076.
6. **Benoit, S. L., N. Mehta, M. V. Weinberg, C. Maier, and R. J. Maier.** 2007. Interaction between the *Helicobacter pylori* accessory proteins HypA and UreE is needed for urease maturation. *Microbiology.* **153**:1474 – 1482.
7. **Björkholm, B., M. Sjölund, P. G. Falk, O. G. Berg, L. Engstrand, and D. I. Andersson.** 2001. Mutation frequency and biological cost of antibiotic resistance in *Helicobacter pylori*. *P. Natl. Acad. Sci. USA.* **98**:14607 – 14612.
8. **Blaser, M. J., Y. Chen, and J. Reibman.** 2008. Does *Helicobacter pylori* protect against asthma and allergy? *Gut.* **57**:561 – 567.
9. **Cases, I., V. de Lorenzo, and C. A. Ouzounis.** 2003. Transcription regulation and environmental adaptation in bacteria. *TRENDS Microbiol.* **11**:248 – 253.
10. **Censini, S. C. Lange, Z. Xiang, J. E. Crabtree, P. Ghiara, M. Borodovsky, R. Rappuoli, and A. Covacci.** 1996. *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *P. Natl. Acad. Sci. USA.* **93**:14648 – 14653.
11. **Contreras, M., J. Thiberge, M. Mandrand-Berthelot, and A. Labigne.** 2003. Characterization of the roles of NikR, a nickel-responsive pleiotropic autoregulator of *Helicobacter pylori*. *Mol. Microbiol.* **49**:947 – 963.

12. **Covacci, A., J. L. Telford, G. Del Giudice, J. Parsonnet, and R. Rappuoli.** 1999. *Helicobacter pylori* Virulence and Genetic Geography. *Science*. **284**:1328 – 1333.
13. **Dailidiene, D., G. Dailide, D. Kersulyte, and D. E. Berg.** 2006. Contraselectable Streptomycin Susceptibility Determinant for Genetic Manipulation and Analysis of *Helicobacter pylori*. *Appl. Environ. Microb.* **72**:5908 – 5914.
14. **Dietz, P., G. Gerlach, and D. Beier.** 2002. Identification of Target Genes Regulated by the Two-Component System HP166-HP165 of *Helicobacter pylori*. *J. Bacteriol.* **184**:350 – 362.
15. **Eppinger, M., C. Baar, G. Raddatz, D. H. Huson, and S. C. Schuster.** 2004. Comparative Analysis of Four Campylobacterales. *Nat. Rev. Microbiol.* **2**:872 – 885.
16. **Eppinger, M., C. Baar, B. Linz, G. Raddatz, C. Lanz, H. Keller, G. Morelli, H. Gressmann, M. Achtman, and S. C. Schuster.** 2006. Who Ate Whom? Adaptive *Helicobacter* Genomic Changes That Accompanied a Host Jump from Early Humans to Large Felines. *PLOS Genet.* **2**:1097 – 1110.
17. **Fabret, C., V. A. Feher, and J. A. Hoch.** 1999. Two-Component Signal Transduction in *Bacillus subtilis*: How One Organism Sees Its World. *J. Bacteriol.* **181**:1975 – 1983.
18. **Falush, D., T. Wirth, B. Linz, J. K. Pritchard, M. Stephens, M. Kidd, M. J. Blaser, D. Y. Graham, S. Vacher, G. I. Perez-Perez, Y. Yamaoka, F. Mégraid, K. Otto, U. Reichard, E. Katzowitsch, X. Wang, M. Achtman, and S. Suerbaum.** 2003. Traces of Human Migrations in *Helicobacter pylori* Populations. *Science*. **299**:1582 – 1585.
19. **Fischer, W., L. Windhager, A. Karnholz, M. Zeiller, R. Zimmer, and R. Haas.** 2008. The complete genome sequence of *Helicobacter pylori* strain P12. Unpublished.
20. **Forsyth, M. H., P. Cao, P. P. Garcia, J. D. Hall, and T. L. Cover.** 2002. Genome-wide transcriptional profiling in a histidine kinase mutant of *Helicobacter pylori* identifies members of a regulon. *J. Bacteriol.* **184**:4630–4635.
21. **Gibson, J. R., E. Slater, J. Xerry, D. S. Tompkins, and R. J. Owen.** 1998. Use of an Amplified-Fragment Length Polymorphism Technique to Fingerprint and Differentiate Isolates of *Helicobacter pylori*. *J. Clin. Microbiol.* **36**:2580 – 2585.

22. **Goodwin, A. C., D. M. Weinberger, C. B. Ford, J. C. Nelson, J. D. Snider, J. D. Hall, C. I. Paules, R. M. Peek Jr., and M. H. Forsyth.** 2008. Expression of the *Helicobacter pylori* adhesin SabA is controlled via phase variation and the ArsRS signal transduction system. *Microbiology*. **154**:2231 – 2240.
23. **Guo, X., and J. Mrázek.** 2008. Long Simple Sequence Repeats in Host-Adapted Pathogens Localize Near Genes Encoding Antigens, Housekeeping Genes, and Pseudogenes. *J. Mol. Evol.* **67**:497 – 509.
24. **Han, S., H. E. Zschausch, H. W. Meyer, T. Schneider, M. Loos, S. Bhakdi, and M. J. Maeurer.** 2000. *Helicobacter pylori*: Clonal Population Structure and Restricted Transmission within Families Revealed by Molecular Typing. *J. Clin. Microbiol.* **38**: 3646 – 3651.
25. **Henderson, I. R., P. Owen, and J. P. Nataro.** 1999. Molecular switches - the ON and OFF of bacterial phase variation. *Mol. Microbiol.* **33**:919 – 932.
26. **Kersulyte, D., A. Kalia, R. H. Gilman, and D. E. Berg.** 2010. Genome Sequence of *Helicobacter pylori* from the Remote Amazon: traces of Asian ancestry of the First Americans. Unpublished.
27. **Kim, S., W. K. Lee, S. H. Choi, S. Kang, H. S. Park, Y. S. Kim, S. G. Lee, E. Y. Byun, J. E. Jeong, Y. H. Park, E. J. Lee, J. S. Kim, B. D. Ryu, Y. S. Lee, Y. Hahn, Y. I. Yeom, S. G. Park, H. S. Youn, G. H. Ko, M. B. Choi, C. H. Park, J. Y. Lim, D. W. Bae, J. Y. Song, J. U. Park, H. L. Kang, S. C. Baik, M. J. Cho, H. S. Yoo, and K. H. Rhee.** 2004. *Helicobacter pylori* 51. <http://www.ncbi.nlm.nih.gov/nuccore/CP000012.1>.
28. **Kim, S., W. K. Lee, S. H. Choi, S. Kang, H. S. Park, Y. S. Kim, S. G. Lee, E. Y. Byun, J. E. Jeong, Y. H. Park, E. J. Lee, J. S. Kim, B. D. Ryu, Y. S. Lee, Y. Hahn, Y. I. Yeom, S. G. Park, H. S. Youn, G. H. Ko, M. B. Choi, C. H. Park, J. Y. Lim, D. W. Bae, J. Y. Song, J. U. Park, H. L. Kang, S. C. Baik, M. J. Cho, H. S. Yoo, and K. H. Rhee.** 2009. *Helicobacter pylori* 52. <http://www.ncbi.nlm.nih.gov/nuccore/261838873>.
29. **Krakovka, S., K. A. Eaton, and D. M. Rings.** 1995. Occurrence of Gastric Ulcers in Gnotobiotic Piglets Colonized by *Helicobacter pylori*. *Infect. Immun.* **63**:2352 – 2355.
30. **Linz, B., and S. C. Schuster.** 2007. Genomic diversity in *Helicobacter* and related organisms. *Res. Microbiol.* **158**:737 – 744.
31. **Loh, J. T., and T. L. Cover.** 2006. Requirement of Histidine Kinases HP0165 and HP1364 for Acid Resistance in *Helicobacter pylori*. *Infect. Immun.* **74**:3052 – 3059.

32. **Loh, J. T., S. S. Gupta, D. B. Friedman, A. M. Krezel, and T. L Cover.** 2010. Analysis of Protein Expression Regulated by the *Helicobacter pylori* ArsRS Two-Component Signal Transduction System. *J. Bacteriol.* **192**:2034 – 2043.
33. **Marcus, E. A., A. P. Moshfegh, G. Sachs, and D. R. Scott.** 2005. The Periplasmic α -Carbonic Anhydrase Activity of *Helicobacter pylori* Is Essential for Acid Acclimation. *J. Bacteriol.* **187**:729 – 738.
34. **Marshall, B.** 2006. Helicobacter Connections. *ChemMedChem.* **1**:783 – 802.
35. **McNulty, S. L., B. M. Mole, D. Dailidienne, I. Segal, R. Ally, R. Mistry, O. Secka, R. A. Adegbola, J. E. Thomas, E. M. Lenarcic, R. M. Peek Jr, D. E. Berg, and M. H. Forsyth.** 2004. Novel 180- and 480-Base-Pair Insertions in African and African-American Strains of *Helicobacter pylori*. *J. Clin. Microbiol.* **42**:5658 – 5663.
36. **Merrell, D. S., and A. Camilli.** 2002. Acid tolerance of gastrointestinal pathogens. *Curr. Opin. Microbiol.* **5**:51 – 55.
37. **Merrell, D. S., M. L. Goodrich, G. Otto, L. S. Tompkins, and S. Falkow.** 2003. pH-Regulated Gene Expression of the Gastric Pathogen *Helicobacter pylori*. *71*:3529 – 3539.
38. **Mollenhauer-Rektoschek, M., G. Hanauer, G. Sachs, and K. Melchers.** 2002. Expression of UreI is required for intragastric transit and colonization of gerbil gastric mucosa by *Helicobacter pylori*. *Res. Microbiol.* **153**:659 – 666.
39. **Müller, S., J. Förster, and D. Beier.** 2005. Repeated sequence motifs in the *Helicobacter pylori* P₁₄₀₈ promoter do not affect its transcription. *Microbiol. Res.* **161**:212 – 221.
40. **Oh, J. D., H. Kling-Bäckhed, M. Giannakis, J. Xu, R. S. Fulton, L. A. Fulton, H. S. Cordum, C. Wang, G. Elliott, J. Edwards, E. R. Mardis, L. G. Engstrand, and J. I. Gordon.** 2006. The complete genome sequence of a chronic atrophic gastritis *Helicobacter pylori* strain: Evolution during disease progression. *P. Natl. Acad. Sci. USA.* **103**:9999 – 10004.
41. **Orsi, R., B. M. Bowen, and M. Wiedmann.** 2010. Homopolymeric tracts represent a general regulatory mechanism in prokaryotes. *BMC Genomics.* **11**:1471 – 2164.
42. **Pantheil, K., P. Dietz, R. Haas, and D. Beier.** 2003. Two-Component Systems of *Helicobacter pylori* Contribute to Virulence in a Mouse Infection Model. *Infect. Immun.* **71**:5381 – 5385.

43. **Parsonnet, J., D. Vandersteen, J. Goates, R. K. Sibley, J. Pritikin, and Y. Chang.** 1991. *Helicobacter pylori* Infection in Intestinal- and Diffuse-Type Gastric Adenocarcinomas. *J. Natl. Cancer I.* **83**:640 – 643.
44. **Pattison, C. P., M. J. Combs, and B. J. Marshall.** 1996. *Helicobacter pylori* and Peptic Ulcer Disease: Evolution to Revolution to Resolution. *Am. J. Roentgenol.* **168**:1415 – 1420.
45. **Pflock, M., P. Dietz, J. Schar, and D. Beier.** 2004. Genetic evidence for histidine kinase HP165 being an acid sensor of *Helicobacter pylori*. *FEMS Microbiol. Lett.* **234**:51-61.
46. **Pflock, M., S. Kennard, I. Delany, V. Scarlato, and D. Beier.** 2005. Acid-induced activation of the urease promoters is mediated directly by the ArsRS two-component system of *Helicobacter pylori*. *Infect. Immun.* **73**:6437 – 6445.
47. **Pflock, M., N. Finsterer, B. Joseph, H. Mollenkopf, T. F. Meyer, and D. Beier.** 2006. Characterization of the ArsRS regulon of *Helicobacter pylori*, Involved in Acid Adaptation. *J. Bacteriol.* **188**:3449 –3462.
48. **Pflock, M., S. Kennard, N. Finsterer, and D. Beier.** 2006. Acid-responsive gene regulation in the human pathogen *Helicobacter pylori*. *J. Biotechnol.* **126**:52 – 60.
49. **Pflock, M., S. Müller, and D. Beier.** 2006. The CrdRS (HP1365-HP1364) Two-Component System Is Not Involved in pH-Responsive Gene Regulation in the *Helicobacter pylori* Strains 26695 and G27. *Curr. Microbiol.* **54**:320 – 324.
50. **Sachs, G., D. L. Weeks, Y. Wen, E. A. Marcus, and D. R. Scott.** 2005. Acid Acclimation by *Helicobacter pylori*. *Physiology.* **20**:429 – 438.
51. **Sanabria-Valentín, E., M. C. Colbert, and M. J. Blaser.** 2007. Role of *futC* slipped strand mispairing in *Helicobacter pylori* Lewis^y phase variation. *Microbes Infect.* **9**:1553 – 1560.
52. **Scarlato, V., I. Delany, G. Spohn, and D. Beier.** 2001. Regulation of transcription in *Helicobacter pylori*: simple systems or complex circuits? *Int. J. Med. Microbiol.* **291**:107 – 117.
53. **Schär, J., A. Sickmann, and D. Beier.** 2005. Phosphorylation-Independent Activity of Atypical Response Regulators of *Helicobacter pylori*. *J. Bacteriol.* **187**:3100 – 3109.

54. **Scott, D. R., E. A. Marcus, D. L. Weeks, A. Lee, K. Melchers, and G. Sachs.** 2000. Expression of the *Helicobacter pylori ureI* Gene Is Required for Acidic pH Activation of Cytoplasmic Urease. *Infect. Immun.* **68**:470 – 477.
55. **Scott, D. R., E. A. Marcus, D. L. Weeks, and G. Sachs.** 2002. Mechanisms of acid resistance due to the urease system of *Helicobacter pylori*. *Gastroenterology.* **123**:187 – 195.
56. **Scott, D. R., E. A. Marcus, Y. Wen, J. Oh, and G. Sachs.** 2007. Gene Expression *in vivo* Shows that *Helicobacter pylori* Colonizes an Acidic Niche on the Gastric Surface. *P. Natl. Acad. Sci. USA.* **104**:7235 – 7240.
57. **Scott, D. R., E. A. Marcus, Y. Wen, S. Singh, J. Feng, and G. Sachs.** 2010. Cytoplasmic Histidine Kinase (HP0244)-Regulated Assembly of Urease with UreI, a Channel for Urea and Its Metabolites, CO₂, NH₃, and NH₄⁺ Is Necessary for Acid Survival of *Helicobacter pylori*. *J. Bacteriol.* **192**:94 – 103.
58. **Snyder, L., and W. Champness.** 2007. Phase variation in *Salmonella* species, p. 415-416. *In* Molecular Genetics of Bacteria, 3rd. ASM Press, Washington, DC.
59. **Snyder, L. A. S., N. J. Loman, J. D. Linton, R. R. Langdon, G. M. Weinstock, B. W. Wren, and M. J. Pallen.** 2010. Simple sequence repeats in *Helicobacter canadensis* and their role in phase variable expression and C-terminal sequence switching. *BMC Genet.* **11**:67.
60. **Solnick, J. V., and D. B. Schauer.** 2001. Emergence of Diverse *Helicobacter* Species in the Pathogenesis of Gastric and Enterohepatic Diseases. *Clin. Microbiol. Rev.* **14**:59 – 97.
61. **Stock, A. M., V. L. Robinson, and P. N. Goudreau.** 2000. Two-Component Signal Transduction. *Annu. Rev. Biochem.* **69**:183 – 215.
62. **Stray-Pedersen, A., A. Vege, and T. O. Rognum.** 2008. *Helicobacter pylori* Antigen in Stool Is Associated With SIDS and Sudden Infant Deaths due to Infectious Disease. *Pediatr. Res.* **64**:405 – 410.
63. **Suerbaum, S., J. M. Smith, K. Bapumia, G. Morelli, N. H. Smith, E. Kunstmann, I. Dyrek, and M. Achtman.** 1998. Free recombination within *Helicobacter pylori*. *P. Natl. Acad. Sci. USA.* **95**:12619 – 12624.
64. **Thiberge, J., C. Boursaux-Eude, P. Lehours, M. Dillies, S. Creno, J. Coppee, Z. Rouy, A. Lajus, L. Ma, C. Burucoa, A. Ruskone-Foumestroux, A. Courillon-Mallet, H. De Reuse, I. G. Boneca, D. Lamarque, F. Megraud, J. Delchier, C. Medigue, C. Bouchier, A. Labigne, and J. Raymond.** 2010. Array-based hybridization of *Helicobacter pylori* isolates to the complete genome

sequence of an isolate associated with MALT lymphoma. *BMC Genomics*. **11**:368.

65. **Throup, J. P., K. K. Koretke, A. P. Bryant, K. A. Ingraham, A. F. Chalker, Y. Ge, A. Marra, N. G. Wallis, J. R. Brown, D. J. Holmes, M. Rosenberg, and M. K. R. Burnham.** 2000. A genomic analysis of two-component signal transduction in *Streptococcus pneumoniae*. *Mol. Microbiol.* **35**:566 – 576.
66. **Tomb, J. F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzgerald, N. Lee, M. D. Adams, E. K. Hickey, D. E. Berg, J. D. Gocayne, T. R. Utterback, J. D. Peterson, J. M. Kelley, M. D. Cotton, J. M. Weidman, C. Fujii, C. Bowman, L. Watthey, E. Wallin, W. S. Hayes, M. Borodovsky, P. D. Karp, H. O. Smith, C. M. Fraser, and J. C. Venter.** 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**:539-547.
67. **Uemura, N., S. Okamoto, S. Yamamoto, N. Matsumura, S. Yamaguchi, M. Yamakido, K. Taniyama, N. Sasaki, and R. J. Schlemper.** 2001. *Helicobacter pylori* Infection and the Development of Gastric Cancer. *N. Engl. J. Med.* **345**:784 – 789.
68. **Waidner, B., K. Melchers, F. N. Stähler, M. Kist, and S. Bereswill.** 2005. The *Helicobacter pylori* CrdRS Two-Component Regulation System (HP1364/HP1365) Is Required for Copper-Mediated Induction of the Copper Resistance Determinant CrdA. *J. Bacteriol.* **187**:4683 – 4688.
69. **Wang, G., M. Z. Humayun, and D. E. Taylor.** 1999. Mutation as an origin of genetic variability in *Helicobacter pylori*. *TRENDS Microbiol.* **7**:488 – 493.
70. **Wen, Y., E. A. Marcus, U. Matrubutham, M. A. Gleeson, D. R. Scott, and G. Sachs.** 2003. Acid-Adaptive Genes of *Helicobacter pylori*. *Infect. Immun.* **71**:5921 – 5939.
71. **Wen, Y., J. Feng, D. R. Scott, E. A. Marcus, and G. Sachs.** 2006. Involvement of the HP0165-HP0166 two-component system in expression of some acidic-pH-upregulated genes of *Helicobacter pylori*. *J. Bacteriol.* **188**:1750 – 1761.
72. **Wen, Y., J. Feng, D. R. Scott, E. A. Marcus, and G. Sachs.** 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 – 2434.
73. **West, A. H., and A. M. Stock.** 2001. Histidine kinases and response regulator proteins in two-component signaling systems. *TRENDS Biochem. Sci.* **26**:369 – 376.

74. **Yamamoto, K., K. Hirao, T. Oshima, H. Aiba, R. Utsumi, and A. Ishihama.** 2004. Functional Characterization *in Vitro* of All Two-component Signal Transduction Systems from *Escherichia coli*. *J. Biol. Chem.* **280**:1448 – 1456.
75. **Yamaoka, Y., O. Ojo, S. Fujimoto, S. Odenbreit, R. Haas, O. Gutierrez, H. M. T. El-Zimaity, R. Reddy, A. Arnqvist, and D. Y. Graham.** 2006. *Helicobacter pylori* outer membrane proteins and gastroduodenal disease. *Gut.* **55**:775 – 781.