

12-2010

Role of eukaryotic Sel-1 like repeat containing genes in *Helicobacter pylori* evolution and pathogenesis.

Kalyani Putty
University of Louisville

Follow this and additional works at: <https://ir.library.louisville.edu/etd>



Part of the [Biology Commons](#)

Recommended Citation

Putty, Kalyani, "Role of eukaryotic Sel-1 like repeat containing genes in *Helicobacter pylori* evolution and pathogenesis." (2010).
Electronic Theses and Dissertations. Paper 1169.
<https://doi.org/10.18297/etd/1169>

This Doctoral Dissertation is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact thinkir@louisville.edu.

**ROLE OF EUKARYOTIC SEL-1 LIKE REPEAT CONTAINING GENES IN
HELICOBACTER PYLORI EVOLUTION AND PATHOGENESIS**

By

Kalyani Putty, B.V.Sc & A.H.

Acharya N.G.Ranga Agricultural University, 2005

A Dissertation

**Submitted to the Faculty of the Graduate School of the University of Louisville in
Partial Fulfillment of the Requirements for the Degree of**

Doctor of Philosophy

**Department of Biology, Division of Molecular, Cellular and Developmental Biology,
Program on Disease Evolution**

University of Louisville

Louisville, Kentucky

December 2010

**ROLE OF EUKARYOTIC SEL-1 LIKE REPEAT CONTAINING GENES IN
HELICOBACTER PYLORI EVOLUTION AND PATHOGENESIS**

By

Kalyani Putty, B.V.Sc & A.H.

Acharya N.G.Ranga Agricultural University, 2005

Dissertation Approved on

November 29, 2010

By the Following Dissertation Committee

Dr. Awdhesh Kalia, Ph.D., Assoc. Professor, Dept. of Biology, University of Louisville.

Dr. Michael H. Perlin, Ph.D., Professor, Dept. of Biology, University of Louisville.

Dr. Micah Worley, Ph.D., Asst. Professor, Dept. of Biology, University of Louisville.

**Dr. Douglas E. Berg, Ph.D., Alumni Professor,
Dept. of Molecular Microbiology, Washington Univeristy in St. Louis.**

**Dr. Yousef Abu Kwaik, Ph.D., Bumgardner Chair of Molecular Pathogenesis,
Dept. of Micribiology and Immunology, University of Louisville**

DEDICATION

To Ajay – my light at the end of the tunnel!

ACKNOWLEDGEMENTS

Firstly, I express my heartfelt gratitude to my advisor Dr. Awdhesh Kalia for his great instructions, patience and diligent guidance. It has been a long road and his continuous support is greatly appreciated. I would like to extend a special thanks to my committee members, Drs. Douglas Berg, Yousef Abu-Kwaik, Michael Perlin and Micah Worley for their suggestions, and wonderful cooperation. I thank the present and past lab mates for lending an excellent helping hand when needed, especially Sarah Marcus for her tremendous contributions to the HcpC study. My deepest thanks to Dr. Palaniappan Sethu, and Dr. Rosendo Estrada, for help with FACS assays. Special thanks to Dr. Christopher Price (Abu-Kwaik Laboratory) and Dr. Michael Perlin for help with Real time PCR experiments. I also thank Dr. Swathi Arur, UT MD Anderson Cancer Center and Dr. Douglas Berg, Washington University in St. Louis for kindly providing several key reagents/tools for my study. A very special thanks to Dr. Micah Worley and his lab members (Josh Thornbrough and Tom Hundley) for sharing their cell culture supplies; my dear friend, Jinny Paul for encouragement, criticism and advice. I thank the Dept. of Biology for giving me the opportunity and financial support to achieve my goals. Last but not the least, none of this would have been possible without the excellent support and strength provided by my family. Very special thanks to my wonderful husband, Dr. Phani Kumar Patibandla for his unending cooperation and for giving me the strength to fulfill my dreams.

ABSTRACT

ROLE OF EUKARYOTIC SEL-1 LIKE REPEAT CONTAINING GENES IN *HELICOBACTER PYLORI* EVOLUTION AND PATHOGENESIS

By

Kalyani Putty, B.V.Sc & A.H

November 29. 2010

Background: *Helicobacter pylori* (*Hp*) establishes life-long gastric infection in billions of humans, and is often responsible for diseases such as peptic ulcer and gastric cancer. Cumulative actions of *genetic drift* and *natural selection* over several millennia sculpted the present *Hp* population structure, which is characterized by extreme genetic diversity and striking geographic clustering of genotypes. Natural selection is more commonly imprinted in DNA sequences of *Hp* proteins that interact with host components; however, in most instances biological relevance of selection during *Hp* infection remains unknown. Here, I attempted to elucidate the consequence of natural selection in two different contexts: (1) on the preservation of duplicated genes in *Hp* genome; and (2) lineage-specific adaptive evolution in *Hp* virulence protein HepC.

Principle Findings: I characterized the molecular evolutionary dynamics of paralogs, *hcpC* and *hcpG*, which belong to the *Hp* Sel1-like gene family. *hcpG* genomic analyses identified three distinct states in natural *Hp* populations, whereby *hcpG* was either

deleted, pseudogenized or encoded highly polymorphic alleles. In contrast, full-length *hcpC* alleles were conserved in all genomes. Although positive selection was detected in the phylogenies of *hcpG* and *hcpC* indicating that both genes had evolved under pressure to diversify, the intensity of selection was much stronger on *hcpG* than *hcpC*. The contribution of *hcpC* to *Hp* fitness, in the AGS cell culture infection model, was significantly greater than *hcpG*; however, both genes together demonstrated an additive effect on *Hp* fitness during infection (24 hrs p.i.: $S_{\Delta hcpC} = 0.264$ vs. $S_{\Delta hcpG} = 0.074$, $P < 0.01$; $S_{\Delta hcpC}$ or $S_{\Delta hcpG}$ vs. $S_{\Delta hcpC::\Delta hcpG} = 0.431$, $P < 0.01$, where S =coefficient of median fitness reduction). Furthermore, HcpC was necessary and sufficient for optimal surface expression of Heat-Shock Protein B (HspB), a major contributor to *Hp* virulence, specifically during infection, and functionally compensated for the lack of HcpG. In contrast, HcpG was only required for optimal HspB expression during early infection, and was unable to compensate for the lack of HcpC during later phases. Thus, a stable, genetically redundant, epistatic and overlapping yet non-reciprocal functional relationship emerged between *hcpC* and *hcpG*: *natural selection* favored retention of the ancestral *hcpC* function and sub-functionalization by fixation of loss-of-function mutations in *hcpG* following its origin in the *Hp* genome.

Earlier studies from my lab showed that the HepC protein, which also belongs to *Hp* Sell1-like gene family, interacted specifically with human cytoskeletal protein Ezrin, and that lineage-specific positive selection changed HepC-Ezrin interaction affinity. How might alterations in HepC-Ezrin interaction affect progression of *Hp* infection? As a first step I established that HepC was indeed biologically relevant and contributed significantly to *Hp* fitness during infection ($P=0.05$). Furthermore, PCR-array analyses

suggested that identical molecules of the human cytoskeletal pathway were differentially up or down-regulated by genetically diverse *Hp* isolates during infection, and that HepC likely inhibited key components of the human cytoskeletal machinery during infection. Thus HepC (possibly via its interaction with Ezrin) likely contributes a key regulatory role that might determine the pace and trajectory of *Hp* infection.

Conclusion: Collectively, my thesis proposes a novel mechanism through which *natural selection* favors the emergence of a stable state of genetic redundancy among duplicated genes, *hcpC* and *hcpG* that contribute significantly to *Hp* infection. This work also establishes a framework within which to further clarify the role of lineage-specific selection in fine-tuning *Hp*-host interactions.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iv
ABSTRACT.....	v
LIST OF FIGURES.....	xi
LIST OF TABLES.....	xiii
INTRODUCTION.....	1
I. Historical aspects of <i>Helicobacter pylori</i>	1
II. Epidemiology and clinical outcome of <i>H. pylori</i> infection.....	1
III. Genetic diversity in <i>H. pylori</i>	2
IV. Population genetic structure of <i>H. pylori</i> : role of positive selection.....	4
V. Eukaryotic like Sel1-like repeat (<i>slr</i>) containing gene family in <i>H.</i> <i>pylori</i>	6
VI. Expansion of <i>H. pylori slr</i> gene family by gene duplication.....	8
VII. Host cytoskeletal dysregulation following <i>H. pylori</i> infection.....	10
SPECIFIC AIMS.....	13
MATERIALS AND METHODS.....	15
I. <i>Helicobacter pylori</i> culture and maintenance.....	15
II. DNA Extraction, PCR-conditions, and DNA sequencing.....	16

III. Computational Biological Analyses.....	19
IV. Genetic Engineering of <i>H. pylori</i>	23
V. Cell Culture and Antibodies.....	34
VI. Immublotting.....	36
VII. Growth Kinetics and Fitness Assays.....	37
VIII. Reverse-transcription PCRs.....	40
IX. Flourescence Activated Cell Sorting Analyses.....	44
RESULTS.....	49
Evolution of Stable Genetic, and Functionally Non-Reciprocal Redundancy Driven by Positive Selection in Duplicated Sel-1 like Genes of <i>H. pylori</i>	49
<i>H. pylori</i> strain – specific <i>slr</i> genes.....	49
Genetic rearrangement at <i>hcpC</i> locus is unique to <i>H. pylori</i> strain 26695.....	52
Strain specific distribution of <i>H. pylori slr gene, hcpG</i>	53
Unique DNA sequence polymorphisms and pseudogenization of <i>hcpG</i>	53
Variations in the number and distribution of Sel-1 domains in uninterrupted <i>hcpG</i> ORFs change the tertiary structure of HcpG.....	58
HcpG is rapidly evolving in diverse <i>H. pylori</i> isolates.....	59
Biological relevance of HcpG in <i>H. pylori</i> growth and AGS cell infection.....	60
Non-neutral evolutionary dynamics of <i>hcpC</i>	64
Biological significance of HcpC adaptive evolution.....	68
Growth kinetics of <i>hcpC</i> single mutant and <i>hcpC-hcpG</i> double mutant.....	68

HcpC and HcpG paralogs are redundant and contribute additively to relative fitness of <i>H. pylori</i> strain G27MA in AGS cell infection	69
Role of HcpC and HcpG in surface translocation of HspB	72
HcpC and HcpG dependent modulation of HspB surface expression requires the cellular infection.....	79
Summary-I.....	80
Role of HcpC during <i>H. pylori</i> Growth and Infection.....	81
Differential regulation of <i>hepC</i> expression in diverse <i>H. pylori</i> isolates in AGS cell infection model.....	81
HepC contributes significantly to the fitness of <i>H. pylori</i>	82
HepC likely targets the host cytoskeletal machinery during late infection in AGS cell culture model of infection.....	84
Genetically diverse <i>H. pylori</i> strains differentially dysregulate cytoskeletal regulators during an early infection of AGS cell line infection.....	87
Summary-II.....	91
DISCUSSION AND FUTURE DIRECTIONS.....	92
REFERENCES.....	102
APPENDIX.....	115
CURRICULUM VITAE.....	119

LIST OF FIGURES

Figure 1: Multiple episodes of positive selection in <i>H. pylori</i> <i>slr</i> gene family expansion.....	9
Figure 2: Construction of insertion and deletion alleles by assembly of three fragments with overlapping ends.....	25
Figure 3: Strategy for knocking out <i>hcpC</i> homolog in <i>H. pylori</i> strain 26695 StrR..	28
Figure 4: Strategy for knocking out <i>hcpG</i> homolog in <i>H. pylori</i> strain G27MA.....	31
Figure 5: SLR domain architecture in encoded proteins of <i>H. pylori</i> <i>slr</i> genes, <i>hcpC</i> and <i>hcpG</i> in the available <i>H. pylori</i> genomes.....	51
Figure 6: Genetic organization of <i>hcpC</i> in the available <i>H. pylori</i> genomes.....	52
Figure 7: <i>hcpG</i> rapidly evolves in diverse <i>H. pylori</i> isolates	55
Figure 8: <i>Indels</i> pattern seen among HcpG homologs in diverse <i>H. pylori</i> isolates..	57
Figure 9: HcpG variants differ in their tertiary structures.....	58
Figure 10: Biological relevance of HcpG during <i>H. pylori</i> growth and infection.....	63
Figure 11: Non neutral evolutionary dynamics of <i>hcpC</i>	66
Figure 12: Growth and relative fitness dynamics of G27MA – <i>hcp</i> mutants.....	70
Figure 13: Role of G27MA HcpC and G27MA HcpG in <i>H. pylori</i> surface translocation of HspB.....	74
Figure 14: Dynamics of CagA and MAPKYT in G27MA – AGS infection.....	76

Figure 15: Synthesis of HspB is not affected in G27MA – <i>hcp</i> mutants.....	78
Figure 16: Cellular infection dependent surface translocation defect of HspB in G27MA – <i>hcp</i> mutants.....	80
Figure 17: <i>hepC</i> appears to be relevant biologically, and required during <i>H. pylori</i> infection.....	83
Figure 18: HepC likely targets the host cytoskeletal machinery during late infection in AGS cell culture model of infection.....	86
Figure 19: Differential dysregulation of cytoskeletal regulators following <i>H. pylori</i> infection with diverse <i>H. pylori</i> strains.....	90
Figure 20: Evolution of stable, non-reciprocal genetic redundancy by diversifying selection following duplication and divergence in <i>H. pylori</i> Sel 1-like gene family.....	94
Figure 21: Causes and consequences of <i>H. pylori</i> molecular evolution.....	100

LIST OF TABLES

Table 1: Parameters for FACS analysis.....	47
Table 2: Primers used in the study.....	48
Table 3: Strain specific distribution of <i>hcpG</i> in the available <i>H. pylori</i> genomes.....	50
Table 4: Maximum-likelihood parameters of selection pressures acting on <i>H. pylori</i> <i>hcpG</i> codons.....	60
Table 5: Maximum-likelihood parameters of selection pressures acting on <i>H. pylori</i> <i>hcpC</i> codons.....	67

INTRODUCTION

I. Historical aspects of *Helicobacter pylori*

H. pylori is a gram-negative, helical, flagellated, microaerophilic bacterium; it was first identified and isolated from human stomach by B. J. Marshall and R. Warren in 1982 [159, 160], for which they were awarded a Nobel prize in Medicine, in 2005. The bacterium belongs to class *Epsilon - proteobacteria*, family *Helicobacteraceae*, and before being grouped as genus *Helicobacter*, it was classified as genus *Campylobacter* [161]. Studies have shown that humans have been colonized by *H. pylori* for at least 60,000 years, and that as ancient humans migrated out of Central Africa and inhabited different geographic regions, the bacterium had coevolved with its human host [1, 2].

II. Epidemiology and clinical outcome of *H. pylori* infection

H. pylori inhabits the gastric mucosa of more than half of world's population, usually colonizing human stomachs in childhood and persisting throughout the life, thus suggesting effective management and perhaps even exploitation of host responses [3]. Transmission usually occurs locally within families or within small populations through oro-faecal route [4, 5]. In developing countries, the prevalence can be as high as 80-90%, where as in industrialized nations, it ranges between 10-50% [6]. The infection can take multiple courses in its progression. Most people infected with *H. pylori* never develop symptomatic disease, 10-15% develop peptic ulcer disease (gastric and

duodenal ulcers), approximately 1% develop gastric adenocarcinoma, and a small group of patients develop gastric MALT lymphoma [7]. To date, *H. pylori* is the only type I definitive human bacterial carcinogen, estimated to be responsible for 5.5% of all human cancer cases, and up to 8% of all non-Hodgkin lymphoma [8, 9]

Another striking feature of *H. pylori* infection is the wide geographical variance seen in the nature and severity of clinical outcome of the disease. Incidence of gastric cancer in Japan is approximately seven-fold higher than in the US among infected persons [10, 11], and is even rarer in South Asia (India) [10]. Similarly, duodenal ulcers are far more common than in many other geographic regions [11,12]. Peptic ulcers are rare among infected Greenland Eskimos [13] and Australian Aborigines [14], relative to that in mainstream US and European populations [15]. Although multiple factors like human genetics, diet, and infections by other pathogens that affect responses to *H. pylori* infection [16, 17] can contribute to these trends, an important role could be attributed to the genetic diversity of *H. pylori* itself.

III. Genetic diversity in *H. pylori*

H. pylori is an extremely diverse species [18, 19], and there is no single *H. pylori* strain that is “typical” for the species as a whole. Great genetic diversity, population subdivisions, and rapid evolvability are hallmarks of *H. pylori* populations. *H. pylori* population genetic structure can be classified as “panmictic”, with very little evidence of clonality or epidemic spread [20]. However, observed “panmixia” is local, and is superimposed on strong geographic differences in predominant genotypes [21-26]. Phylogenetic studies of *H. pylori* housekeeping gene sequences and insertion sequences

(*IS605* and *IS607*) revealed strong geographic clustering, wherein types of alleles from different geographic regions (e.g. East Asia, Europe, and Africa) are each distinct, and not overlapping [24-26].

Even greater geographic differences are seen in the virulence-associated *cytotoxin antigen-A* (*cagA*) and *vacuolating toxin A* (*vacA*) genes [24, 27-31]. Their encoded proteins each interact with target cells and disrupt different sets of normal cellular signal transduction pathways, with strengths and specificities that seem to vary geographically [32]. For example, East-Asian and Western-type CagA proteins differ most markedly in sequence in the domain responsible for phosphorylation (EPIYA motif) and resulting interaction with host SHP-2 phosphatase, an intracellular regulator of cell proliferative, morphogenetic and motility signaling pathways [33, 34]. Similarly, a highly active variant of the *vacA* toxin gene termed *s1, m1* (*s*: region encoding signal peptide; *m*: region encoding toxin that determines cell type specificity of VacA toxin) predominates in Japanese isolates, whereas the nontoxicogenic *s2, m2* type is relatively common in the West [29, 31]. An intermediate form designated *s1, m2* is common in coastal China [35]. Strong geographic clustering was also found in the functionally active middle region of *H. pylori* adhesin *babA* (*blood group antigen binding adhesin A*) [22]. Adherence of *H. pylori* to gastric mucosa is important for its long-term survival in gastric niche, and contributes significantly to the risk of gastric disease. These geographic differences could reflect genetic drift, or more likely types of selection pressures imposed by host physiologies that predominate (d) in the various human populations, either currently or centuries ago [17]. Given this reasoning, it seems logical that additional *H. pylori* genes whose products interact with host components might exhibit equivalent geographic

differences, also affecting how they interact with cognate host factors, which might influence clinical outcome of the gastric disease.

Thus, what drive the extraordinary genetic diversity and geographic subdivisions in *H. pylori* populations? Usually, genetic diversity in populations arises through accumulation of point mutations, recombination and genetic exchange. *H. pylori* strains: 1) have high recombination and mutation rates, which can be ascribed to a lack of *mutHLS* like pathway for DNA mismatch repair [36], 2) *H. pylori* are naturally competent for DNA transformation, and can acquire genetic material from other *H. pylori* strains or even from other species in its niche [36], and 3) *H. pylori* have extensive non-randomly distributed DNA repeat sequences that facilitate frequent intragenomic recombination, resulting in deletion or duplication of intervening DNA fragments [37]. Superimposed on these, the extraordinary chronicity of infection along with the barriers to gene flow due to geographic isolation of ancestral *H. pylori* populations cause considerable genetic drift, and thus genetic differentiation of various *H. pylori* subpopulations. Such genetically isolated subpopulations are more likely to “adapt” to variations in local environments [25, 42], and such adaptations could potentially explain the geographic variance found in the nature and severity of infection. However, is there a signature of such adaptations in the population genetic structure of *H. pylori*?

IV. Population genetic structure of *H. pylori*: role of positive selection

In general, genetic drift may allow subpopulations to explore adaptive landscapes and respond to environmental or other changes, then allowing natural selection for particular mutant or recombinant types to help fine-tune genotypes [38]. An effective

way to map selection pressures for protein coding genes is to contrast the rates at which synonymous (silent; d_s) and nonsynonymous (amino acid altering; d_N) mutations are fixed in the population [39]. The ratio d_N/d_s ($= \omega$) indicates whether an amino acid change is unaffected, inhibited or promoted by natural selection. Since most synonymous substitutions have no or very little effect on fitness, d_s is often equated to the rate of neutral nucleotide substitution and hence provides a benchmark against which to measure if d_N is accelerated or diminished by selection. Under neutral evolution, d_N is assumed to be equal to d_s ($\omega = 1$) suggesting a relaxed functional constraint. Functionally critical genes (e.g. housekeeping genes, genes responsible for metabolic functions) are expected to show very low d_N ($\omega < 1$; negative or purifying selection) [39]. However, in certain genes, nonsynonymous mutations are in excess ($\omega > 1$; positive Darwinian selection or diversifying selection), because they provide an advantage in a given environmental context, and such substitutions are more likely to be selected to change the activity or structure of the encoded protein [39, 40].

The first study that supported the notion that *H. pylori* subpopulations adapt in response to the differences in local host physiologies was the study on *H. pylori* adhesin, *babA* [22]. *H. pylori* binds to fucosylated histo-blood group antigens (Lewis b, LeB) by an outer membrane protein BabA, which is expressed by most disease causing *H. pylori* strains [41]. Strong geographic partitioning was found in BabA binding affinities to its cognate LeB receptors. *babA* alleles can be classified as generalist and specialist [22]. Generalist alleles (> 95% of *H. pylori* strains) confer binding to fucosylated blood group antigens A, B, and O [22]. However, in Amerindian populations where the predominant host blood group is O, 60% of *H. pylori* strains in these populations bound only to blood

group antigen O (specialists) [22]. Given that, diversifying selection ($\omega = 3.5$) had contributed significantly to *babA* divergence, it is plausible that the BabA protein had undergone adaptations in response to human population specific selective pressures [22].

Signatures of positive selection ($\omega > 1$) were also evident in the evolutionary history of *H. pylori* *slr* (*sel 1* like repeat) gene family expansion [42]. Sequence analysis of six of the nine known *slr* genes (*hp0160*, *hp0211*, *hp0235*, *hp0519*, *hp0628* and *hp1117*) from representative East Asian, European, and African strains revealed that all but *hp0628* had been subject to positive selection, with different amino acids often being selected in different geographic regions [42]. Most striking was a divergence of Japanese and Korean alleles of *hp0519*, with Japanese alleles having undergone particularly strong positive selection ($\omega > 25$), whereas Japanese and Korean alleles of other *slr* and housekeeping genes were intermingled [42]. Homology-based structural modeling localized most residues under positive selection to SLR protein surfaces where they would potentially interact with host components [42]. Rapid evolution of certain *slr* genes in specific *H. pylori* lineages was interpreted as reflecting geographic isolation favoring adaptive divergence of bacterial proteins driven by selection for fine tuning of host responses [42].

V. Eukaryotic like Sel1-like repeat (*slr*) containing gene family in *H. pylori*

Sel1-like repeat (SLR) motif was first identified in a *Caenorhabditis elegans* extracellular receptor protein, Sel-1 (*sel*: suppressor-enhancer of *lin*), which is a key negative regulator of the Notch pathway and regulates cell proliferation, cell fate specification, differentiation, cell death, and endosomal sorting [43-46]. Deregulated

Notch activity is oncogenic in many cases, including gastric cancer [47, 48]. Some eukaryotic SLR proteins are known to function as adaptor proteins for the assembly of membrane-bound macromolecular complexes [49-52]. Studies on bacterial SLR proteins suggest that they can aid in the adaptation of bacteria to different eukaryotic hosts [52-57]. Based on the evidence available so far, the SLR motif is absent in archaea and viruses suggesting that SLR domains have been acquired by horizontal gene transfer between bacteria and eukaryotes [58, 59] [60]. It is also possible that SLR domains have evolved in the last common ancestor of eukaryotes and bacteria playing a role in cellular differentiation [60].

The *H. pylori* *slr* gene family has eight members ('*slr*' genes) in sequenced *H. pylori* genomes [61, 62] and they belong to cluster of orthologous group (COG) 0790. Their encoded proteins are likely secreted and contain a variable (2-8) number of highly degenerate 34 amino acid Sell- repeats [63]. Six of the nine *H. pylori*'s SLR proteins are rich in cysteines, and had been designated as '*Helicobacter* cysteine rich proteins' (Hcp) [64, 65]. A disulfide bond bridges the α -helices of each SLR repeat, which is a unique feature of Hcps [60]. Although the exact *in vivo* function of *H. pylori* SLR proteins is not currently known, some Hcps bind β -lactam compounds [64, 66], which suggests possible interactions with immunomodulatory peptidoglycan fragments, that could affect the innate immune response [67]. High antibody titres against four SLR proteins [Hp0211 (HcpA), Hp0235 (HcpE), Hp0336 (HcpB), and Hp1098 (HcpC)] were found in sera of *H. pylori*-infected people [65], indicating *in vivo* expression and immune recognition. Furthermore, recombinant HcpA elicited IL12-dependent IFN- γ secretion in a naïve mouse splenocyte model [67], and induced differentiation of freely circulating monocytes

into adherent macrophages [68]. It was also shown that HP1117 elicited protective antibodies during a mouse infection [69]. *H. pylori* SLR protein, HepC (Hp0519), interacts with the multifunctional human cytoskeletal proteins Ezrin and Vinculin, and different geographic variants of HepC differ in their binding affinity to Ezrin [Putty K *et al.*, *in preparation*]. Such interaction differences likely manifest in altered Ezrin-dependent signal transduction and may influence the progression and outcome of gastric disease in different geographic regions.

Among the genomes of other related Epsilon-proteobacteria, hepatocarcinogenic *Helicobacter hepaticus* has six ORFs, and the relatively nonpathogenic *Campylobacter jejuni*, *Campylobacter lari*, *Campylobacter upsaliensis* and *Wolinella succinogenes* each have only one ORF corresponding to *slr* gene homologs [70-72]. This suggests an *H. pylori* specific expansion of *slr* gene family, perhaps reflecting a gastric niche specific adaptive role.

VI. Expansion of *H. pylori* *slr* gene family by gene duplication

A duplication event leads to relaxed selective constraints on one copy, such that it is free to diverge. In the absence of a new selective constraint such as new environment, the duplicated copy may accumulate deleterious mutations and suffer functional loss and ultimately get degraded [74-81]. However, in the presence of a positive selective pressure, randomly generated mutations by genetic drift may encode a functional variant that is critical for survival under new conditions [73-82].

Taken together, the evidence of diversifying/positive selection during the evolutionary history of a gene family is strongly suggestive of functional divergence

among related members. Maximum likelihood phylogenetic analysis using codon-based models of sequence evolution revealed that *H. pylori* *slr* gene family expansion was driven by diversifying selection ($\omega > 1$) (Fig. 1), implying functional divergence among *slr* paralogs [42]. Superimposed on this, individual *slr* genes (*hp0160*, *hp0211*, *hp0235*, *hp0519*, and *hp01117*) evolve rapidly, often in distinct human-host lineages, suggesting a key role in mediating host-pathogen interactions [42].

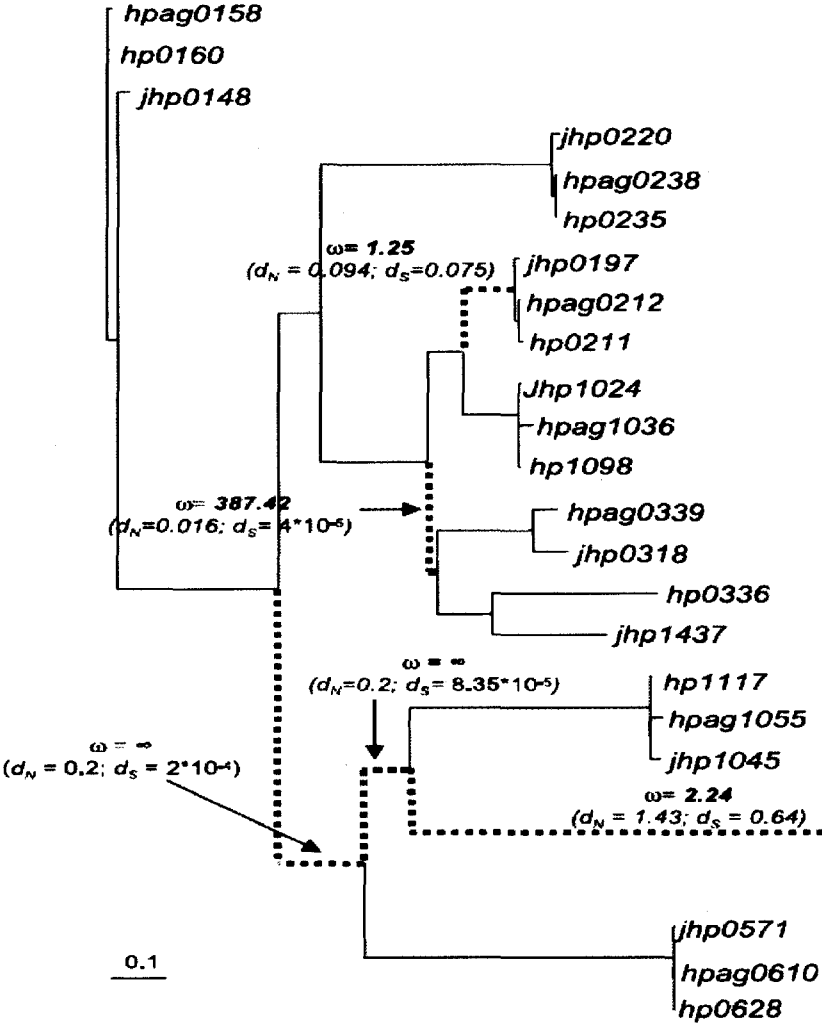


Figure 1: Multiple episodes of positive selection in *H. pylori* *slr* gene family expansion [42].

A structure-based alignment of *H. pylori* SLR proteins from strains 26695, J99, and HPAG1 was derived using EXPRESSO. A corresponding *slr* gene family sequence alignment was derived manually. An initial ML tree, used as input for selection analysis, was generated assuming the TrN + I + model of sequence evolution. Phylogeny shown above was estimated under the FR model implemented in PAML version 3.14. Branches that experienced positive selection during their evolution are indicated by dotted lines, and dN, dS, and ω values are indicated. A ω ratio = ∞ indicates branches that only accumulated nonsynonymous mutations during their divergence. Scale indicates number of substitutions per codon.

VII. Host cytoskeletal dysregulation following *H. pylori* infection

An important feature of cultured gastric epithelial cells infected with *H. pylori* is the development of a humming bird phenotype characterized by loss of cell-cell contacts, cell elongation, and eventual acquisition of motility [83, 84]. Regulated cell motility can mediate embryonic morphogenesis, tissue repair and regeneration, whereas deregulation in this highly integral process can lead to disease progression in cancer [84]. Important biological events associated with cell migration are breaking apart of intercellular complexes and cytoskeletal rearrangements, leading to elongation of the leading edge, adhesion of this protrusion to the matrix, movement of the cell body, and release of the trailing edge of the cell [84-86]. Focal adhesions (FA) and actin cytoskeleton are the important structures that regulate cell motility. FAs are comprised of a transmembrane α - β integrin heterodimer, the extracellular domain of which binds to extra cellular matrix proteins and anchors actin cytoskeleton on the cytoplasmic side of the membrane, mediating various intracellular signaling pathways [87]. Integrin activation induces recruitment and stimulation of a variety of signaling and adapter proteins (FAK, Src) that

target and activate members of Rho GTPases, which thereby regulate actin rearrangements that control cell elongation and migration [83].

Injection of CagA into epithelial cells, its phosphorylation at FAs, and subsequent deregulation of downstream signaling pathways affecting cell spreading, cell movement, and cell survival requires binding of *H. pylori* protein CagL to the β 1 integrin receptor of epithelial cells [88-91]. Phosphorylated CagA binds to: 1) Shp-2 (Src homology 2 domain-containing tyrosine phosphatase), activating Shp-2/Rap1/B-Raf/Erk signaling pathway, inducing elongated cell morphology [92-98]; 2) CrkII (v-crk sarcoma virus CT10 oncogene homolog), activating the DOCK180/Rac1/WAVE/Arp2/3 pathway, leading to the actin cytoskeletal rearrangement [99]. *H. pylori* infection also leads to the activation of Rho GTPase Rac1, activating WAVE (WASP family verprolin-homologous protein) and the Arp2/3 (actin-related protein 2/3) complex, leading to formation of broad sheet-like protrusions containing a network of branching actin filaments, lamellipodia, which are usually found at the leading edge of migrating cells [100, 101].

However, CagA independent mechanisms of actin cytoskeletal rearrangements in *H. pylori* infection were also reported [102-105], suggesting a role for other, as yet unidentified secreted *H. pylori* factors in the pathogenesis of this bacterium. Understanding the pathogenesis of *H. pylori* and identification of *H. pylori* population specific virulence determinants are areas of active research. Key questions that remain to be fully answered are the following:

1. What are the molecular signals and evolutionary forces that enable *H. pylori* to establish a chronic infection in the gastric niche in the face of continuous innate and adaptive immune responses?

2. What are the *H. pylori* induced signaling events that initiate gastric carcinogenesis?
3. Why clinical outcomes of *H. pylori* infection vary geographically?
4. Are there any clear-cut markers that can confidently predict clinical disease outcome?

SPECIFIC AIMS

Emerging data from recent studies suggest a possible role for proteins encoded by the *slr* genes in mediating and / or managing *H. pylori* - host interaction. However, the molecular and evolutionary dynamics of strain-specific *slr* genes remain to be fully characterized. Strain-specific genes often contribute to strain-specific traits. Based on this premise I posit that strain-specific *slr* genes in *H. pylori* genomes likely contribute importantly to strain-specific features during *H. pylori*'s extraordinarily chronic infection. Thus, a first goal of my thesis is to characterize the molecular evolutionary dynamics of *hcpC* and *hcpG*, each demonstrating unique strain-specific features, in order to establish a framework within which to further understand their role in *H. pylori* pathogenesis.

Gene duplication is an important driving force for biological innovation in any species. Our data showed *H. pylori* genome specific expansion of *slr* gene family, driven by strong diversifying selection [42], and led to the identification of the paralogous genes, *hcpC* and *hcpG*. However, the evolutionary forces that contributed to the origin, maintenance and preservation of these paralogous genes in natural *H. pylori* populations is not known. Furthermore, the influence of these paralogous genes upon *H. pylori* - host interaction remains to be determined.

Phylogenetic and protein structure analyses had shown that *H. pylori* *slr* gene *hp0519* (*hcpC*) alleles in different *H. pylori* lineages evolve rapidly at different rates thereby leading to selection of different amino acids in different populations, that amino

acids under positive selection in affected HepC lineages were exposed on HepC surfaces, and that positive selection altered specific amino acids in the Japanese HepC lineage and helped drive its divergence from the Korean HepC lineage; whereas in all the other *slr* genes, Japanese and Korean alleles were invariably intermingled (42). Ongoing work in the lab had identified that HepC interacts directly with the multifunctional human cytoskeletal protein Ezrin, and that different geographic variants of HepC differ in their binding affinity to Ezrin. Thus, I hypothesized that geographically distinct *H. pylori* strains likely differ in their ability to deregulate host cytoskeletal dynamics.

Taken together, although available data strongly implicate selection as a potent force during *H. pylori*'s evolutionary history it remains unclear how selective forces might have shaped *H. pylori* - host interactions. Thus, the specific aims of my dissertation are the following:

Aim 1: To characterize the molecular and evolutionary dynamics of strain-specific, paralogous *slr* genes, *hcpC* and *hcpG*.

Aim 2: To determine how *hcpC* and *hcpG* each affect *H. pylori* growth and fitness, and identify their role in mediating *H. pylori* - host interaction.

Aim 3: To determine the functional relevance of HepC during *H. pylori* - host interaction.

Aim 4: To quantify *H. pylori* population-specific variations in human cytoskeletal dysregulation.

MATERIALS AND METHODS

I. *Helicobacter pylori* culture and maintenance

A. Standard *H. pylori* culture, growth and maintenance

Bacteria were grown on plates of selective brain heart infusion (BHI) agar (Difco, KS) supplemented with 7 % (v / v) defibrinated horse blood (Cleveland Scientific, OH), isovitalex (BBL Medium enrichment for Fastidious Microorganisms, BD, France), *H. pylori* selective supplement, Dent (vancomycin 10 mg / L, cefsulodin 5 mg / L, polymyxin B 2,500 U / L, trimethoprim 5 mg / L and amphotericin B 7.5 mg / L) (Oxoid Ltd, Basingstoke, Hants, England). Erythromycin (10 µg / ml), streptomycin (10 µg / ml), and/ or chloramphenicol (10 µg / ml) were added to the agar as needed to select for mutant colonies. Cultures were incubated at 37°C, for 3-4 days in a GasPak jar (BBL Microbiology Systems, Cockeysville, Md.) with a micro aerobic gas mixture (CampyPak, BBL Microbiology Systems, Cockeysville, MD) composed of 5 % oxygen, 10 % carbon dioxide, and 85 % nitrogen [106]. Strains were maintained as frozen stocks by scraping bacterial growth from agar plates with a sterile loop in BHI broth (Difco, KS) with 15 % glycerol, aliquoted in 1ml freezer tubes, and stored at -74°C.

B. Culture of *H. pylori* in liquid media

H. pylori was grown in liquid medium using BHI broth supplemented with 1% (v / v) Isovitalex, 1% *H. pylori* selective supplement, and 10% fetal bovine serum (FBS)

(GIBCO, CA) [107, 108]. Liquid medium in 50ml tissue culture flask was inoculated with the starting bacterial suspension (OD_{600} 0.05/ml and incubated at 37°C for 30 min in CO₂ incubator. Then the flasks were transferred to GasPak jars and incubated at 37°C with shaking (125 rpm) for a maximum of 56hrs [106].

II. DNA Extraction, PCR-conditions, and DNA sequencing.

A. Genomic DNA isolation

H. pylori genomic DNA was isolated from confluent cultures on agar plates by the hexadecyltrimethylammonium bromide (CTAB) method [109, 110]. Briefly, 1×10^8 bacteria were collected and washed twice in 1X TE buffer. To the cell pellet, SDS and proteinase K were added to a final concentration of 0.5% (v / v) and 100 µg / ml, respectively and the mixture was incubated at 37°C for 1 hr. 80 µl of CTAB / NaCl solution was then added and incubated for 20 min at 65°C. Proteins were then removed using phenol / chloroform / isoamyl alcohol method. DNA was precipitated with 0.6 volumes of 100% v / v isopropanol, washed twice with 70% ethanol to remove residual CTAB, dried and suspended in 1X TE buffer. Genomic DNA was stored at -20°C until required.

B. PCR conditions and DNA sequencing

i) General PCR conditions:

Specific PCR reactions were carried out in 25 µl mixtures containing 5 to 10 ng of DNA, 1 U of *Taq* polymerase (Biolase; Midwest Scientific, St. Louis, Mo.), 1.5 mM of MgCl₂ (Biolase; Midwest Scientific, St. Louis, Mo.), 0.8 pmol of each forward and

reverse primers, and each deoxynucleoside triphosphate at a concentration of 0.1 mM in a standard buffer for 30 cycles with the following cycling parameters: denaturation at 94°C for 30 s, annealing at a temperature appropriate for the primer sequence (generally 54°C) for 30 s, and DNA elongation at 72°C for an appropriate time (1 minute per kb).

ii) *hcpC* and *hcpG* amplifications:

To amplify complete nucleotide sequences of *hcpC* and *hcpG* homologs, and to detect their chromosomal conservation in genetically diverse *H. pylori* isolates, one set of primers was designed located within *jhp1023* and *jhp1025*, the genes flanking *jhp1024* (the J99 homolog of *hcpC*), and in *jhp1436* and *jhp1438*, the genes flanking *jhp1437* (the J99 homolog of *hcpG*), respectively. To help prevent primers from being designed in segments with high frequencies of mutations, *jhp1023* was aligned with *hpag11035* from strain HPAG1, and *jhp1025* was aligned with *hpag1_1037* from HPAG1 and *hp1099* from 26695. Similarly, *jhp1436* was aligned with *hpg27_1468*, and *jhp1438* was aligned with *hpg27_1470* from strain G27MA. This alignment was done in MegAlign (DNASTAR package; Lasergene Inc., USA). Once the flanking genes were aligned, primers were designed in regions of the sequence conserved in the alignment. The resulting primers, *hcpCF1* (Primer #1; Table 2) located in *jhp1023* and *hcpCR1* (Primer #2; Table 2) located in *jhp1025* amplified an approximately 1643 base pairs long segment; *hcpGF1* (Primer #3; Table 2) located in *jhp1436* and *hcpGR1* (Primer #4; Table 2) located in *jhp1438* amplified 988 bp long segment, in strain J99. To increase sequence redundancy coverage of *hcpC* sequences missing 30 or more base pairs due to a failure of the sequences from both primers to overlap, a third internal primer was designed

(1098Fseq) (Primer #5; Table 2). This primer was designed 200 base pairs within *hp1098* at a site found to be well conserved when several of the first sequences to be analyzed were aligned with *hp1098*. This primer was then used to sequence *hcpC* homologs that showed incomplete or ambiguous data. Since no amplification was observed in approximately 50% of *H. pylori* strains tested using *hcpG* flanking primers, a second set of primers (*hcpGF2* and *hcpGR2*) (Primer #8 & 9; Table 2) was designed within *jhp1437* to amplify a 750bp product at sites found to be well conserved when several sequences obtained initially were analyzed. To confirm either the absence of *hcpG* homolog in the genome or its chromosomal rearrangement, *hcpGF2* and *hcpGR2* primers were then used to amplify those *H. pylori* strains that showed no amplification or a smaller amplicon than expected fragment size using primers designed in the flanking genes of *jhp1437*. These primers were then used to amplify *hcpC* (by Sarah Marcus under my guidance as part of her undergraduate honors thesis) and *hcpG* homologs from 166 strains of *H. pylori* from Spain, Japan, Korea, Gambia, South Africa, India, Peru, and Shima (Peruvian Amazon). These strains had been isolated from patients with gastric complaints that underwent a diagnostic endoscopy with informed consent. Amplification of these genes and the specificity of PCR amplification were confirmed using 1% agarose gel electrophoresis. Only amplifications that yielded a single specific product were processed further.

iii) DNA sequencing: *hcpG* and *hcpC* sequence determination:

The DNA concentration of PCR products was determined using a BioPhotometer (Eppendorf) to ensure quality products for sequencing. 10 µl of the amplified product was

then placed in two 96 - well plates for sequencing. Each plate was paired with a second plate containing either 10 μ l of primer hcpCF1 or hcpCR1, and hcpGF1 or hcpGR1; these primers are from either end and were used to sequence *hcpC* and *hcpG* homologs, respectively. The plates were then shipped to High-Throughput Genomics Unit, Seattle, WA, where the PCR products were purified and sequenced on both strands using an ABI377 automated sequencer.

III. Computational Biological Analyses.

A. DNA Sequence Data Processing and Management

Sequences from both forward and reverse strands were assembled and edited in Seqman (DNASTAR, Lasergene Inc., WI). *hcpC* and *hcpG* sequences from the available *H. pylori* genome sequences in Genbank were included in the respective alignments to ensure proper alignment, and to properly trim the sequence data to only the sequence of the gene of interest. Seqman was also used to view the sequence data to ensure that the sequence data was not contaminated, due to nonspecific amplification for example, or incorrectly interpreted by the computer. Any errors were noted and corrected if the data clearly showed a peak for a different nucleotide than the one indicated. Those sequences that had poor data, possibly from contamination, were removed from the group, and sequencing was repeated.

B. Multiple Sequence Alignments

Complete sequences of the *hcpC*, and *hcpG* homologs were then translated into amino acids and aligned initially in MegaAlign using Clustal W and default settings. A

preliminary phylogenetic tree was then reconstructed in MEGA version 5 using the Neighbor-joining method, and bootstrap analysis with 1000 replicates. The Mega5 alignment was saved as a nexus file for use in PAUP*4b10 (<http://paup.csit.fsu.edu/>). The alignment was then used to create a neighbor-joining tree used as input for MODELTEST, version 3.7 (<http://darwin.uvigo.es/software/modeltest.html>). MODELTEST determines the model of DNA evolution that best describes the given data. This test is important for choosing an appropriate model, because although more complex tests add more parameters to make a model more realistic, additional parameters also leave more room for errors. Therefore a model should not be chosen if it is more complex than the data provided merits.

C. Phylogenetic Reconstruction

The MODELTEST program [111] was used to find the best model by measuring the likelihood that a null model fits the given data compared to the likelihood that an alternative model fits given the same data. The first test in the program, hierarchical likelihood ratio tests (hLRTs) compares the likelihoods of nested models, or models which build off one another, becoming more complex with more parameters as the program continues the comparisons [112]. These models can easily be compared for goodness of fit using a chi square test and the model with the highest likelihood of a pair is used for the next comparison through a series of more complex models, until the model of best fit is determined. The second test run in MODELTEST is the Akaike information criterion (AIC). This test can function without consecutive models being nested, and it again tests for goodness of fit. It also adds a penalty for unnecessary parameters. Both of

these tests will not only determine the model of best fit, but will give the estimated values for each of the parameters used in that model.

Once the model of best-fit was determined, a maximum likelihood (ML) phylogeny was reconstructed under this model by using a combination of heuristic searches and branch swapping to further optimize the likelihood score and substitution parameters. ML generates a tree that most likely explains the given data. To find which tree is most likely, a heuristic search was conducted starting with a single tree, and altering it in a single way. If the new tree is more likely than the previous tree, the new tree was kept and the old tree was discarded, and the process was repeated sometimes over 100,000 times. If the original tree is more likely, it would be kept until an improved tree was found. This process was continued to search through trees until the most likely tree was found. Although a heuristic search is not an exhaustive search through all possible trees, leaving the possibility that only suboptimal tree is given as the most likely tree, this approach is necessary due to the exponential number of possible trees and the many weeks that would be needed to sort through all of them computationally.

The significance of the observed phylogenetic groupings was assessed by a bootstrap analysis performed with 1,000 replicates under the distance optimality criterion, while incorporating the ML-optimized model and parameters. Bootstrapping was performed in order to estimate sampling error by sampling from within the provided data creating pseudoreplicates, which is used to generate a new tree. This method is then repeated, here 1000 times, and a bootstrap consensus tree displaying the most frequent splits labeled at their nodes with their frequencies. These frequencies help determine the precision of the methods used to estimate the tree with the given data, providing an

estimate of confidence in the phylogeny. Phylogenetic trees were visualized with TreeView version 1.6.6 (<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>), and were edited in Adobe photoshop CS8 software.

D. Analyses of Selection Pressures

The selective pressures operating on *hcpC* and *hcpG* were measured using an ML method that takes into account the sequence phylogeny and assesses the fit to the data of various models of codon evolution that differ in how ω varies across the sequence or across the phylogeny. Site specific models (SSMs) in codon-based analysis (M0, M1, M2, M3, M7, M8, and M8a) assume a single ω for all branches of the tree, but allow ω to vary among individual codon sites, thereby providing a measure of heterogeneity in selection pressures acting across the gene sequence [113, 114]. Positive selection was inferred when codons with ω of >1 were identified and the likelihood score (-lnL) of the codon substitution model in question was significantly higher than the likelihood of a nested model that did not take positive selection into account. The probability that a specific codon belonged to the neutral, negative, or positively selected class was calculated using empirical Bayes methods implemented in PAML version 4.14 (<http://abacus.gene.ucl.ac.uk/software/paml.html>). Multiple runs, assuming different initial ω and κ values, and different models for estimating equilibrium codon frequencies (to be calculated from the average nucleotide frequencies at the three codon positions tables (F3X4) or used as free parameters), were analyzed for *hcpC*, and *hcpG* to verify the convergence optima for each model.

E. Domain Architecture Analyses

Identification and annotation of domain architectures of HcpC, and HcpG variants was performed using a Simple Modular Architecture Research Tool (SMART), PFAM and Signal IP. Domains were extensively and confidently annotated with respect to phyletic distributions, functional class, tertiary structures and functionally important residues. Parameters are given in Appendix I.

F. Homology Modeling

Homology modeling of HcpG variants was performed using a web based version of Modeller v 9.8, webmod. Available HcpB and HcpC crystal structures were used as template models for the analysis.

IV. Genetic Engineering of *H. pylori*

A. General Method for Introducing Marked Deletions in *H. pylori*

The general principle for generating *H. pylori* *slr* gene deletions in this study was described previously [115]. Briefly, a streptomycin resistant (str^{R}) allele of *rpsL* (str^{S}) was generated by PCR by introducing A-to-G changes at each of the two sites most frequently responsible for streptomycin resistance (codons 43 and 88; Lys to Arg in each case) [116]. This allele was then used to transform str^{S} 26695 *H. pylori* strain to generate a str^{R} 26695. Genomic DNA of str^{R} 26695 was then used to transform G27MA (cell culture adapted *H. pylori* strain) to generate str^{R} G27MA. *rpsL-erm* or *rpsL-cat* cassettes conferring streptomycin susceptibility (*rpsL* gene from *Campylobacter jejuni*, codes for ribosomal protein S12, a dominant allele for streptomycin sensitivity) and erythromycin

or chloramphenicol resistance, respectively were then used to generate PCR based constructs where the flanking genes of antibiotic cassettes were homologous to genes flanking the targeted gene deletion. These constructs were generated by assembling individual PCR products which have overlaps of ≥ 20 bp at the ends of fragments to be joined together, which in turn, result from the design of PCR primers used in amplification (sewing - PCR) [117, 118] (Fig. 2). Primers 2 and 5, each on the inside ends of their respective fragments were designed not only to match the sequence at the end of the fragments amplified, but also to have approximately 20 bp overlap with fragment B. Primer 2 began at its 5' end with a segment that is the reverse complement of primer 3, and primer 5 begins with a segment that is the reverse complement of primer 4. These overlaps are essential for assembly of fragments A, B, and C. Once the primers were designed, standard PCR amplification was used to separately generate PCR products A, B, and C. These three products were then combined and amplified using PCR with primers 1 and 6. The overlapping segments at the ends of fragments A and B and at the starts of fragments B and C allow the fragments to be "sewn" together as amplification occurs, resulting in a final contiguous PCR product ABC (Fig. 1) [115]. Such constructs were then used to transform str^R strains to select for erythromycin or chloramphenicol resistance and screen for str^S mutant colonies.

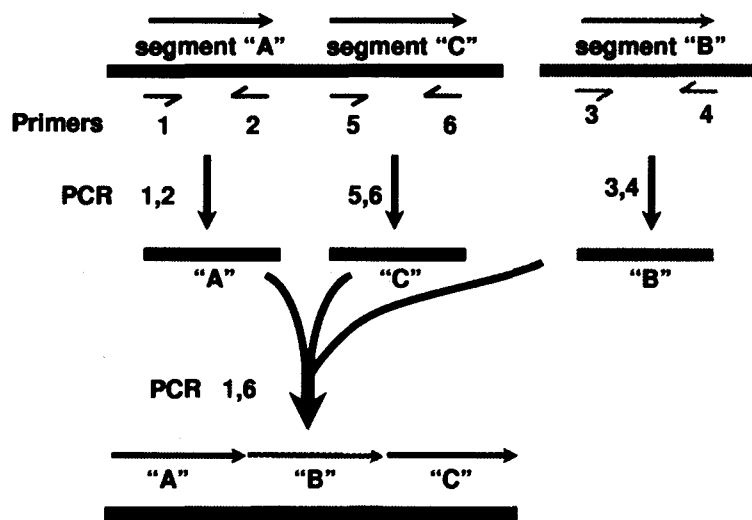


Figure 2: Construction of insertion and deletion alleles by assembly of three fragments with overlapping ends [115].

Half arrows indicate the positions of the primers. Segments A and C represent DNA segments flanking the locus to be deleted or the site of insertion (depending on needs of experiment). Segment B represents (i) the *erm* (resistance) gene used initially to replace the *mdaB* locus; (ii) the *rpsL* streptomycin susceptibility gene from *C. jejuni* that was inserted just upstream of *erm* in a strain carrying *erm* in place of *mdaB*; or (iii) the *rpsL,erm* cassette, which can be moved to many loci [115] (with permission from DEB).

B. Engineering $\Delta hcpC$ mutation in *H. pylori* strain G27MA.

The overall strategy used for generating $\Delta hcpC$ mutation was similar to that described above with a few modifications as noted below. Three sets of primers were designed. The first set for amplification included 1098A1n and 1098A2n as primers 1 and 2, respectively, for the amplification of a 437bp long fragment A. Fragment A began at the end of *hp1098* with hp1098A1n (primer # 8; table 2) starting 804 bp within the gene and ending at the start of primer hp1098A2n (primer # 9; table 2) 368 bp within the gene. This fragment was not positioned in the gene downstream of *hp1098* (*hp1097*), as is

described in the general procedure above, because homologs to *hp1097* are not present in all genomes from which homologs to *hp1098* may need to be amplified to transform back into the knockout. Fragment B, as described above, was the *rpsL-erm* cassette and primers *rpsL-F-1* (primer # 10; table 2) and *erm-R* (primer # 11; table 2), corresponding to primers 3 and 4, respectively, were used for its amplification. Their sequences are also the reverse complements for the overlapping sequences found at the 5' ends of 1098A2n, which overlaps with *rpsL-F-1*, and 1098C5n, overlapping with *erm-R*. The final set of primers, 1098C5n (primer # 12; table 2) and 1098C6n (primer # 13; table 2), corresponding to primers 5 and 6, amplified a 440 bp long fragment C. This fragment began at the start of primer 1098C5n, 20 bps within *hp1098* and extended to the start of 1098C6n placed 276 bps from the start of *hp1099*. Primers were synthesized by Integrated DNA technologies, Inc. IA. Using the primers described above, fragments A, B, and C were each amplified using a standard PCR. Fragments A and C were run with an annealing temperature of 55°C and fragment B with an annealing temperature of 58°C. All of the PCR products were purified by spin column purification. 3 µl of each of the products from the first round of amplification were used in an assembly PCR using primers 1098A1n and 1098C6n with an annealing temperature of 54°C. Similarly, 3 µl of product B and the PCR products A and C from the alternative primers were used for two more assembly PCRs with annealing temperatures of 55°C and 58°C. Poor amplification of the PCR assembly led to an attempt at a two part assembly, which has been found to be successful when the single three fragment assembly fails [119]. For this PCR assembly, 3 µl of alternative fragment A and fragment B were amplified together, and alternative fragment C and fragment B were amplified together to form two combined

fragments: AB and BC. 3 μ l of each of these PCR products were then used for a second PCR assembly to give the final ABC assembly which contained the *rpsL-erm* cassette within *hp1098* (Fig. 2). The final assembly was then used to transform *H. pylori* strain 26695 Str^R which was grown overnight and re streaked in the center of a fresh plate in the morning. The transformation was completed by adding 10 μ l of the ABC assembly to the cells and mixing them with a loop at the center of the plate. These cells were then grown overnight and the following day they were streaked on a fresh plate containing erythromycin to select for Ery^R and presence of the *rpsL-erm* cassette. After growing several days, individual Ery^R colonies were picked using toothpicks and streaked in a line on corresponding erythromycin- and streptomycin-containing plates. Colonies observed on the erythromycin plate and not on the streptomycin plate were then streaked across individual erythromycin plates so that the cells could be used to make frozen stocks and streptomycin plates as a final check for Str^S. Representative colonies that passed the antibiotic screening as both Ery^R and Str^S were then tested by PCR using primers 1098A1n and 1098C6n. Colonies showing PCR products matching the expected PCR product size of 2388bp were considered to be successful knockouts with the *rpsL-erm* cassette inserted in and replacing part of *hp1098* (*hcpC*). The *26695- Δ hcpC* genomic DNA was then used to transform *H. pylori* strain G27MA Str^R to generate a *G27MA- Δ hcpC-str^S* and erm^R strain. This *hcpC* mutant strain in Hp26695 was generated by Sarah Marcus in the lab of Dr. Douglas E. Berg as a part of her undergraduate honors thesis.

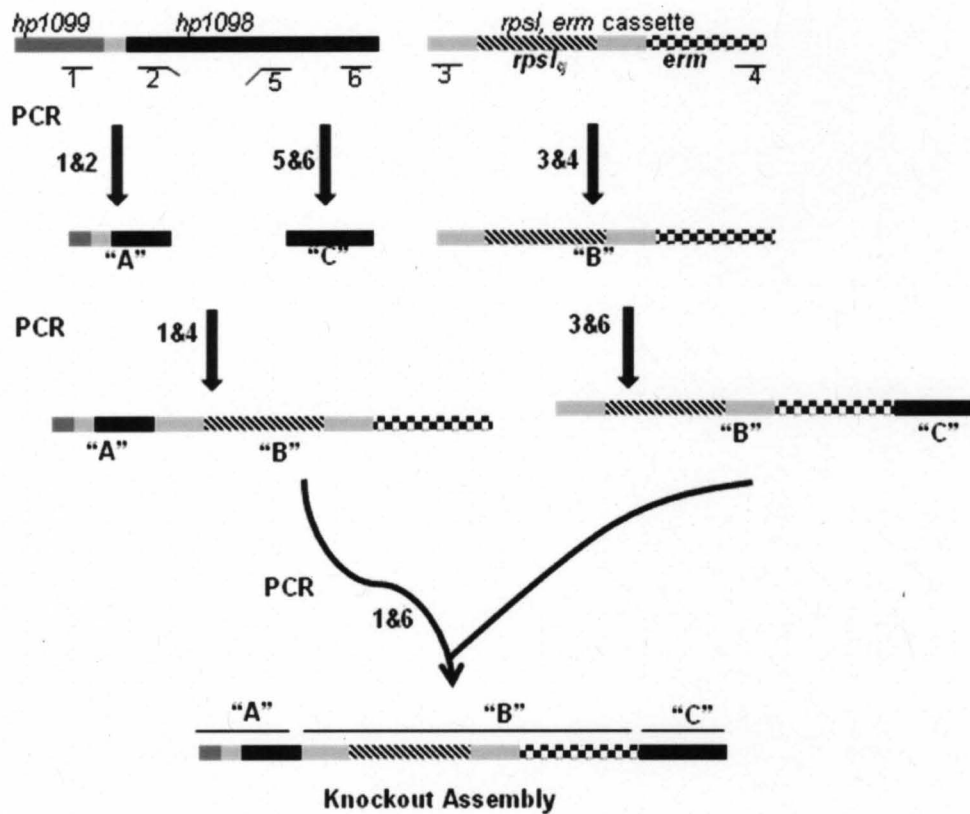


Figure 3: Strategy for knocking out *hcpC* homolog in *H. pylori* strain 26695 StrR.

Primer 2 overlaps (hanging line) with primer 3 and 5 overlaps with 4 so that fragments A and C, on either side of *hcp1098* overlaps with the ends of B in the final assembly. Light gray filled boxes represent non coding sequences. Fragment B (*rpsL-erm*) confers streptomycin susceptibility and erythromycin resistance. Genomic DNA of *H. pylori* strain 26695-*hcpC* Δ was then used to transform G27MA-StrR *H. pylori* strain to generate G27MA:*hcpC* Δ -StrS and *ermR*.

C. Engineering Δ *hcpG* mutation in *H. pylori* strain G27MA

Three sets of primers were designed. The first set for amplification included *hcpGAF* and *hcpGAR* as primers for the amplification of a 700 bp long fragment A. Fragment A began at 24 bp of *jhp1436* and ends at the beginning of *jhp1437*. Fragment B, was the *rpsL,cat* cassette and primers *rpsL-F-1* and *cat-R*, were used for its

amplification. Their sequences are also the reverse complements for the overlapping sequences found at the 5' ends of hcpGAR, which overlaps with rpsL-F-1, and hcpGCF, overlapping with erm-R. The final set of primers, hcpGCF and hcpGCR would amplify a 640bp long fragment C. This fragment began at the start of primer hcpGCF, at the end of *jhp1437* and extended to the 640 bp into *jhp1438*. Primers were synthesized by Integrated DNA technologies, Inc. IA. Using the primers described above, fragments A, B, and C were each amplified using a standard PCR. All of the PCR products were purified by spin column purification. 3 µl of each of the products from the first round of amplification were used in an assembly PCR using primers hcpGAF and hcpGCR with an annealing temperature of 54°C. Poor amplification of the PCR assembly led to the alternate procedure as described above where fragments AB and BC were generated first and then “sewn” together resulting in ABC. However, even this procedure resulted in poor amplifications. As a result, this procedure was used with few modifications. An alternate set of four primers were then designed. Primer hcpGAR was replaced with hcpGAR2 (primer 2) (primer # 15; Table 2) with the same sequence as hcpGAR along with an added restriction site of *EcoRI*-GAATTC at its 5' end. Primer rpsL F2 (primer 3) (primer # 16; table 2) had restriction site GAATTC added to its 5' end. Primer catR2 (primer 4) (primer # 17; table 2) had restriction site GCGGCCGC-*NotI* at its 5' end which matches the 5' end of primer hcpGCF2 (primer 5) (primer # 18; table 2). Forward primer of fragment A (hcpGAF) (primer 1) (primer # 14; table 2) and reverse primer of fragment C (hcpGCR) (primer 6) (primer # 19; table 2) were the same as before. First, fragments A, B, and C were amplified using alternate set of primers. PCR products were then purified using Qiagen PCR purification kit. 1 µg of each fragment was then taken and restriction

enzymes *EcoRI* and *NotI* were then added to the mix along with 1X digestion buffer making a final volume of 20 μ l. Restriction digestion was then performed at 37°C for 3hrs. Digested products were then purified and ligated with the addition of T4 DNA ligase in a 30 μ l total reaction volume and incubated overnight at 16°C. Primers hcpGAF and hcpGCR were then used to amplify fragment ABC from the ligated mix and the PCR product was run on a 1 % agarose gel to confirm the presence of ABC fragment at the expected size 2950 bp. The PCR construct was then used to transform *H. pylori* strain G27MA str^R which was grown overnight and restreaked in the center of a fresh plate in the morning. The transformation was completed by adding 10 μ l of the ABC assembly to the cells and mixing them with a loop at the center of the plate. These cells were then grown overnight and the following day they were streaked on a fresh plate containing chloramphenicol to select for Cat^R and presence of the *rpsL,cat* cassette. After growing several days, individual Cat^R colonies were picked using toothpicks and streaked in a line on corresponding chloramphenicol- and streptomycin-containing plates. Colonies observed on the chloramphenicol plate and not on the streptomycin plate were then streaked across individual chloramphenicol plates so that the cells could be used to make frozen stocks and streptomycin plates as a final check for Str^S. Colonies that passed the antibiotic screening as both Cat^R and Str^S then had their genomes amplified by PCR using primers hcpGAF and hcpGCR. Colonies showing PCR products matching the expected PCR product size of 2950bp were considered to be successful knockouts with the *rpsL-cat* cassette inserted in place of *hcpG* (Fig. 3). The PCR products were finally sequenced to check for any unwanted mutations in flanking regions other than the restriction sites that were introduced immediately before and after the *rpsL-cat* cassette.

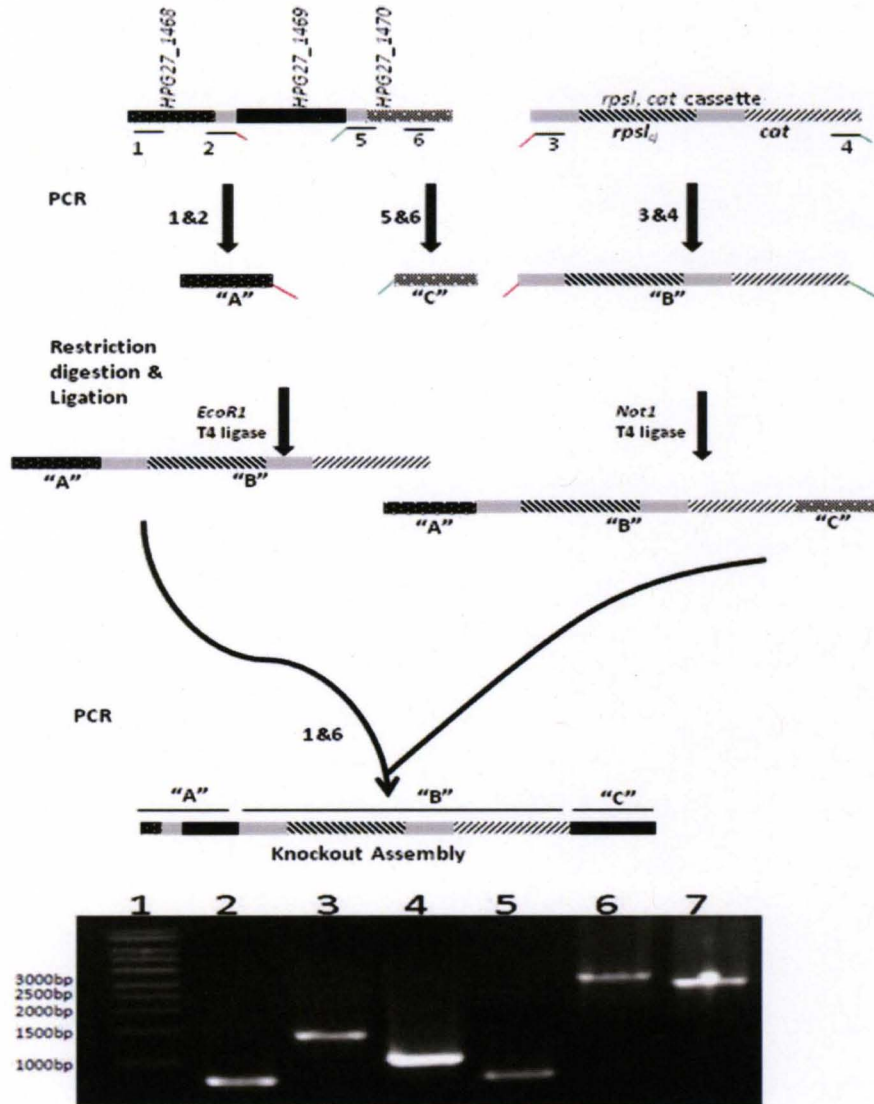


Figure 4: Strategy for knocking out *hcpG* homolog in *H. pylori* strain G27MA.

Fragment A has restriction site *GAATTC* at its 3' end which matches with restriction site at 5' end of fragment B (red overhang). Fragment C has restriction site *GCGGCCGC* at its 5' end which matches with restriction site at 3' end of fragment B (green overhang). Fragment B (*rpsL-cat*) confers streptomycin susceptibility and chloramphenicol resistance. Restriction digestion with enzymes *EcoR1* (red), *Not1* (green); ligation with T4 DNA ligase and subsequent PCR with primers 1 and 6, yields knockout assembly ABC, which was then used to transform *G27MA-StrR H. pylori* strain, to generate *G27MA-hcpGA-StrS and catR*. The same strategy was used to

generate *H. pylori* strain G27MA-HcpG 6XHis where fragment B (*rpsl-cat*) was replaced with *hcpG::6XHis*. Lanes 1, 2, 3, 4, 5, 6, and 7 in the gel have molecular weight marker, fragments A, B (*rpsl-cat*), B (*hcpG::His*), C, ABC (B-*rpsl-cat*), ABC (B-*hcpG::His*), respectively.

D. Engineering HcpG::6XHis insertion in G27MA Δ hcpG derivative

The strategy described in section IVC was also used to generate *H. pylori* strain G27MA-HcpG::6X His, expressing HcpG with a 6X Histidine tag at its C - terminus. Primers hcpGBF and hcpGBR (Table 2) were designed to amplify *hcpG* homolog in *H. pylori* strain G27MA with a 6X His tag at its 3' end. This was achieved by adding a 6X His tag sequence to the 5' end of hcpGBR sequence. Primers hcpGBF (primer # 20; table 2) and hcpGBR (primer # 21; table 2) also contained restriction sites for EcoR1 and Not1 at their 5' ends so that fragment B (*hcpG::6X His*) matched with fragments A and C at its 5' end and 3' end, respectively. Using fragments A, B, and C, construct ABC (2250 bp) was generated as described above and the construct was then used to transform G27MA- Δ hcpG Cat^R strain, which was grown overnight and restreaked in the center of a fresh plate in the morning. The transformation was completed by adding 10 μ l of the AB2C assembly to the cells and mixing them with a loop at the center of the plate. These cells were then grown overnight and the following day they were streaked on a fresh plate containing streptomycin to test for Cat^S and presence of the *hcpG::His*. After growing several days, individual Str^R colonies were picked using toothpicks and streaked in a line on corresponding streptomycin and chloramphenicol containing plates. Colonies observed on the streptomycin plate and not on the chloramphenicol plate were then streaked across individual streptomycin plates as a final check for Str^R so that the cells could be used to make frozen stocks. Expression of functional HcpG::6XHis in the

G27MA-HcpG::6XHis strain was confirmed by monitoring expression of HcpG in *in vitro* grown G27MA-HcpG::6XHis by reverse transcription PCR, and by detecting HcpG::6X His by immuno blotting with an anti-His antibody (described below).

E. Engineering $\Delta hcpC\Delta hcpG$ mutations in *H. pylori* strain G27MA

Genomic DNA was extracted as described in earlier section from the *H. pylori* G27MA- $\Delta hcpC$ strains generated above. 10 μ l of this genomic DNA was used to transform G27MA- $\Delta hcpG$ Cat^R strain which was grown overnight on a chloramphenicol plate and restreaked in the center of a fresh plate in the morning. Genomic DNA was mixed with bacteria on the plate and mixed with a loop at the center of the plate. These cells were then grown overnight and the following day they were streaked on a fresh plate containing both erythromycin and chloramphenicol to select for Erm^R and Cat^R respectively (and absence of *hcpC* and *hcpG*, respectively). After growing several days, individual Erm^R and Cat^R colonies were picked using toothpicks and streaked in a line on corresponding chloramphenicol-erythromycin and streptomycin containing plates. Colonies observed on the chloramphenicol-erythromycin plate and not on the streptomycin plate were then streaked across individual chloramphenicol-erythromycin plates as a final check for Erm^R and Cat^R so that the cells could be used to make frozen stocks. Colonies that passed the antibiotic screening as both Cat^R and Erm^R then had their chromosomal regions amplified by PCR using primers 1098A1n and 1098C6n and, *hcpGAF* and *hcpGCR*. Colonies showing PCR products matching the expected PCR product size of 2388bp (with primer pairs 1098A1n and 1098C6n) and 2950 bp (with primer pairs *hcpGAF* and *hcpGCR*) were considered to be successful knockouts with the

rpsL-erm and *rpsL-cat* cassettes inserted in respectively and replacing part of *hcpC* and whole *hcpG*, respectively in *H. pylori* strain G27MA.

F. Engineering Δ *hcpC* mutations in *H. pylori* strains G27MA and 26695

hcpC deletions in *H. pylori* strains 26695 and G27MA were generated as described in section IVA and were kindly provided by Dr. Douglas E. Berg and Dr. Ozge Darka (Washington University School of Medicine, St. Louis, MO).

V. Cell Culture and Antibodies

A. Culture and maintenance of cell line, AGS

ATCC CRL 1739, a human gastric adenocarcinoma epithelial cell line (AGS) was cultured and maintained in antibiotic free DMEM-high glucose (Sigma), supplemented with 10 % heat inactivated fetal bovine serum (GIBCO). Cells were grown and maintained in 50 ml tissue culture flasks at 37°C in a humidified atmosphere of 5% CO₂.

B. AGS cell culture for assaying infection dynamics, including cytoskeletal assays

48 hrs before infection, AGS cells were seeded in 6-well plates at a density of 0.5×10^5 cells / well after a viability count in haemocytometer, using the trypan blue exclusion assay. 12 hrs before infection, the cells were maintained in serum free medium. For the infection assays, bacteria were harvested and washed in PBS (pH 7.4), and were diluted to a multiplicity of infection (MOI) of 100 i.e., 100 *H. pylori* / AGS cell. To synchronize the infection, the plates were centrifuged for 10 min at 1000 X g using a SORVALL Legend RT centrifuge and incubated for indicated time intervals at 37°C in a

humidified atmosphere of 5% CO₂. To quantify the AGS cell scattering phenotype 6 hrs post infection, images of infected and uninfected AGS cells were taken at 20X magnification in a Nikon Eclipse TE-2000U fluorescent microscope. Cell extensions were measured using Metamorph software and extensions measuring over 40 µm from the center of the cell were considered “scattered” or “hummingbird” phenotype [120]. A total of 100 cells were counted in each field and a minimum of three such fields were counted in each experiment. Uninfected AGS cells were used as controls wherever necessary.

C. AGS cell culture for assaying bacterial growth and fitness

24 hrs before infection, AGS cells were seeded in 6-well plates at a density of 10⁵ cells / well, after doing a viability count in haemocytometer using trypan blue exclusion assay. Since *H. pylori* growth dynamics was monitored here, AGS cells were not serum starved before the infection. Instead, wells were supplemented with 20% FBS during infection. For the infection, desired bacterial strains were harvested in PBS (pH 7.4), washed with PBS, and were diluted to a multiplicity of infection (MOI) of 100. To synchronize the infection, the plates were centrifuged for 10 min at 1000 X g using a SORVALL Legend RT centrifuge and incubated for indicated time points at 37°C in a humidified atmosphere of 5% CO₂.

D. Antibodies used in the study

Primary antibodies used in this study are: anti-CagA (goat IgG) sc-34039, anti-VASP (mouse IgG) sc-46668, anti-*Helicobacter pylori* HSP (mouse IgG) sc-57779

[purchased from Santacruz Biotechnology, Santa Cruz, CA]; anti-phospho-Map kinase kinase (rabbit IgG) M-76783, anti-HIS (mouse IgG), anti-FLAG (mouse IgG), anti-GST (mouse IgG), anti-tubulin (mouse IgG) [kindly provided by Dr. Swathi Arur, MD Anderson Cancer Center, TX]. Secondary antibodies used in this study were: horseradish peroxidase conjugated secondary (HRP - mouse and HRP - goat) [kindly provided by Dr. Swathi Arur, MD Anderson Cancer Center, TX], Goat polyclonal Secondary Antibody to Mouse IgG - H&L (FITC) ab6785, Donkey polyclonal Secondary Antibody to Goat IgG - H&L (FITC) ab6881, Donkey F(ab')₂ polyclonal Secondary Antibody to Rabbit IgG - H&L (PE) ab7007, Goat polyclonal Secondary Antibody to Mouse IgG - H&L (TR) ab6787 [purchased from Abcam, MA or kindly provided by Dr. Palaniappan Sethu, Dept. of Bioengineering, University of Louisville, KY].

VI. Immunoblotting

Crude cell lysates or purified protein preparations were separated on 5-10% Tris-HCl polyacrylamide gels and then transferred to polyvinylidene difluoride (PVDF) membranes (Millipore) with the Bio-Rad Mini-Protean system. Membranes were blocked in TBS-T (140 mM NaCl, 2.7 mM KCl, 25 mM Tris-HCl, 0.1% (V/V) Tween) with 5% non-fat dry milk for 1h at room temperature, and washed 3 times with TBST. Membranes were then incubated in desired primary antibodies overnight at 4°C. After washing thrice with TBST, primary antibodies were detected using horseradish peroxidase conjugated secondary antibodies and visualized by Western Lightning Chemiluminescence Reagent (Pierce) according to the manufacturer's instructions, using Versadoc 4000. Between

blotting procedures, the membranes were stripped for 30 min at 50°C in stripping buffer (62.5 mM Tris-HCl, pH 6.7, 100 mM 2-mercaptoethanol, 2% SDS).

VII. Growth Kinetics and Fitness Assays.

A. *In vitro* Growth Kinetics and Fitness Assays

i) *In vitro* growth dynamics of engineered HpG27MA derivatives:

To assess the *in vitro* growth dynamics of G27MA WT and its *hcp* mutants (*hcpC* mutant or *hcpG* mutant or *hcpC*, *hcpG* double mutant) tested, bacterial strains were grown on BHI agar plates for 3 days as described in section IA with selective antibiotics. Individual *H. pylori* strains were then inoculated and grown in BHI broth as described in section IB. Serial dilutions of bacterial aliquots from the inoculums were plated on selective plates: G27MA WT - streptomycin, *G27MAΔhcpC* - erythromycin, *G27MAΔhcpG* - chloramphenicol, *G27MAΔhcpC* : *ΔhcpG* - erythromycin and chloramphenicol to obtain 0 hr bacteria counts. At 6 hr (lag growth phase), 24 hr (log growth phase), 48 hr (stationary phase), and 56 hr (late stationary phase), bacterial aliquots from the cultures were taken, serial dilutions made and plated on selective plates for counting colony forming units (CFUs). The plates were incubated for 4-5 days as described in section IB. At the end of incubation, colonies were counted and the number of bacteria was expressed as the number of CFU / ml. The experiment was repeated three more times and statistical significance between groups was calculated using a Student's *t*-test.

ii) Competition and relative fitness assays with HpG27MA derivatives and G27MA WT:

To assess the relative fitness of G27MA WT and each of the *hcp* mutants, (*hcpC* mutant or *hcpG* mutant or *hcpChcpG* double mutant) *in vitro*, each of the mutant was competed with the WT. Bacterial strains were grown on BHI agar plates for 3 days as described in section IA. Each BHI broth tube was inoculated with G27MA str^R along with either *hcpC* mutant or *hcpG* mutant or *hcpChcpG* double mutant with a starting bacterial OD₆₀₀ of 0.05 / ml and cultured as given in section IB. Serial dilutions of bacterial aliquots from the initial inoculums were plated on selective plates-G27MA WT-streptomycin, *G27MAΔhcpC*-erythromycin, *G27MAΔhcpG*-chloramphenicol, *G27MA-ΔhcpC: ΔhcpG* - erythromycin and chloramphenicol to obtain 0 hr bacteria counts. At time intervals 12 hr and 56 hr, bacterial aliquots from the cultures were taken, serial dilutions made and plated on selective plates (streptomycin for WT, erythromycin, chloramphenicol, erythromycin and chloramphenicol depending on the mutant tested for fitness) for CFUs. For example, when fitness of *G27MAΔhcpC* strain was tested against G27MA WT, at each time interval serial dilutions from the bacterial culture were plated on both erythromycin plates and streptomycin plates. BHI plates were incubated for 4-5 days as described in section IB. At the end of incubation, colonies were counted and the number of bacteria was expressed as the number of CFU / ml. The competitive Index (CI) was then calculated as the ratio of mutant CFU / ml to wild type CFU / ml at each time interval / ratio of mutant CFU / ml to wild type CFU / ml in the inoculums. CI value less than 1 indicates that wild type is favored over the mutant; a value more than 1 indicates that mutant is favored over the wild type. The CI is equivalent to fitness (F) of a given

bacterial strain relative to the WT [121]. The median reduction in fitness (S) was computed as $(1 - CI)$. The experiment was repeated three more times and statistical significance between groups was calculated using a Student's *t*-test.

B. Growth Kinetics and Fitness Assays in AGS cell culture infection model

i) Growth dynamics of engineered HpG27MA derivatives in AGS infection model:

AGS cells were seeded in 6-well plates at a density of 10^5 cells / well, 24hrs before infection as described in section VC. Each well was infected with a single desired bacterial strain (*G27MA* WT, or *G27MAΔhcpC*, or *G27MAΔhcpG*, or *G27MAΔhcpC:ΔhcpG*, or *G27MAΔhcpC*) at MOI of 100. To synchronize the infection, the plates were centrifuged for 10 min at 1000 X g using a SORVALL Legend RT centrifuge and incubated at 37°C, 5% CO₂. Serial dilutions of bacterial aliquots from the initial inoculums were plated on selective antibiotic resistant plates as described above to obtain 0 hr bacteria counts. At the end of time intervals 6 hr and 24 hr, the whole well was scraped gently and mixed well. Serial dilutions were prepared and aliquots were then plated on selective antibiotic resistant BHI plates for counting. The number of bacteria was expressed as CFU / ml. Each experiment was repeated three more times and statistical significance between groups (WT and mutant) was calculated using a Student's *t*-test. (Note: CFUs of each growth dynamics experiment performed in this study are given in the Appendix).

ii) Competition and Relative Fitness Assays with HpG27MA derivatives and G27MA WT in AGS infection model:

To assess the relative fitness of G27MA WT and each of the *slr* mutants, (*hcpC* mutant or *hcpG* mutant or *hcpC*, *hcpG* double mutant or *hepC* mutant) in an AGS cell infection, each of the mutant was competed with the WT. The experimental strategy used in VIIB(i) was used with the only exception being that each AGS well was infected with the two bacterial strains. For example when the fitness of *G27MA-hepCΔ* strain was tested against G27MA WT, each AGS cell well was infected with both the strains with a MOI of 100 each. At the end of each time interval, 6 hr and 24 hr, serial dilutions from the bacterial culture were plated on both chloramphenicol (for selecting *hepC* mutant) and streptomycin (for selecting WT) plates. Plates were incubated and CI was calculated as described in VIIA(ii). Each experiment was repeated three more times and statistical significance between groups (WT and mutant) was calculated using a Student's *t*-test.

VIII. Reverse-transcription PCRs.

A. RNA extraction, purification and preparation for RT-PCR analyses

For RNA extraction, cells were lysed in buffer RLT (Qiagen), homogenized and purified using RNeasy mini kit (Qiagen) following manufacturer's instructions. RNA was dissolved in 30 μ l of RNAase free water and checked for integrity using Agilent Bioanalyzer 2100. Determining RNA integrity is a critical step in gene expression analysis. Thus we determined RNA integrity and quality using three different measures, which included electropherogram analysis, determination of 16S:23S rRNA ratio and the RNA integrity number (RIN). RIN is a powerful new tool and is expressed as a scale that

ranges from 1 – 10. A RIN value closer to 10 indicates greater RNA integrity, and thus potentially high reproducibility for high throughput gene expression analysis (eg., microarrays). These analyses showed that the extracted RNA had RIN numbers that ranged from 7.5 to 9.2 (Appendix V). cDNA was then synthesized using the purified total RNA and the cDNA synthesis kit (Qiagen, USA), and used as template for reverse-transcription PCR and real-time PCR. Briefly, DNA was removed from 500 ng-1 µg RNA by adding 2 µl of 7X gDNA wipe out buffer, and incubated at 42°C for 10 minutes. Then, 1 µl of Quantitect Reverse transcriptase, 1 µl of random primer mix or desired primers (1 µM), 4 µl of 5X Quantiscript RT Buffer was added and incubated at 42°C for 15 min. Reverse transcriptase was then inactivated by incubating at 95°C for 3 min. The generated cDNA was used immediately or stored at -20°C.

B. Qualitative Assays for detecting *slr* gene transcripts, *in vitro* and in AGS cell culture infection model

To detect expression of *hcpG* in genetically diverse *H. pylori* strains: J99 (European), 26695 (European), JS7 (Japanese), G27MA (cell-culture adapted), K17 (Korean), R7 (South African), Pecan32 (Peruvian), and S46 (Spanish) *in vitro*, bacterial growth was collected for each strain grown on BHI agar plate; RNA was then extracted and converted to cDNA as described in section VIII A. A set of primers were designed in the internal region of *hcpG* (*hcpGF*-cDNA and *hcpGR*-cDNA), which is conserved among all the strains tested, and which amplified a 250 bp product. As a positive control, a primer set was designed within the *recA* gene (*recAF* and *recAR*) of *H. pylori*, amplifying a 206 bp product. To detect expression of *hcpG* in diverse *H. pylori* strains

after an AGS cell culture infection, AGS cells were infected with the indicated *H. pylori* strains as described in section VB. 6 hr post infection, RNA was extracted as described before and converted to cDNA. As a positive control for AGS infection, a primer set (gapdhF and gapdhR) was designed in the internal region of *gapdh* gene amplifying a 278 bp product. Since genomic DNA contamination of cDNA can produce false positive results, RNA alone was used as a template in the PCR reactions as a negative control. Using the designed primers PCR reactions were performed as described in section IIB(i) using cDNA as template. To detect the expression of *hepC* in *H. pylori* strain G27MA, a set of primers were designed in the internal region of *hepC* (*hepCF*-cDNA and *hepCR*-cDNA) amplifying a 150bp product using procedures described above.

C. Quantitative-RT PCRs for determining *slr* gene expression levels *in vitro* and in AGS cell culture infection models

To quantify the transcript expression level of *slr* genes in this study grown *in vitro* and in AGS cell culture infections, experiments were set up as desired and cDNA was synthesized as described above section VIII B. Since comparison of gene expression in real-time PCR assay assumes that the all the primer sets used in the analysis have equal amplification efficiency, all the primer sets were designed using ABI primer express software version 3.0, so that they should have the same amplification efficiency. Primers *jhp1024F* and *jhp1024R*, *jhp1437F* and *jhp1437R*, *jhp468F* and *jhp468R* were used to amplify *hcpC*, *hcpG*, and *hepC* homologs, respectively. The generated cDNA was diluted fivefold with RNase free water. All real time PCR reactions were performed in 20 μ l mixture containing 1 / 10 volume of cDNA preparation (2 μ l of 1:5 diluted cDNA), 1X

Power SYBR green PCR master mix (10 μ l) (Applied Biosystems, Foster City, CA, USA), 5 μ M of each forward and reverse primer (1 μ l) and making up the reaction volume to 20 μ l with water. RNA quantitations were performed in Applied Biosystems Step-One thermocycler using the following PCR conditions: 95°C for 10 min followed by 95°C for 15 s and 60°C for 1 min for 40 cycles. Appropriate no-RT and non-template controls were included in each reaction plate and melting curve analysis was performed at the end of each run to confirm the specificity of the reaction. The concentration was determined by the comparative C_T method (threshold cycle number at the cross-point between amplification plot and threshold) and values were normalized to *ureA/recA*. Negative and positive values were considered as down-regulation or up-regulation of expression of genes of interest, respectively, represented by a minimum of two-fold difference.

D. Pathway-specific PCR arrays

The Human cytoskeleton regulators pathway-specific PCR arrays (PAHS-088A) developed by SABiosciences permits simultaneous interrogation of gene expression of 84 genes controlling the intracellular scaffolding's biogenesis, organization, polymerization, and depolymerization.

AGS cells were infected with a representative European (J99, and G27MA), Japanese (JS7), Amerindian (SHI470) *H. pylori* strains, for 3 hrs and 6 hrs. Next, RNA was extracted from AGS cells, equal amount (500 ng) of RNA was used in all the arrays to convert to cDNA as described above and quantitative PCRs were conducted using the pathway-specific PCR array with appropriate controls for human DNA contamination

(*higX1A*), housekeeping genes (β 2M, *hgp*, *rpl13A*, *g3pd* and β -actin), and multiple positive and negative controls (supplied by the manufacturer & given in the appendix). The expression of cytoskeletal regulators was then compared with an uninfected control to identify genes that were differentially up or down regulated by $\Delta\Delta C_t$ method, following infection with diverse *H. pylori* strains. The analysis was performed through an integrated web-based software package for the PCR Array System which performs all quality checks and $\Delta\Delta C_t$ based fold-change calculations from the uploaded raw threshold cycle data obtained for each condition. Fold up or down regulations were depicted as scatter plots with a cut off of two fold change.

IX. Fluorescence Activated Cell Sorting Analyses

A. Determination of Variation in Heat Shock Protein B (HspB) surface expression

by *H. pylori* G27MAWT and $\Delta hcpC$ and $\Delta hcpG$ derivatives *in vitro*

To detect the variation in HspB surface expression among G27MA WT and *hcp* mutants grown *in vitro*, indicated strains were grown in BHI broth as described in section IB. BHI broths inoculated with *hcpC* mutant, *hcpG* mutant, and *hcpC*, *hcpG* double mutant contained 10 μ g / ml of erythromycin, chloramphenicol, and erythromycin with chloramphenicol, respectively, to provide selection. At 3 hr, 24 hr, and 56 hr post infection, 3 ml of bacterial cultures were aliquoted, spun at 3500 rpm for 10 min and washed twice with 1X PBS (pH:7.4). Cells collected were immediately fixed using 4% paraformaldehyde at room temperature for 15 min. The cells were then washed twice with 1X PBS and stored immediately at 4^oC. Once the cells were fixed, all the washing steps were done at a speed not more than 2000 rpm until a cell pellet is visible (or for a

maximum of 10 min). After the cell pellets from all the desired time points were obtained, surface staining of HspB was conducted by skipping the cell permeabilization step. Anti-HSP primary antibody was then added to the cell suspension at a final concentration of 1 $\mu\text{g} / \text{ml}$ and incubated on ice for 30 min, followed by washing twice with wash buffer (1X PBS + 0.2% BSA). Secondary antibody conjugated with texas red (TR) flourophore was then added to cell suspension at a final concentration of 0.5 $\mu\text{g} / \text{ml}$ and incubated for 30 min on ice, in dark. Cells were then washed twice with 1X wash buffer and analyzed in a BD FACS CaliburTM Flow Cytometer (BD Biosciences) using channel FL4 for detecting emitted fluorescence from 10^4 cells / sample. Unstained bacteria were used to optimize the scatter (Forward scatter-FSC and Side scatter-SSC) and to establish back ground fluorescence. After obtaining mean fluorescence for each sample, the data was analyzed in WinMDI software version 2.9 to obtain mean geometric fluorescence values and depicting fluorescent intensities as histograms or dot plots. Parameters used in the detection of fluorescence were given in Table 1.

B. Determination of Variation in Heat Shock Protein B (HspB) surface expression by *H. pylori* G27MAWT and $\Delta hcpC$ and $\Delta hcpG$ derivatives in AGS cell culture infection model

After infecting AGS cells (as described in section VB) with either the *hcpC* mutant or *hcpG* mutant or *hcpC*, *hcpG* double mutant for 3hr, 6hr, 12hr, and 24hr, AGS cells were washed twice with 1X PBS to remove unattached bacteria and detached using 2mM EDTA. Cells collected were immediately fixed using 4% paraformaldehyde at room temperature for 15 min. Once the cells from all the desired time points were

obtained, surface staining of HspB was conducted and analyzed as described above in section IXA. Unstained AGS cells with adhered bacteria were used to optimize the scatter and to establish back ground fluorescence. The experiment was done in triplicate and a student *t*-test was used to calculate statistical significance between experimental groups.

C. Determination of CagA, VASP and activated MAPK expression analyses in the context of AGS infection model

To ensure the specificity of HspB dynamics in *hcp* mutants and *G27MA WT H. pylori* strain, expression dynamics of CagA, VASP, and activated MAPK were monitored as controls. CagA expression is used as a control for equal bacterial inoculums. Increase in host cell VASP expression and activation of MAPK are hallmarks of *H. pylori* infection. Thus, they were used as controls for monitoring host cell dynamics. CagA, VASP, and activated MAPK antibodies used in this study were raised against three different hosts: anti goat, anti mouse, and anti rabbit, respectively. After infecting AGS cells (as described in section VB) with either the *hcpC* mutant or *hcpG* mutant or *hcpChcpG* double mutant for 3 hr, 6 hr, 12 hr, and 24 hr, AGS cells were washed twice with 1X PBS to remove unattached bacteria and detached using 2 mM EDTA. Cells collected were immediately fixed using 4% paraformaldehyde at room temperature for 15 min. Once the cells from all the desired time points were obtained, they were permeabilized using 0.01% Triton-X (Sigma Aldrich, MO) for 3 min at room temperature. Cells were immediately washed twice with 1X wash buffer. The three primary antibodies were then added to each of the tubes containing permeabilized AGS cells at a final concentration of 1 µg/ml and incubated on ice for 30 min, followed by

washing twice with wash buffer (1X PBS + 0.2% BSA). Secondary antibodies conjugated with either texas red (anti VASP) or phycoerythrin (PE) (anti-activated MAPK), or FITC (anti-CagA) flourophores were then added to cell suspension at a final concentration of 0.5 µg / ml and incubated for 30 min on ice, in dark. Cells were then washed twice with 1X wash buffer and analyzed in a BD FACS Calibur Flow Cytometer (BD Biosciences) using channel FL1 (FITC), channel FL2 (PE) and channel FL4 (TR) for detecting emitted fluorescence from 10⁴ cells/ sample. Unstained AGS cells with adhered bacteria were used to optimize the scatter and to establish back ground fluorescence. Data was then analyzed in WinMDI 2.9 as described above.

Table 1: Parameters for FACS analysis.

	α-HSP	α-CagA/VASP/pSer MAPK
FSC	E00	E00
SSC	387	408
FL1	412	364
FL2	150	400
FL3	150	150
FL4	150	819

FSC - forward scatter, SSC – side scatter, FL1, FL2, FL3, FL4 – fluorescent channels 1, 2, 3, and 4, respectively to detect fluorescence emitted from FITC, PE, and TR conjugated flourophores, respectively.

Table 2: Primers used in the study						
Primer #	Primer name	Sequence 5'-3'	Location	Product size	Intended use (added sequences and restriction sites italicized and underlined)	Reference
Primers used in <i>hpcC/hpcG</i> part of the study						
1	hpcCF1	5'CGCCTTCGCCATTGTATTGCATA	jhp1023: 747-769	1643bp	Amplification of <i>hpcC</i> with flanking gene fragments	This study
2	hpcCR1	5'GCCCCGGGCTTACGATAAAGCTTTTAGAACA	jhp1025: 269-299			This study
3	hpcGF1	5'CATTGTGGATATTGTGGG	jhp1436: 274-293	988bp	Amplification of <i>hpcG</i> with flanking gene fragments	This study
4	hpcGR1	5'AGGACAAAGGGTTTGT	jhp1438: 26-42			This study
5	1098 Fseq	5'GGGTGTTTTAATTTAGGGGTGCT	jhp1024: 200-223	650bp	Amplification of internal fragment of <i>hpcC</i>	This study
6	hpcGF2	5'TCAAAAAAGCGGTTTTAGGG	jhp1437: 11-31	750bp	Amplification of <i>hpcG</i> internal fragment	This study
7	hpcGR2	5'TTGTCTTTGAGAATATCCG	jhp1437: 743-762			This study
8	1098A1n	5'-CGCCTTCGCCATTGTATTGCATA	hp1098: 747-769	509bp	Amplification of fragment A for <i>hpcC</i> knockout	This study
9	1098A2n	5'- <u><i>GTGTTAATCCATAGTATAAAGCAT</i></u> CCAGCAATGGGTGTCATTTGCTAGGGA	hp1098: 288-313			This study
10	rpsl-F	5'-GATGCTTTATAAAGTATGGATTAAACAC	rpsl-erm: 1-27	1539bp	Amplification of <i>rpsl-erm</i> cassette	[115]
11	erm-R	5'-TTACTTATTAATAAATTTATAGCTATTGAA	rpsl-erm: 1510-1539			[115]
12	1098C5n	5'- <u><i>TC AATAGCTATAAATTAATTAATAAGTAA</i></u> CAAGCACAAAACCTAAAAAAGGACT	hp1098: 20-47	500bp	Amplification of fragment C for <i>hpcC</i> knock out	This study
13	1098C6n	5'-GCCCGGGCTTACGATAAAGCTTTAGAACA	hp1099: 269-299			This study
14	hpcGAF	5'-GAAAGTCTGTATTATAAAGGATT	jhp1435: 987-1008	700bp	Amplification of fragment A for <i>hpcG</i> knockout	This study
15	hpcGAR2	5'- <u><i>GATCGAATTC</i></u> CCGACACTCTCCTCCCTTTGTAA	J99: 1584828-1584852			This study
16	rpslF2	5'- <u><i>GATCGAATTC</i></u> GGATGCTTTATAACTATAACTATGATT	rpsl-cat: 1-26	1500bp	Amplification of fragment B (<i>rpsl-cat</i>) for <i>hpcG</i> knockout	This study
17	catR2	5'- <u><i>GATCGCGGCGC</i></u> TTATCAGTGGCAGAACTGGG	rpsl-cat: 1395-1420			This study
18	hpcGCF2	5'- <u><i>GATCGCGGCGC</i></u> AAATCGTAGCCCTATTAAAGCCCAT	J99: 1585624-1585648	750bp	Amplification of fragment C for <i>hpcG</i> knockout	This study
19	hpcGCR	5'-TATTTAAAAGCGTGGGGCTTA	jhp1438: 716-737			This study
20	hpcGBF	5'ATGTTAAGGGGTGTCAAAAAAGC	hpg27: 1469: 1-23	984bp	Amplification of hpcG-His	This study
21	hpcGBR	<u><i>ATGATGGTGGATGAT</i></u> GTTATTGTTTTGTCTTTGAG	hpg27: 1469: 960-984			This study
22	hpcCF-c DNA	5'-TGGCAGAGCAAAGCCCTAAAAGA	hp1098: 782-803	245bp	Amplification of fragment of <i>hpcC</i> for performing RT-PCR	This study
23	hpcCR-c DNA	5'-ATTCCTAGCAATACACC	hp1098: 559-578			This study
24	hpcGF-c DNA	5'-AGAAGGCTGGGCTCATTTG	jhp1437: 152-172	247bp	Amplification of fragment of <i>hpcG</i> for performing RT-PCR	This study
25	hpcGR-c DNA	5'-TACAAGATTTAGAGTAATATT	jhp1437: 377-398			This study
26	jhp1024F	5'-GTGCCTCTGCCTGCATCAT	jhp1024: 378-396	109bp	Amplification of fragment of <i>hpcC</i> for performing qRT-PCR	This study
27	jhp1024R	5'-TTTGAACGATGGCGATGGT	jhp1024: 468-487			This study
28	jhp1437F	5'-AGGCGTGGCAAAGGATGA	jhp1437: 668-690	109bp	Amplification of fragment of <i>hpcG</i> for performing qRT-PCR	This study
29	jhp1437R	5'-AATCCTAGCTTGCAACTCTTTCAA	jhp1437: 756-777			This study
30	groelF	5'-TGTGGTTAAAAGCAGGCTTAAA	groelF:99-121	102bp	Amplification of fragment of <i>H. pylori groel</i> for performing qRT-PCR	This study
31	groelR	5'-GGCATTATTGACCCCTTAAAAGTAAA	groelR:179-201			This study
Primers used in <i>hpcC</i> part of the study						
32	jhp468F	5'CCAAGCCCGATGCTGTATTC	jhp468: 528 -548	89bp	Amplification of fragment of <i>hpcC</i> for performing RT-PCR	This study
33	jhp468R	5'GCCAATACAAGCAAGCCCTTAAA	jhp468: 594 -617			This study
House keeping gene primers used in the study						
34	ureA F	5'-ATGAAACTCACCCTAAAAGAGTT	ureA: 695-771	271bp	Amplification of fragment of <i>ureA</i> for performing RT-PCR	This study
35	ureA R	5'-CACATCATCCGGTTTTAAAAGA	ureA: 501-529			This study
36	recA F	5'-GCAATAGATGAAGACAAAACA	recA: 23-4	206bp	Amplification of fragment of <i>recA</i> for performing RT-PCR	This study
37	recA R	5'-GACTCTGCCCATAAATTTCA	recA: 180-209			This study
38	GAPDH F	5'-AGAAAGGCTGGGCTCATTTG	GAPDH: 369-389	278bp	Amplification of fragment of <i>GAPDH</i> for performing RT-PCR	This study
39	GAPDH R	5'-AGGGCCATCCACAGTCTTC	GAPDH: 626-646			This study
40	recAF	5'-TGCGAATGCCAAAAAATGG	recA: 924-944	100bp	Amplification of fragment of <i>recA</i> for performing qRT-PCR	This study
41	recAR	5'-AAGGCATGCTCAGCGTCAA	recA: 1004-1024			This study
42	ureAF	5'-CCATCCATCACATCATCTGGTT	ureA: 502-523	104bp	Amplification of fragment of <i>ureA</i> for performing qRT-PCR	This study
43	ureAR	5'-TGGGCTGAATTGATGCA	ureA: 587-606			This study
44	ureBF	5'-GTTCGCATCACCCATTTGACT	ureB:318-339	110bp	Amplification of fragment of <i>ureB</i> for performing qRT-PCR	This study
45	ureBR	5'-GCGTGAACCTAACATGATCATCA	ureB:408-428			This study

RESULTS

I. Evolution of Stable Genetic, and Functionally Non-Reciprocal Redundancy Driven by Positive Selection in Duplicated Sel-1 like Genes of *H. pylori*.

***H. pylori* strain – specific *slr* genes**

To determine the distribution and genetic organization of *H. pylori* strain – specific *slr* genes, *hcpC*, and *hcpG* in the available complete genomes of *H. pylori*, we conducted a BLAST search using the representative homologs in *H. pylori* strain J99. These analyses revealed that while *hcpC* was present in all the available *H. pylori* genomes, only four of ten harbored *hcpG* homologs (Table 3). The prototype HcpC from *H. pylori* strain 26695 contained 7 Sel-1 domains with a predicted signal peptide, whereas encoded proteins of *hcpG* homologs in the available *H. pylori* genomes demonstrated a much more variable domain architecture with alleles containing 4-7 Sel-1 domains and with a predicted signal peptide (Fig. 5). When compared to the other available *H. pylori* genomes, genomic location of *hcpC* is not conserved in *H. pylori* genome 26695 (Fig 6). Investigation of the genes upstream of *hcpC* in available *H. pylori* genomes revealed that, except in the genome of Hp26695, the flanking gene (*hp1097*; *short chain alcohol dehydrogenase*) was conserved. Additionally, two genes encoding for transposases, *tnpA* (*hp1096*) and *tnpB* (at *hp1095*) were identified 1 kb to 3 kbs upstream of *hcpC*, in the genome of strain 26695. It is likely that these transposases encode a functional two-element insertion sequence of the IS605 family, several of which are

known to be present in *H. pylori* genomes [26] and, which could potentially account for the genetic rearrangement observed in the genome of strain 26695 at the *hcpC* locus. This also suggested that although the *hp1097* homolog may not be in the same chromosomal location with respect to *hp1098* (*hcpC*), it might not have been lost altogether from the 26695 genome. Therefore, I conducted a BLAST (*blastn*) search with *jhp1023* (J99 homolog of *short chain alcohol dehydrogenase*) sequence against the 26695 genome. This analysis revealed that, *hp0357* was homologous to *jhp1023* and also confirmed that the *hcpC* locus had indeed been subject to genetic rearrangements in strain 26695.

Table 3: Strain specific distribution of *hcpG* in the available *H. pylori* genomes.

Genome	<i>hcpC</i>	<i>hcpG</i>
J99	<i>jhp_1024</i>	<i>jhp_1437</i>
26695	<i>hp1098</i>	-
G27	<i>hpg27_1039</i>	<i>hpg27_1469</i>
HPAG	<i>hpag1_1036</i>	-
SHI470	<i>hpsh_05640</i>	-
P12	<i>hpp12_1063</i>	-
B8	<i>hpb8_406</i>	<i>hpb8_1690</i>
HP51	<i>khp_0998</i>	-
HP52	<i>hpkb_1028</i>	-
HPV225d	<i>hpv225_1116</i>	<i>hpv225_1471</i>

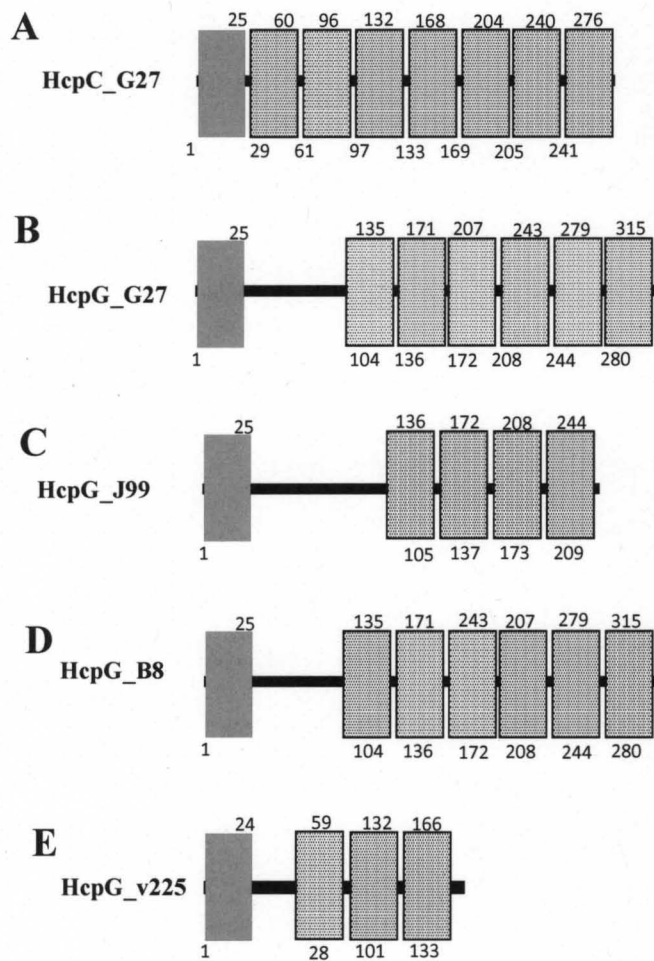


Figure 5: SLR domain architecture in encoded proteins of *H. pylori* *slr* genes, *hcpC* and *hcpG* in the available *H. pylori* genomes.

The schematic here shows predicted SLR domains (dotted box), signal peptide motifs (grey box), and their amino acid positions. HcpC domain architecture is identical in all the available *H. pylori* genomes, whereas *hcpG* variants encode proteins with different domain architectures

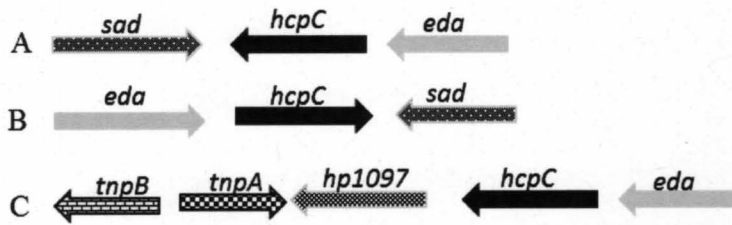


Figure 6: Genetic organization of *hcpC* in the available *H. pylori* genomes.

hcpC is flanked by 2-keto-3-deoxy-6-phosphogluconate aldolase (*eda*) and short chain alcohol dehydrogenase (*sad*) in all the available genomes except *H. pylori* 26695, where it is flanked by *eda* and a hypothetical protein (*hp1097*). Type A is seen in *H. pylori* genomes J99, P12, G27, HPAG, SHI470, 51, 52 and V225d; type B in B8 and type C in 26695. Genes encoding for transposases (*tnpA* and *tnpB*) are found 1kb-3kb downstream *hcpC* in 26695, which could possibly explain the genetic rearrangement seen.

Genetic rearrangement at *hcpC* locus is unique to *H. pylori* strain 26695

Next, I asked if the genetic rearrangement with respect to the *hcpC* locus was unique to strain 26695 or was also present in the genomes of other diverse *H. pylori* strains. To determine this, two primers (primers 1 and 2 in Table 2) designed in genes flanking *jhp1024* (J99 homolog of *hcpC*) were used to PCR amplify *hcpC* homologs in genetically diverse *H. pylori* strains. 166 out of 166 *H. pylori* isolates tested variously from Africa (Gambia and South Africa), East Asia (Japan and Korea), India, Peru, Spain, and Shima were positive for *hcpC* locus specific PCR amplification with products of nearly the same size as expected from HpJ99 (1643 bp). This outcome indicated that the *hcpC* locus rearrangement seen in strain Hp26695 genome was not common in *H. pylori* strains.

Strain specific distribution of *H. pylori* *slr* gene, *hcpG*

Next, I sought to determine the chromosomal organization and distribution of *H. pylori* strain – specific *slr* gene, *hcpG* in genetically diverse *H. pylori* isolates. To determine this, I used primers (primers 3 and 4 in Table 2) designed in the genes flanking *jhp1437* (J99 homolog of *hcpG*) to PCR amplify *hcpG* homologs in a worldwide collection of 166 *H. pylori* strains. I found that 47% (79 / 166) of the isolates tested were positive for PCR amplification, although with random size variations deviating from the expected amplicon size (988 bp), when viewed on a 1% agarose gel (Fig. 7A). Next, I asked if the chromosomal location of *hcpG* was rearranged in the strains that appear to lack *hcpG* (87 / 166). If this were true, it could explain the absence of *hcpG* in these strains tested for its presence using primers designed in the flanking genes. To determine this, I designed primers (primers 6 and 7 in Table 2) in the conserved internal region of *hcpG* homologs and used them to PCR amplify *hcpG* internal fragment in *H. pylori* strains that resulted in no amplification using flanking primers. My results indicated that *hcpG* was indeed absent in 53% (87 / 166) of *H. pylori* isolates tested.

Unique DNA sequence polymorphisms and pseudogenization of *hcpG*

Next, I asked if the size variations seen in *hcpG* homologs affect the open reading frame (ORF) of their respective encoded proteins. To determine this, full-length PCR products of *hcpG* homologs were sequenced using primers from either end. *hcpG* DNA sequence analysis revealed the following key features: 1) Encoded ORFs of 36% (28 / 79) of *hcpG* homologs were interrupted by premature stop codons forming truncated proteins (< 100 amino acids) or no protein altogether. Premature stop codons resulted

mostly from a frame shift of the ORF that originated several base pairs upstream of the stop codon (Fig 7B), and also due to base substitutions in a few homologs. 2) In *hcpG* homologs with uninterrupted ORFs (> 100 amino acids), striking *indel* (insertions and deletions of nucleotides) polymorphisms were observed (Fig. 8). Taken together, I identified striking genetic variation in *hcpG* homologs, among diverse *H. pylori* isolates. This finding was in accord with a previous study where it was shown that *hcpG* (designated as *hsp12* in the study) exhibited high genetic variation [122].

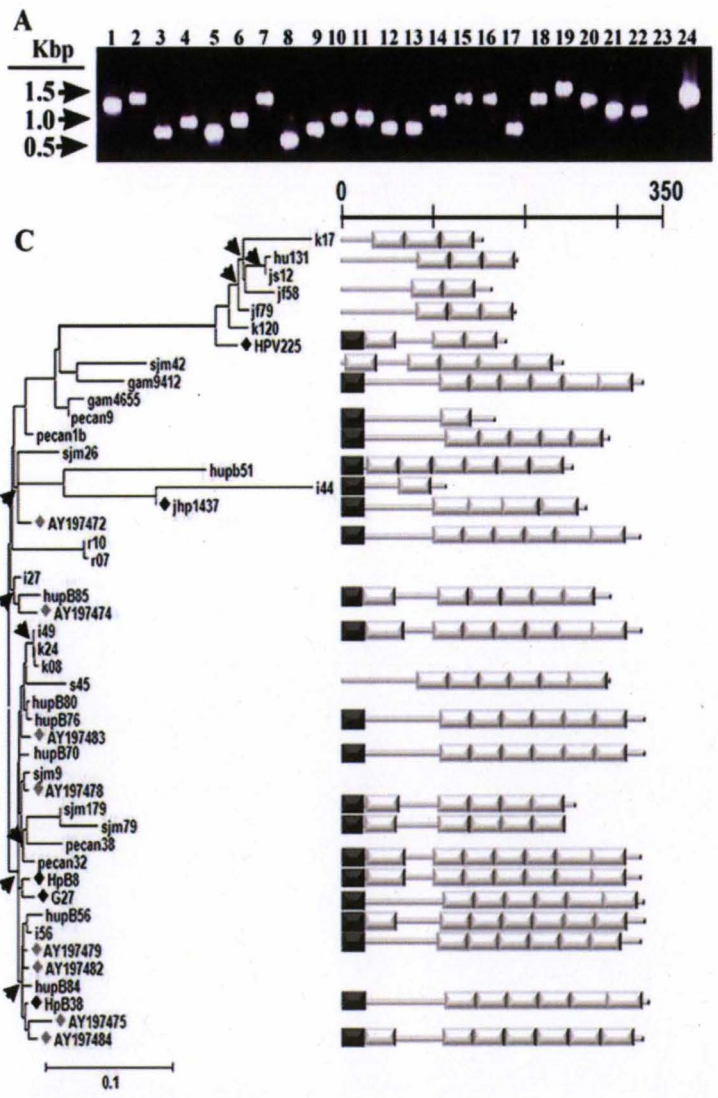


Figure 7: *hcpG* rapidly evolves in diverse *H. pylori* isolates.

(A) Size variations of *hcpG* homologs in diverse *H. pylori* isolates.

Depicted here is a 1% agarose gel with *hcpG* homologs amplified from Gambia (1 - 3), Japan (4 - 12), Korea (13 - 19), and Spain (20 - 23). Lane 24 harbors *hcpG* homolog amplified from *H. pylori* strain J99.

(B) Pseudogenization in *hcpG* homologs, as a result of frame shift mutations.

Chromatograms shown here depict nucleotide sequences of two different *hcpG* homologs amplified from Japanese *H. pylori* strains *Hu131* and *Hu56*. Single base deletion (T-indicated by a star) in *Hu56* resulted in a frame shift of *Hu56 hcpG* homolog, leading to the formation a premature stop codon (TAA), 17 bp downstream of the single base pair deletion. Such frame shift resulted in pseudogenization of *hcpG* homolog in *H. pylori* strain, *Hu56*.

(C) ML phylogeny of HcpG (n = 46) with random geographical distribution of HcpG Sel1 domain variants.

The phylogeny was reconstructed assuming the TVM + G substitution model and was optimized to the following parameters using heuristic searches and tree bisection-reconnection algorithm: rate matrix: $A \leftrightarrow C = 1.8144$, $A \leftrightarrow G = 8.4557$, $A \leftrightarrow T = 0.4947$, $C \leftrightarrow G = 1.2924$, $C \leftrightarrow T = 8.4557$, and $G \leftrightarrow T = 1$; base frequencies: A=0.3623, C=0.1295, G=0.2197, T=0.2885, proportion of invariable sites, I = 0 and gamma distribution shape parameter = 0.3955. Phylogeny is rooted using the outgroup method implemented in PAUP. Bar scale = 0.1 nucleotide substitutions per site. Lineages under positive selection were indicated with arrow heads.

(D) Maximum-likelihood parameters of selection pressures acting on *H. pylori* HcpG codons.

InL, Log-likelihood score; dN/dS, rate ratio of non-synonymous to synonymous changes averaged over all sites; LRT, likelihood ratio test.

(E) Adaptive evolution in HcpG.

Frequency distribution of three codon classes (k1, k2, and k3) and their associated d_N/d_S ratios were computed under the SSM M3 for HcpG. 11 Codons under positive selection (codon class k3) are shown.

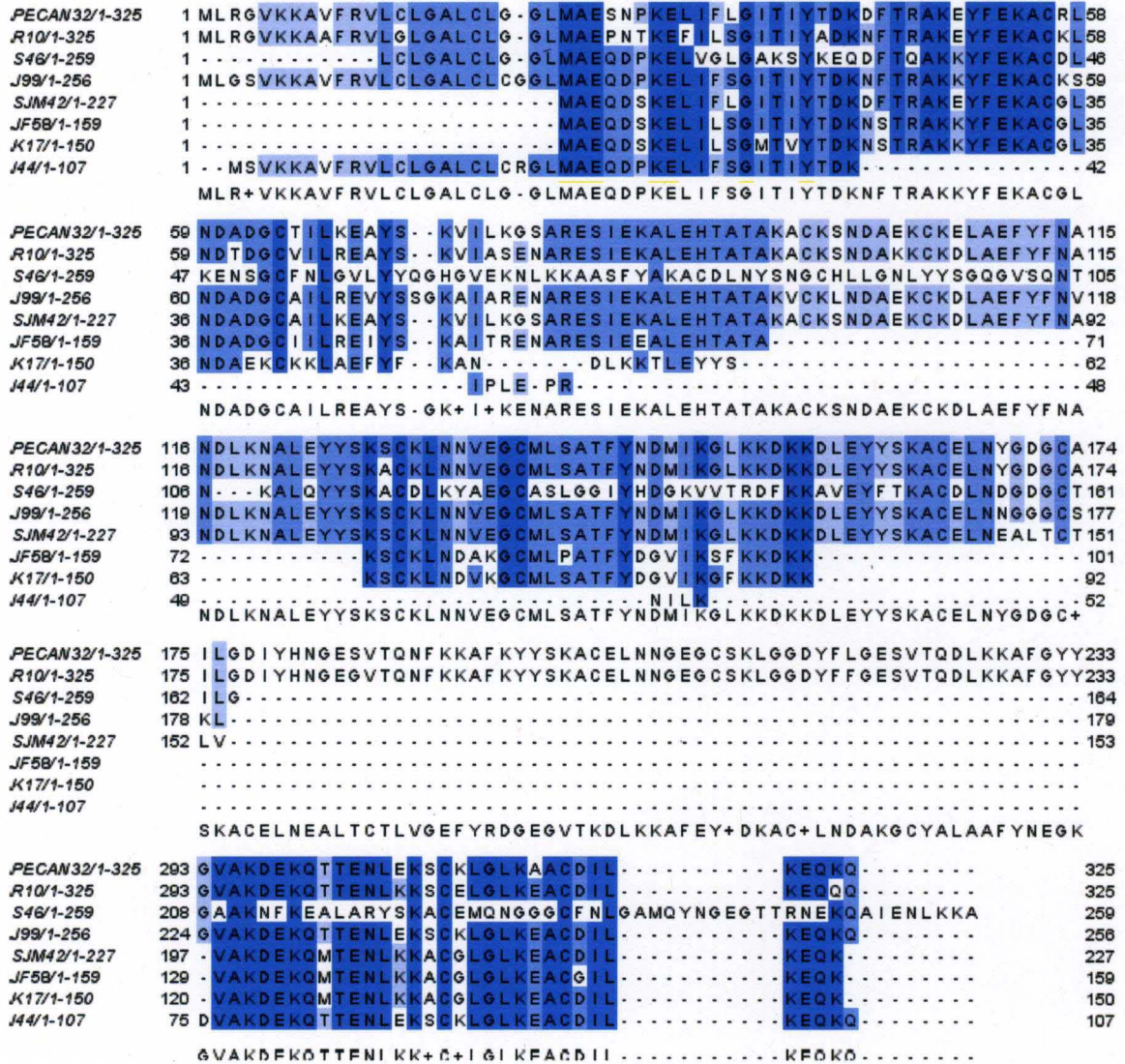


Figure 8: *Indels* pattern seen among HcpG homologs in diverse *H. pylori* isolates.

Alignment shown here is generated in Jalview 2.5.1 (ClustalW algorithm) using HcpG homologs from representative populations-Peru (PECAN32), South Africa (R10), Spain (S46, SJM42), Japan (JF58), Korea (K17), India (I44), and European (J99). Deletions of amino acids can be seen as dotted lines throughout the alignment. Alignment is colored based on percentage identity.

Variations in the number and distribution of Sel-1 domains in uninterrupted *hcpG* ORFs change the tertiary structure of HcpG variants.

Next, I tested if the domain architecture of the encoded proteins was affected as a result of *indels* in *hcpG* homologs, using SMART prediction tool. This analysis revealed that *indels* in uninterrupted ORFs resulted in variations in the number (1 – 7) and distribution of Sel-1 domains in each encoded proteins with a random geographic association in the distribution of HcpG variants (Fig. 7C). Next, I asked if the encoded protein's tertiary structure was affected due to variations in the number and distribution of Sel-1 domains among diverse HcpG homologs. Comparative protein structure modeling of HcpG variants using the related HcpB and HcpC crystal structures as templates suggested that the tertiary structure of HcpG variants likely differed from each other (Fig. 9). This suggests that the activity or functions of HcpG variants likely differ from each other in subtle or perhaps dramatic manners.

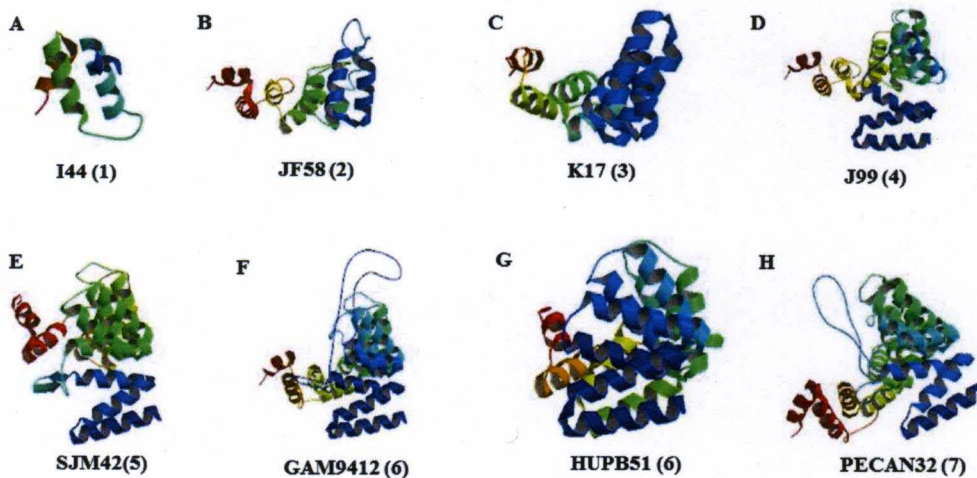


Figure 9: HcpG variants differ in their predicted tertiary structures.

Tertiary structures of HcpG variants from representative populations, varying in the number and distribution of Sel1 domains are depicted here. Numbers in the parenthesis indicate the number of Sel1 domains confidently predicted by SMART analysis. Tertiary structure modeling of HcpG variants was performed using Modeller 9v8; version modweb using the available HcpB and HcpC crystal structures as template.

***hcpG* rapidly evolves in diverse *H. pylori* isolates**

To better understand *hcpG* evolutionary dynamics, a maximum-likelihood (ML) phylogeny was reconstructed using the sequence data. Selection pressures on HcpG individual codons and branches of its phylogenetic tree were next studied in detail using two groups of codon-based models of sequence evolution and ML-based LRTs: 1) Site-specific models [SSMs], which examine variation in selection pressures across codons and assume a single ω across the phylogeny; and 2) Lineage-specific models (LSMs), which allow ω to vary among lineages, while assuming a single rate across all codons (M1bra). SSMs confidently identified 11 sites under positive selection ($\omega = 3.365$) (Fig 7D & 7E), which suggested different selective pressures at different sites in HcpG. The M1bra model assuming free ω for each branch was next used to assess the overall variation in ω in all *hcpG* lineages. The M1bra model also fit the data significantly better than M0 ($p = 0.002$), which suggested that *hcpG* alleles had been subject to different selective pressures in different populations (Fig 7C). Detailed estimates of model parameters are shown in Table 4.

Table 4: Maximum-likelihood parameters of selection pressures acting on *H. pylori* *hcpG* codons

Model Code	<i>lnL</i> *	Tree-length	κ^{\dagger}	d_N/d_S	Parameter Estimates	D [d.f.] [¶]	χ^2	P-value	Positively selected sites [Reference sequence: Hpl49]
M0 (1-ratio)	-5242.663	3.2215	4.768	0.325	$\omega=0.325$	NA	NA	NA	NA
M1a (Nearly Neutral)	-5130.388	3.4698	4.867	0.341	$p_0=0.742, \omega_0=0.111; p_1=0.258, \omega_1=1$	NA	NA	NA	NA
M2a (Positive Selection)	-5108.087	3.5508	5.228	0.434	$p_0=0.723, \omega_0=0.117; p_1=0.256, \omega_1=1; [p_s]=0.021, [\omega_s]=4.46$ $p_0=0.627, \omega_0=0.083; p_1=0.334, \omega_1=0.674; [p_s]=0.039, [\omega_s]=3.365$	M0 vs M2a [3] M1a vs M2a [2]	269.152 44.602	<0.0001 <0.0001	28S, 30P, 36S, 58K, 66A, 79L, 262A
M3 (Discrete)	-5105.965	3.551	5.125	0.409	$p_0=0.627, \omega_0=0.083; p_1=0.334, \omega_1=0.674; [p_s]=0.039, [\omega_s]=3.365$	M0 vs M3 [4] M1a vs M3 [4]	273.396 44.846	<0.0001 <0.0001	21G, 28S, 30P, 36S, 58K, 66A, 79L, 221L, 224S, 262A, 313K
M7 (β)	-5135.398	3.4751	4.731	0.321	$p=0.323, q=0.684$	NA	NA	NA	NA
M8 ($\beta, \omega_S > 1$)	-5106.781	3.556	5.096	0.405	$p=0.465, q=1.159; p_0=0.961, [p_s]=0.393, [\omega_s]=3.334$	M7 vs M8 [2]	57.234	<0.0001	21G, 28S, 30P, 36S, 58K, 66A, 79L, 221L, 224S, 262A, 272S, 313K
M8a ($\beta, \omega_S = 1$)	-5128.752	3.525	4.659	0.362	$p=1.516, q=9.983; p_0=0.733, [p_s]=0.267, \omega_s=1$	M8a vs M8 [1]	43.942	<0.0001	
M1bra	-5177.115	3.2376	4.649	NA	Free ω for each lineage	M0 vs M1bra	131.1	<0.002	NA

* *lnL*, Log Likelihood Score

\dagger , κ , kappa, ratio of transition to transversions

\ddagger , D, hierarchical Likelihood Ratio Test statistic and d.f., degrees of freedom

Biological relevance of HcpG in *H. pylori* growth and AGS cell infection

Next, I sought to characterize the biological relevance of HcpG in *H. pylori* growth and infection. For this, I first characterized the expression dynamics of HcpG variants *in vitro* and in gastric epithelial cell line (AGS) infection model by qualitative assessment of *hcpG* transcript using Reverse transcription PCR. These experiments revealed that transcripts of representative HcpG variants were detectable *in vitro* (BHI medium) and in AGS cell infection model (Fig 10A and 10B). Using quantitative real time PCR experiments, I next identified that *hcpG* transcript is differentially up regulated by 3hr and 6hr post AGS cell infection, in *H. pylori* strains G27MA and J99 (Fig. 10C). Differential regulation of *hcpG* in different genomic contexts in the AGS infection suggested that *hcpG*'s role in *H. pylori* strains was likely tightly regulated, which in turn indicated that *hcpG* is likely biologically relevant during *H. pylori* infection, at least with certain strains. However, previous studies with *hcpG* (*hsp12*) had concluded that the

deletion of *hcpG* did not affect *H. pylori* growth [122]. To further characterize the function of HcpG, I generated *hcpG* deletion in *H. pylori* strain G27MA (cell culture adapted strain). Using this strain, I sought to determine the role of HcpG in *H. pylori*'s growth *in vitro* and in AGS cell infection model. My initial findings (Fig. 12B) were in accord with the previous study where it was shown that HcpG did not appear essential for growth of *H. pylori in vitro* [122]. However, since human gastric epithelium constitutes the natural niche for *H. pylori*, I examined the role of HcpG in *H. pylori*'s growth and survival 6 hr and 24 hr post infection in the AGS cell-culture infection model. When compared to the G27MA WT, no significant difference ($p > 0.1$) was found in the growth dynamics of *G27MAΔhcpG* strain in the AGS cell infection assay (Fig. 12A). These growth dynamics, whereby the growth of parent and the mutant derivative strain was measured in pure culture, provide a measure of absolute fitness. However, such absolute measures of performance can be unreliable in predicting evolutionary success or failure [123]. Competing various strains (here, WT and mutants) in a given environment and monitoring the fitness of each strain relative to the other identifies strains that can perform “better” when both the strains encounter adverse conditions (e.g., nutrient limitation) [123]. Hence, measures of relative fitness should provide more meaningful insights into the contributions of a gene to bacterial growth. In this context, I conducted competition assays between *G27MAΔhcpG* with *G27MAWT in vitro*, and in AGS cell-culture infection model and determined the relative fitness of *G27MAΔhcpG* by comparing its growth dynamics with *G27MAWT*. The results obtained from these experiments indicated that there was no significant difference ($p > 0.1$) in the relative fitness of *H. pylori* strain *G27MAΔhcpG* when competed with *G27MAWT* (Fig. 10D).

Given that, *hcpG* was only slightly up regulated (~2 fold) in G27MA compared to J99 (~15 fold), it is reasonable to expect that the effect of *hcpG* deletion will likely be more dramatic in J99 *H. pylori* strain. Thus, attempts at engineering J99 strain to generate *hcpG* mutation however, were unsuccessful (other investigators have had similar difficulties in transforming J99; Dr. Douglas Berg, personal communication). For the purpose of ascertaining that HcpG was expressed during infection with HpG27MA, I conducted infection assays using the G27MA strain wherein the native *hcpG* copy was replaced by an *hcpG::6XHis* insertion (See Materials and Methods section IVD). Even though the *hcpG* transcript was moderately up regulated in G27MA (~2 fold), detectable amounts of HcpG were identified during AGS cell-culture infection (10E), suggesting that the role of HcpG in the context of *H. pylori* strain G27MA cannot be completely eliminated. I monitored the expression of control proteins CagA (translocated into host cells via the CAG-PAI encoded Type IV secretion system [124]), and activated Mitogen-Associated Protein Kinase, MAPK (which is a hallmark of *H. pylori* infection; [125] [126]), and tubulin during infection with *H. pylori* G27MA::*hcpG-His*, to ensure that I replicated known *H. pylori* infection dynamics and to ascertain that equal MOI of *H. pylori* strains was used during this experiment (Fig. 10E). Thus, I conclude that detectable amounts of HcpG are produced by *H. pylori* strain G27MA during infection of AGS cells. However, if HcpG is produced during infection then why didn't its deletion from G27MA result in any obvious growth defect?

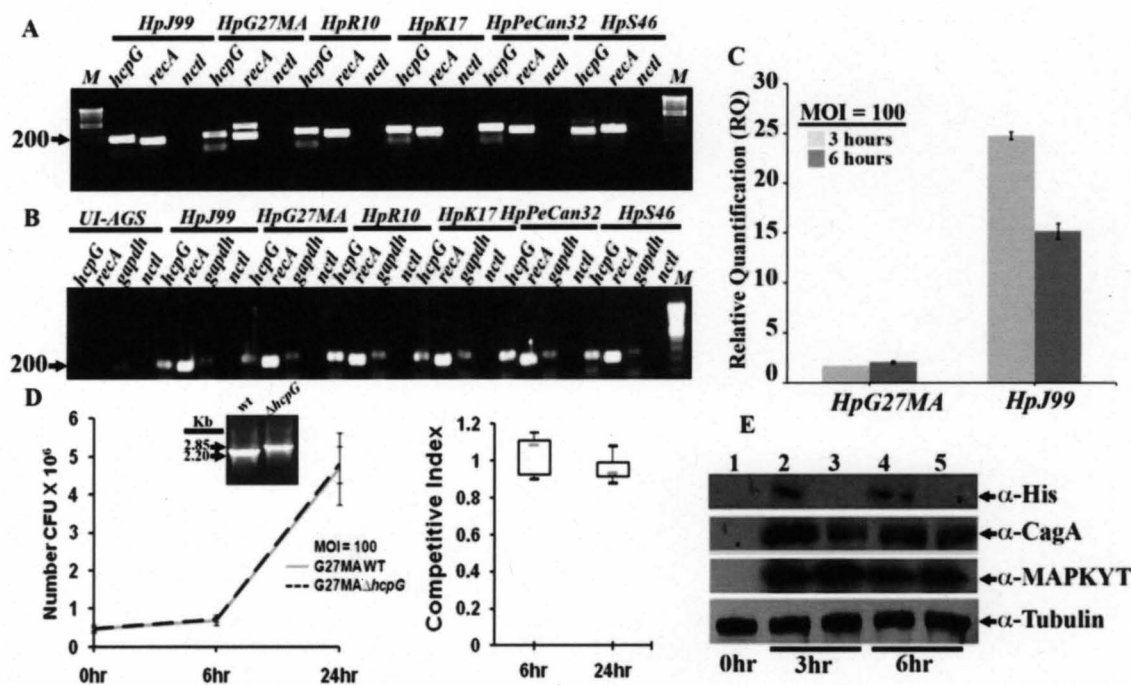


Figure 10: Biological relevance of HcpG during *H. pylori* growth and infection.

(A) Expression dynamics of *hcpG* homologs in diverse *H. pylori* strains grown *in vitro*.

Depicted here is a 1% agarose gel with *hcpG* homologs amplified by reverse transcription PCR of cDNA synthesized from RNA extracted from indicated *H. pylori* strains grown *in vitro* on Brain heart infusion agar plates. *recA* was amplified as a positive control for *H. pylori*, and *nctI* lanes indicate no amplifications from PCR using RNA as a template, suggesting no genomic DNA contamination in cDNA samples.

(B) Expression dynamics of *hcpG* homologs in diverse *H. pylori* strains in an AGS cell infection.

Depicted here is a 1% agarose gel with *hcpG* homologs amplified by reverse transcription PCR of cDNA synthesized from RNA extracted from AGS cells infected with indicated *H. pylori* strains. Infection was proceeded for 6 hrs before RNA was extracted. *recA* and *gapdh* were amplified as a positive controls for *H. pylori* and AGS cells respectively, and *nctI* lanes indicate no amplifications from PCR using RNA as a template, suggesting no genomic DNA contamination in cDNA samples.

(C) Differential expression of *hcpG* in *H. pylori* strains, J99 and G27MA in an AGS cell infection.

3 hr and 6 hr post AGS infection, strain J99 shows a striking up regulation of *hcpG* transcript by 25 fold and 16 fold, respectively. Compared to J99, strain G27MA shows a mild up regulation (~2 fold) of its *hcpG* transcript, 3 hr and 6 hr post infection.

(D) Relative fitness measures of G27MAΔ*hcpG*.

No significant reduction in the relative fitness of *H. pylori* strain G27MAΔ*hcpG* was observed when competed with G27MA WT in an AGS cell infection for 6 hr and 24 hr. Graph in the left

panel depicts colony forming units (CFUs) obtained for each WT and mutant strain when plated on antibiotic resistant BHI agar plates after competing for indicated time points. Gel image in the inset shows fragment ABC amplified from G27MA WT (str^R) and G27MA Δ hcpG (cat^R) *H. pylori* colonies. hcpG mutant ABC fragment is 600bp heavier than WT ABC fragment due to insertion of *rpsL-cat* cassette at hcpG locus. Error bars represent standard deviations. Panel to the right depicts box plot representation of competitive indices calculated from the CFUs in the left panel. Bar located in the box represents the median value. CI > 1 indicates that mutant is favored over the WT and CI < 1 indicates that WT is favored over the mutant. Values of CI at 6 hr = 1.08, 24 hr = 0.92 indicated that there is no significant reduction in the relative fitness of G27MA Δ hcpG ($p > 0.1$) in competition with G27MA WT.

(E) Detection of G27MA HcpG in an AGS cell infection.

Depicted here is an immuno blot detecting G27MA HcpG-His in an AGS cell infection. Lane 1 has cell lysates from uninfected serum starved AGS cells. Lanes 2 and 5 have cell lysates from AGS cells infected with G27MA-hcpG:His at 3hr and 6 hr respectively. Lanes 3 and 5 have cell lysates from AGS cells infected with G27MA-hcpG Δ at 3hr and 6 hr respectively. α - His antibody detects HcpG-His. CagA blot serves as a control for equal MOIs of G27MA-hcpG:His and G27MA-hcpG Δ in AGS cell infection. MAPKYT blot serves as a control for *H. pylori* infection dynamics in AGS cells (activated MAPK is a hall mark of *H. pylori* infection), and Tubulin blot serves as a loading control.

Non-neutral evolutionary dynamics of hcpC

Phylogenetic analysis of *slr* gene homologs from closely related ϵ -proteobacteria revealed a *H. pylori* genome specific *slr* gene family expansion by gene duplication, driven by positive selection, possibly for functional diversity [42]. I hypothesized that one reason for the lack of any obvious contribution by HcpG to bacterial growth was the likely presence of its paralog, HcpC, in *H. pylori* genome. Thus, as a first step towards elucidating any dynamic interplay between HcpC and HcpG during *H. pylori* infection, I sought to elucidate the evolutionary dynamics of the closest homolog or ancestor of hcpG in *H. pylori* genomes. G27MA hcpG is 61.4% (DNA) and 44.5% (amino acid) identical to G27MA hcpC homolog (Fig. 11A & 11B). To better understand hcpC evolutionary dynamics, complete nucleotide sequences from European (Spain), East Asian (Japan and

Korea), South American (Peru, Shima) and African (Gambia) *H. pylori* strains (N=100) were determined. The ML phylogeny reconstructed using the sequence data of a subset of *hcpC* alleles (N=81) revealed strong geographic clustering among *H. pylori hcpC* sequences, which was strikingly absent among *hcpG* alleles. (Fig.11C). Such strong subdivision is very typically seen in most *H. pylori* genes. There were, however, several examples wherein alleles of one geographic origin clustered with those from another, which suggested a history of recent admixture. Selection pressures on HcpC individual codons and branches of its phylogenetic tree were next studied in detail. SSMs confidently identified 23 sites under positive selection ($\omega = 1.4$) (Fig. 11D & 11E), which suggested different selective pressures at different sites in HcpC. The M1bra model was next used to assess the overall variation in ω in all *hcpC* lineages. This model also fit the data significantly better than M0 ($p = 0.003$) (Fig. 11D), which suggested that *hcpC* alleles had been subject to different selective pressures in different populations (Fig. 11C). Detailed estimates of ML parameters associated with each model are shown in Table 5.

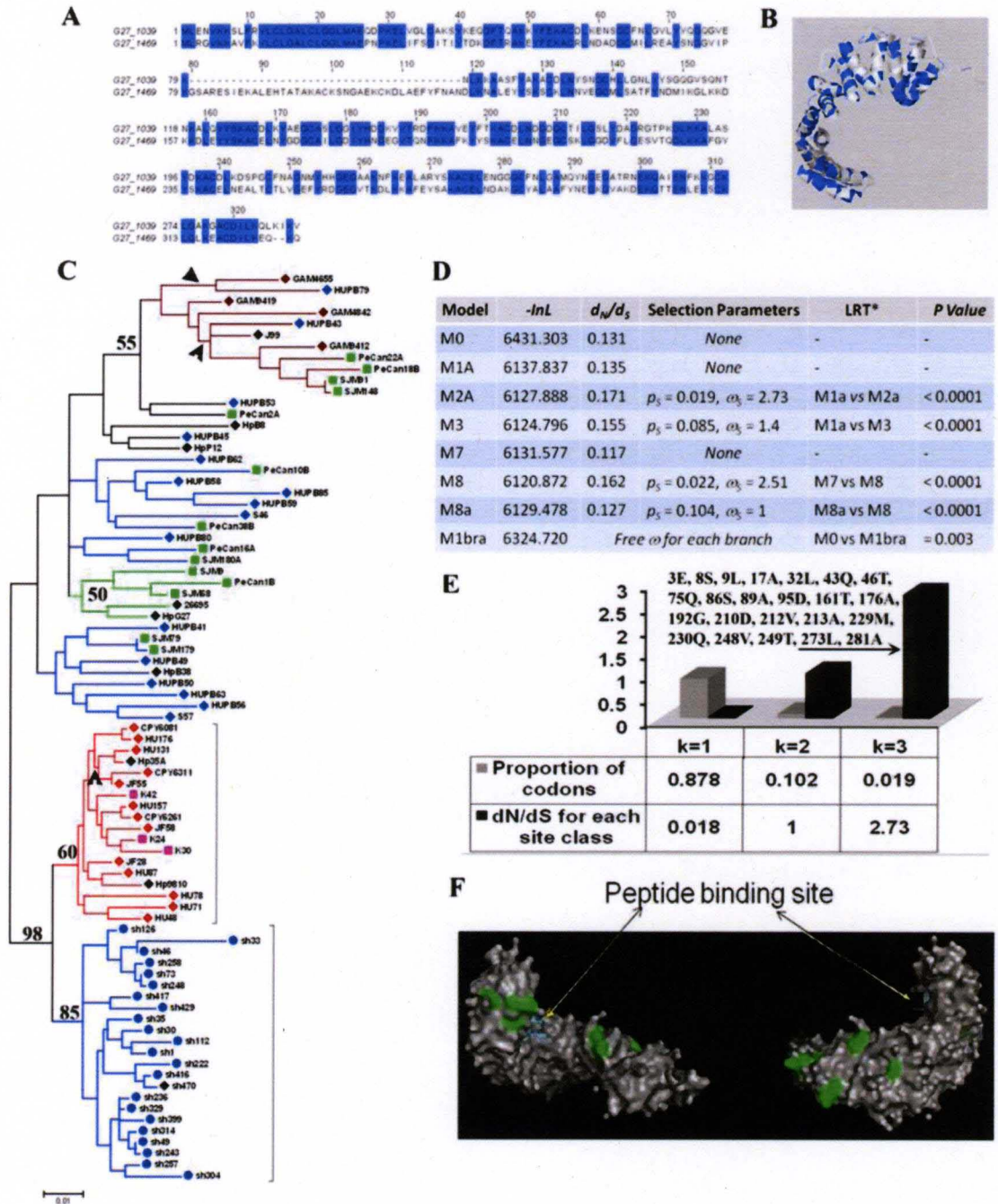


Figure 11: Non neutral evolutionary dynamics of *hcpC*.

(A) G27MA HcpG is 44.5% identical to G27MA HcpC.

Depicted here is a pairwise clustalW alignment generated in Jalview 2.5.1 using HcpC (G27_1039) and HcpG (G27_1469) homologs in G27MA. Blue shaded regions show similarity between the two protein sequences.

(B) Depicts HcpG amino acid residues identical with HcpC, mapped onto HcpC crystal structure.

(C) ML phylogeny of *hcpC* (N=81).

The phylogeny was reconstructed assuming the TrN +I+G substitution model and was optimized to the following parameters using heuristic searches and tree bisection-reconnection algorithm: rate matrix: $A_{AC} = 1$, $A_{AG} = 6.5621$, $A_{AT} = 1$, $C_{CG} = 1$, $C_{CT} = 11.8894$, and $G_{GT} = 1$; base frequencies: A=0.31, C=0.15, G=0.27, T=0.25 proportion of invariable sites, I = 0.5367 and Γ distribution shape parameter = 0.3544. Significant bootstrap values (≥ 50) are shown. Phylogeny is rooted using the out group method implemented in PAUP. Bar scale = 0.01 nucleotide substitutions per site. Lineages under positive selection were indicated with arrow heads.

(D) Maximum-likelihood parameters of selection pressures acting on *H. pylori* HcpC codons.

InL, Log-likelihood score; dN/dS, rate ratio of non-synonymous to synonymous changes averaged over all sites; LRT, likelihood ratio test.

(E) Adaptive evolution in HcpC.

Frequency distribution of three codon classes (k1, k2, and k3) and their associated d_N/d_S ratios were computed under the SSM M3 for HcpC. 23 codons under positive selection (codon class k3) are shown.

(F) Biological relevance of positive selection in *hcpC*.

Most adaptive amino acids under positive selection mapped to the molecular surface of the HcpC (shown in green). The peptide-binding site on HcpC is indicated. Both concave and convex surfaces are shown.

Table 5: Maximum-likelihood parameters of selection pressures acting on *H. pylori* *hcpC* codons

Model Code	InL*	Tree-length	κ^{\dagger}	d _N /d _S	Parameter Estimates	D [d.f.] [‡]	χ^2	P-value	Positively selected sites [Reference sequence: HpSh30]
M0 (1-ratio)	-6431.304	3.358	3.448	0.1311	$\omega = 0.1311$	NA	NA	NA	NA
M1a (Nearly Neutral)	-6137.837	3.447	3.478	0.135	$p_0 = 0.881$, $\omega_0 = 0.018$; $p_1 = 0.119$, $\omega_1 = 1$	NA	NA	NA	NA
M2a (Positive Selection)	-6127.888	3.499	3.82	0.171	$p_0 = 0.878$, $\omega_0 = 0.018$; $p_1 = 0.102$, $\omega_1 = 1$; [p_s]=0.019, [ω_s]=2.73	M0 vs M2a [3] M1a vs M2a [2]	606.832 19.898	<0.0001 <0.0001	9L, 43Q, 75Q, 161T, 229M, 248V
M3 (Discrete)	-6124.796	3.483	3.695	0.155	$p_0 = 0.794$, $\omega_0 = 0.004$; $p_1 = 0.119$, $\omega_1 = 0.258$; [p_s]=0.085, [ω_s]=1.4	M0 vs M3 [4] M1a vs M3 [4]	613.016 26.082	<0.0001 <0.0001	3E, 8S, 9L, 17A, 32L, 43Q, 46T, 75Q, 86S, 89A, 95D, 161T, 176A, 192G, 210D, 212V, 213A, 229M, 230Q, 248V, 249T, 273L, NA
M7 (β)	-6131.577	3.389	3.452	0.117	$p = 0.021$, $q = 0.145$	NA	NA	NA	NA
M8 (β , $\omega_s > 1$)	-6120.872	3.547	3.742	0.162	$p = 0.045$, $q = 0.375$; $p_0 = 0.978$, [p_s]=0.022, [ω_s]=2.51	M7 vs M8 [2]	21.41	<0.0001	8S, 9L, 17A, 43Q, 46T, 75Q, 86S, 89A, 95D, 161T, 176A, 192G, 210D, 212V, 213A, 229M, 230Q, 248V, 249T, 281A
M8a (β , $\omega_s = 1$)	-6129.478	3.464	3.426	0.127	$p = 0.102$, $q = 3.291$; $p_0 = 0.895$, [p_s]=0.104, $\omega_s = 1$	M8a vs M8 [1]	17.212	<0.0001	NA
M1bra	-6324.72	3.362	3.458	NA	Free ω for each lineage	M0 vs M1bra	213.168	<0.004	NA

* InL, Log Likelihood Score

\dagger , κ , kappa, ratio of transition to transversions

\ddagger , D, hierarchical Likelihood Ratio Test statistic and d.f., degrees of freedom

Biological significance of HcpC adaptive evolution

To examine adaptive evolution in HcpC in a protein-structure function context, amino acids identified to be under positive selection were mapped on the crystal structure of HcpC in collaboration with Dr. Peer Mittl at the University of Zurich. Most adaptive amino acids mapped to the molecular surface of the HcpC protein (Fig. 11F), which was reminiscent of our findings with other *H. pylori* *slr* genes [42] and suggested that positive selection may affect the strength or specificity of interaction. Taken together, the presence of *hcpC* in all *H. pylori* strains tested, and evidence of positive selection superimposed upon strong geographic clustering suggest a likely essential role of *hcpC* in *H. pylori* infection.

Growth kinetics of G27MA Δ *hcpC* and G27MA Δ *hcpC* Δ *hcpG*

To further characterize the functional dynamics of HcpC, I assessed the role of HcpC in *H. pylori*'s survival and reproduction *in vitro* and in AGS cell-culture infection model. For this purpose, I conducted growth assays with *hcpC* single mutant, and a double mutant lacking both *hcpC* and *hcpG* in *H. pylori* strain G27MA. No significant defect ($p > 0.1$) was observed in the growth rate of G27MA Δ *hcpC* 6 hr (lag phase), 24 hr (mid exponential phase), 48 hr (stationary phase), and 56 hr (death phase) grown *in vitro*. However, the G27MA Δ *hcpC* Δ *hcpG* strain exhibited a significant growth defect ($0.01 > p < 0.05$) 48 hr and 56 hr when grown *in vitro* (Fig. 12B). This finding suggests that nutrient scarcity poses adverse effects to the absolute fitness of double mutant G27MA Δ *hcpC* Δ *hcpG* grown *in vitro*. However, no significant reduction ($p > 0.1$) was observed in the relative fitness of both G27MA Δ *hcpC* (Fig. 12D) and G27MA Δ *hcpC* Δ *hcpG* mutant strains (Fig. 12F), when competed with G27MA WT *in*

vitro. Next, I examined the role of HcpC and HcpG in *H. pylori*'s growth and survival 6 hr and 24 hr post AGS cell infection. When compared to the G27MA WT, significant difference ($0.01 > p < 0.05$) was found in the growth dynamics of *G27MAΔhcpC* and *G27MAΔhcpCΔhcpG* (Fig. 12A) strains 24 hr post infection. This finding suggested a possible role of HcpC relatively late infection in the AGS cell culture model.

HcpC and HcpG paralogs are redundant and contribute additively to relative fitness of *H. pylori* strain G27MA in AGS cell infection

Next, I assessed for relative fitness defects (if any) in *G27MAΔhcpC* and *G27MAΔhcpCΔhcpG* mutants in AGS cell infection model. Results showed that both the tested mutants showed a significant ($p < 0.01$) reduction (30% decreased relative fitness in *hcpC* mutant and 40% in *hcpC*, *hcpG* double mutant) in the relative fitness of respective bacterial strains at 24 hr post infection, but only the double mutant showed a significant reduction ($p < 0.01$) (25% reduced relative fitness), 6 hrs post infection ($p < 0.01$) (Fig. 12C & 12E). Thus my results from 6 hrs post infection show that while loss of either *hcpC* or *hcpG* alone does not adversely affect *H. pylori* fitness, lack of both *hcpC* and *hcpG* significantly affected the ability of *H. pylori* to grow during infection, thereby revealing that *hcpG* and *hcpC* are genetically redundant. Similarly, while *hcpG* deletion had no significant impact on *H. pylori* growth, lack of *hcpC* significantly reduced *H. pylori*'s ability to compete with the WT parent strain at 24 hrs post infection. Strikingly, the mutant derivative lacking both *hcpC* and *hcpG* performed worse than the WT or *hcpG* and *hcpC* mutants themselves. This outcome suggested that HcpC and HcpG were not only genetically redundant, their contribution to *H. pylori* fitness was additive.

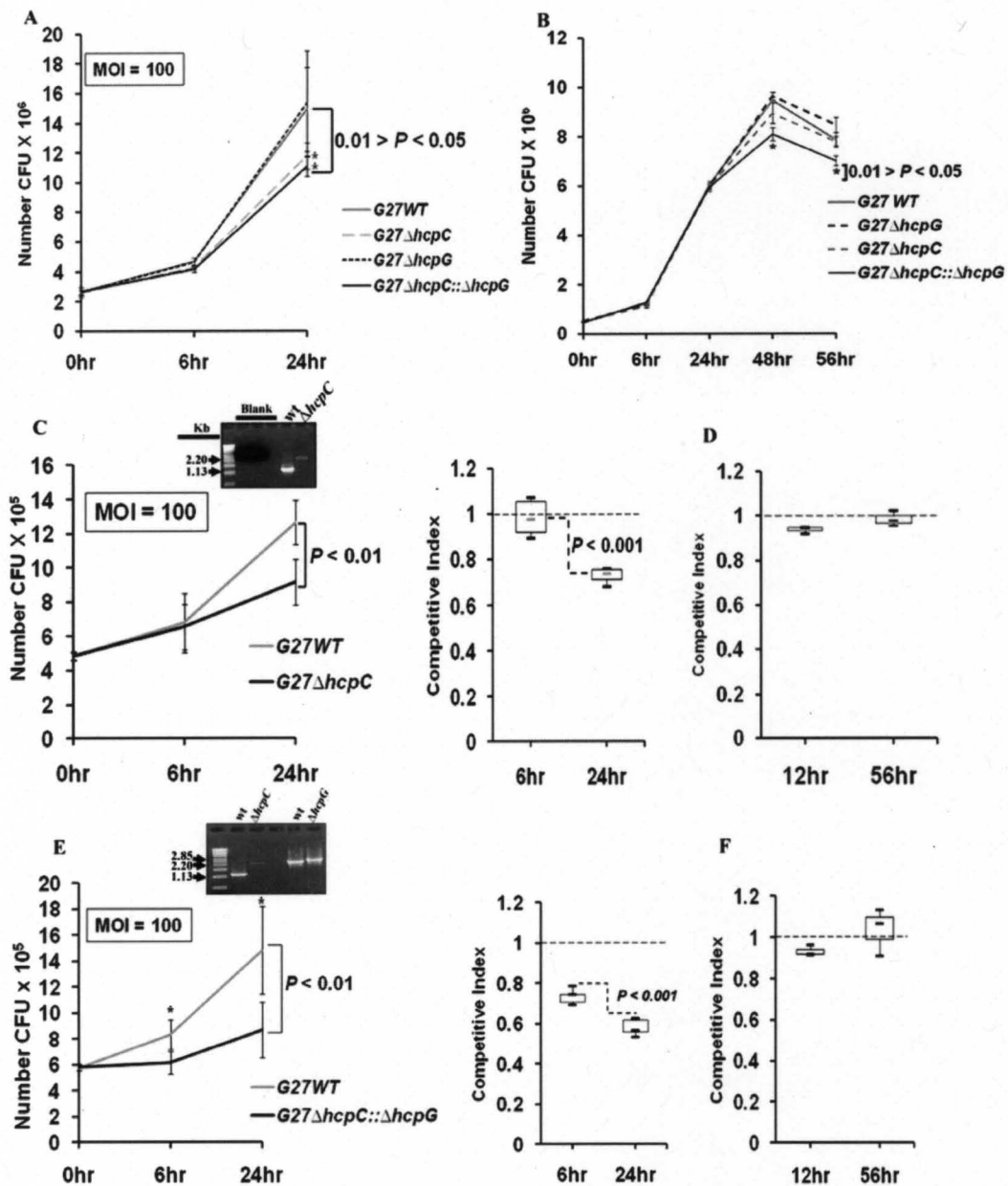


Figure 12: Growth and relative fitness dynamics of G27MA – *hcp* mutants.

(A) Growth kinetics of G27MA – *hcp* mutants in an AGS cell infection.

The infection was carried out in triplicate using the indicated G27MA mutants and G27MA WT with a MOI of 100. G27MA Δ hcpC and G27MA Δ hcpC Δ hcpG mutant strains exhibited a significant ($0.01 > p < 0.05$) decrease in the growth rate 24 hr post infection when compared with G27MA WT. Error bars indicate standard deviations.

(B) Growth kinetics of G27MA – *hcp* mutants grown *in vitro*.

With a starting bacterial density OD_{600} of 0.05 / ml, BHI broths were inoculated with the indicated G27MA – *hcp* mutants and G27MA WT and grown for indicated time points. Only the *G27MAΔhcpCΔhcpG* double mutant exhibited a significant growth defect ($0.01 > p < 0.05$) when grown for 48 hr and 56 hr *in vitro*. Error bars represent standard deviations from four independent experiments.

(C) Relative fitness measures of *G27MAΔhcpC* in AGS cell infection.

Significant reduction in the relative fitness of *H. pylori* strain *G27MAΔhcpC* was observed when competed with G27MA WT in an AGS cell infection for 24 hr. Graph to the left depicts colony forming units (CFUs) obtained for each WT and mutant strain when plated on antibiotic resistant BHI agar plates after competing for indicated time points. Gel image in the inset shows fragment ABC amplified from G27MA WT (*str^R*) and *G27MAΔhcpC* (*erm^R*) *H. pylori* colonies. *hcpC* mutant ABC fragment is 1400bp larger than WT ABC fragment due to insertion of *rpsI- erm* cassette at *hcpC* locus. Error bars represent standard deviations of four independent experiments. Graph to the right depicts box plot representation of competitive indices calculated from the CFUs in the left panel. Bar located in the box represents the median value. $CI > 1$ indicates that mutant is favored over the WT and $CI < 1$ indicates that WT is favored over the mutant. Values of CI at 6 hr = 0.97, 24 hr = 0.73 indicated that there is a 27% reduction in the relative fitness of *G27MAΔhcpC* ($p < 0.001$) in competition with G27MA WT.

(D) Relative fitness measures of *G27MAΔhcpC* grown *in vitro*.

Graph depicted here is a box plot representation of competitive index (CI). Values of CI at 6 hr = 0.92, 24 hr = 0.98 indicated that there is no significant reduction in the relative fitness of *H. pylori* strain *G27MAΔhcpC* when competed with G27MA WT *in vitro*, and grown for 12 hr and 56 hr.

(E) Relative fitness measures of *G27MAΔhcpCΔhcpG* in AGS cell infection.

Significant reduction ($p < 0.001$) in the relative fitness of *H. pylori* strain *G27MAΔhcpCΔhcpG* was observed when competed with G27MA WT in an AGS cell infection for 6hr and 24 hr. Graph to the left depicts colony forming units (CFUs) obtained for each WT and mutant strain when plated on antibiotic resistant BHI agar plates after competing for indicated time points. Gel image in the inset shows fragment ABC amplified from G27MA WT (*str^R*), *G27MAΔhcpC* (*erm^R*), and *G27MAΔhcpG* (*cat^R*) *H. pylori* colonies. Graph to the right depicts box plot representation of competitive indices calculated from the CFUs in the left panel. Bar located in the box represents the median value. Values of CI at 6 hr = 0.74, 24 hr = 0.56 indicated that there is a 27% reduction in the relative fitness of *G27MAΔhcpCΔhcpG* ($p < 0.001$) in competition with G27MA WT.

(D) Relative fitness measures of *G27MAΔhcpCΔhcpG* grown *in vitro*.

Graph depicted here is a box plot representation of competitive index (CI). Values of CI at 6 hr = 0.94, 24 hr = 1.09 indicated that there is no significant ($p > 0.1$) reduction in the relative fitness of *H. pylori* strain *G27MAΔhcpCΔhcpG* when competed with G27MA WT *in vitro*, and grown for 12 hr and 56 hr.

Role of HcpC and HcpG in surface translocation of HspB

Even though the SEL1-like domains are often generally involved in protein-protein interactions, [43, 55, 127-130] little is known about the biological function of Helicobacter SLR Hcp proteins. The crystal structure of HcpC had suggested a peptide binding site in its crystal contact I surface, which was similar to the binding site seen in eukaryotic TPR protein Hsp70/Hsp90 organizing protein (Hop) [131]. *H. pylori* genomes encode 10 - 11 heat shock proteins; among these the 58 kDa heat shock protein B (HspB or GroEL) is unique in that, unlike other bacteria where it is cytoplasmic, in *H. pylori* HspB is also found on the bacterial surface in association with Urease subunit UreB [132-135]. Given these findings, and that eukaryotic Sel-1 proteins are known to function as chaperones in assembling macro molecular complexes [49-52], I hypothesized that *hcpC/hcpG* might function as chaperone proteins, likely involved in translocation of HspB to *H. pylori* membrane. Thus, HspB surface expression in AGS infection assays with G27MAWT or *G27MAΔhcpC* or *G27MAΔhcpG* or *G27MAΔhcpCΔhcpG* was quantified using a BD FACS Calibur (Beckton Dickenson Inc., USA) at 3 hr, 6 hr, 12 hr and 24 hr post AGS infection. This analysis clearly indicated that all the G27MA - *hcp* mutants tested showed a significant ($p < 0.01$) defect in the surface localization of HspB compared to the WT G27MA strain at 3 hr post infection (Fig. 13A). However, only the strains lacking *hcpC* and *hcpC* & *hcpG* showed a significant defect ($p < 0.01$) in surface localization of HspB at 6 hr (Fig. 13B), 12 hr (Fig. 13C) and 24 hr (Fig. 13D) post infection. These findings suggest that *hcpC* and *hcpG* can each modulate HspB surface expression, and that HcpC can compensate for the loss of HcpG in *G27MAΔhcpG* strain by 6hr post infection. However, HcpG is unable to compensate for the lack of HcpC as

illustrated by HspB expression dynamics during infection with G27MA Δ *hcpC* infection, whereby the *hcpC* mutant showed significant reduction in HspB surface expression throughout the infection time course (Fig. 13). Significantly, the HspB dynamics during infection with G27MA Δ *hcpC* Δ *hcpG* and G27MA Δ *hcpC* were quite similar. These findings suggest the following conclusions: 1) *hcpC* is necessary and sufficient for optimal surface expression of HspB; 2) that *hcpC* and *hcpG* are functionally redundant but in a non-reciprocal fashion; and 3) that non-reciprocal functional redundancy stems from the relatively higher efficiency of *hcpC* to regulate HspB surface expression. Dynamics of CagA and phosphorylated MAPK were monitored as controls to ensure the specificity of HspB dynamics in G27MA - *hcp* mutants (Fig. 14). These data showed that parameters of *H. pylori* infection, independent of HcpG and HcpC, remained unaffected, thereby supporting my observation that HcpC and HcpG function to specifically modulate surface exposure of HspB antigen.

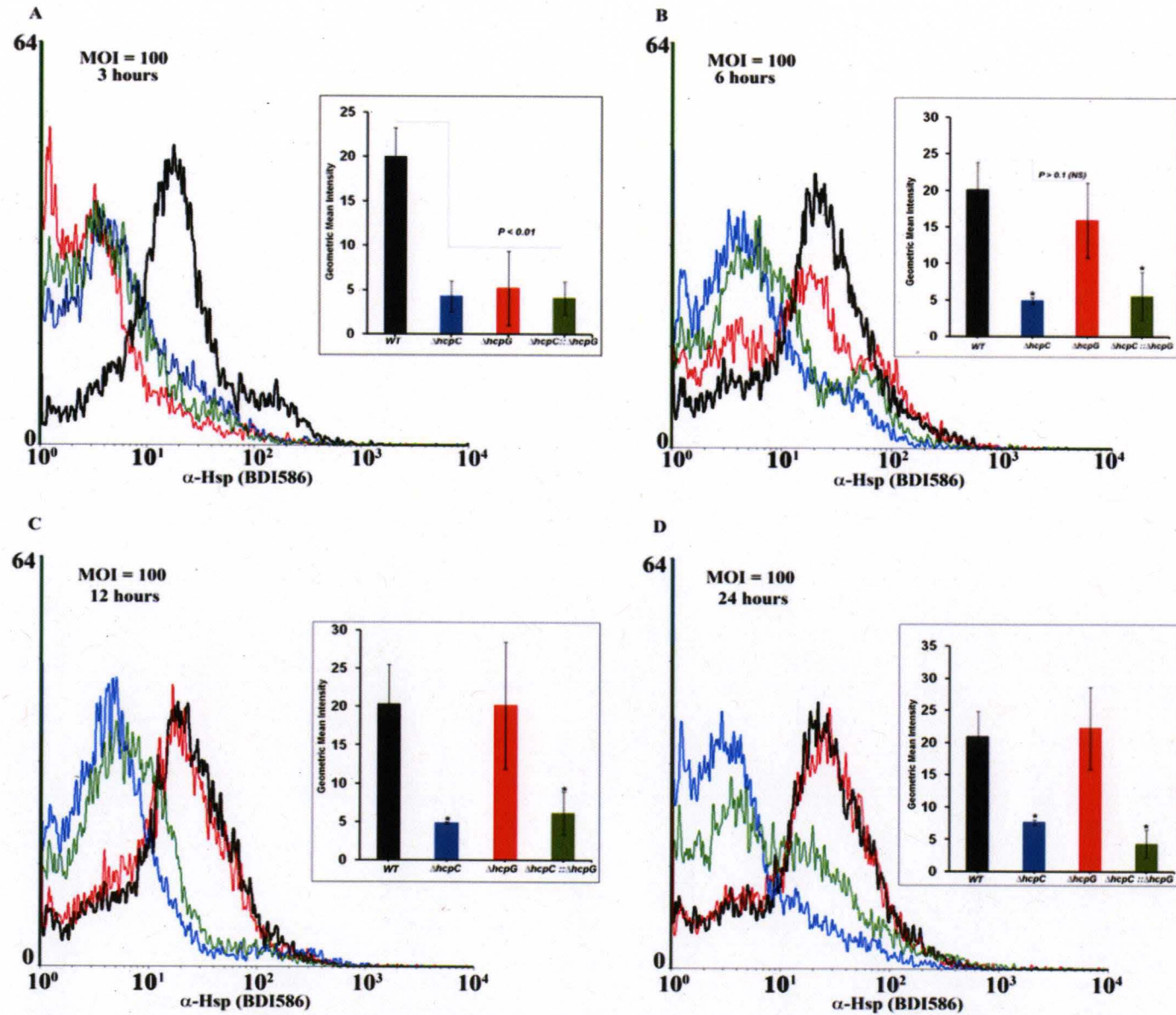


Figure 13: Role of G27MA HcpC and G27MA HcpG in *H. pylori* surface translocation of HspB.

(A) Defect in surface localization of HspB in all the tested G27MA – *hcp* mutants, 3 hr post AGS infection.

Depicted here is a histogram representation of mean fluorescence of surface expressed HspB; black – G27MA WT, blue – G27MAΔ*hcpC*, red – G27MAΔ*hcpG*, green – G27MAΔ*hcpC*Δ*hcpG*. x – axis indicates HspB fluorescence; y- axis indicates number of events. Decrease in the amount of fluorescence shifts the peaks to left (here, blue, red and green peaks compared to the black peak). Graph in the inset is plotted using geometric mean intensity of HspB fluorescence to calculate statistical significance. Significant difference ($p < 0.01$) in geometric mean intensity of HspB fluorescence was detected in all the G27MA – *hcp* mutants compared to G27MA WT. Error bars represent standard deviations from three independent experiments.

(B) Defect in surface localization of HspB in G27MAΔ*hcpC*, and G27MAΔ*hcpC*Δ*hcpG*, 3 hr, 6hr (C), and 24 hr (D) post AGS infection.

Following 6 hr post AGS infection, G27MAΔ*hcpG* strain did not exhibited a defect in surface translocation of HspB, whereas G27MAΔ*hcpC*, and G27MAΔ*hcpC*Δ*hcpG* strains consistently exhibited the defect. Error bars represent standard deviations from three independent experiments. * indicates statistical significance.

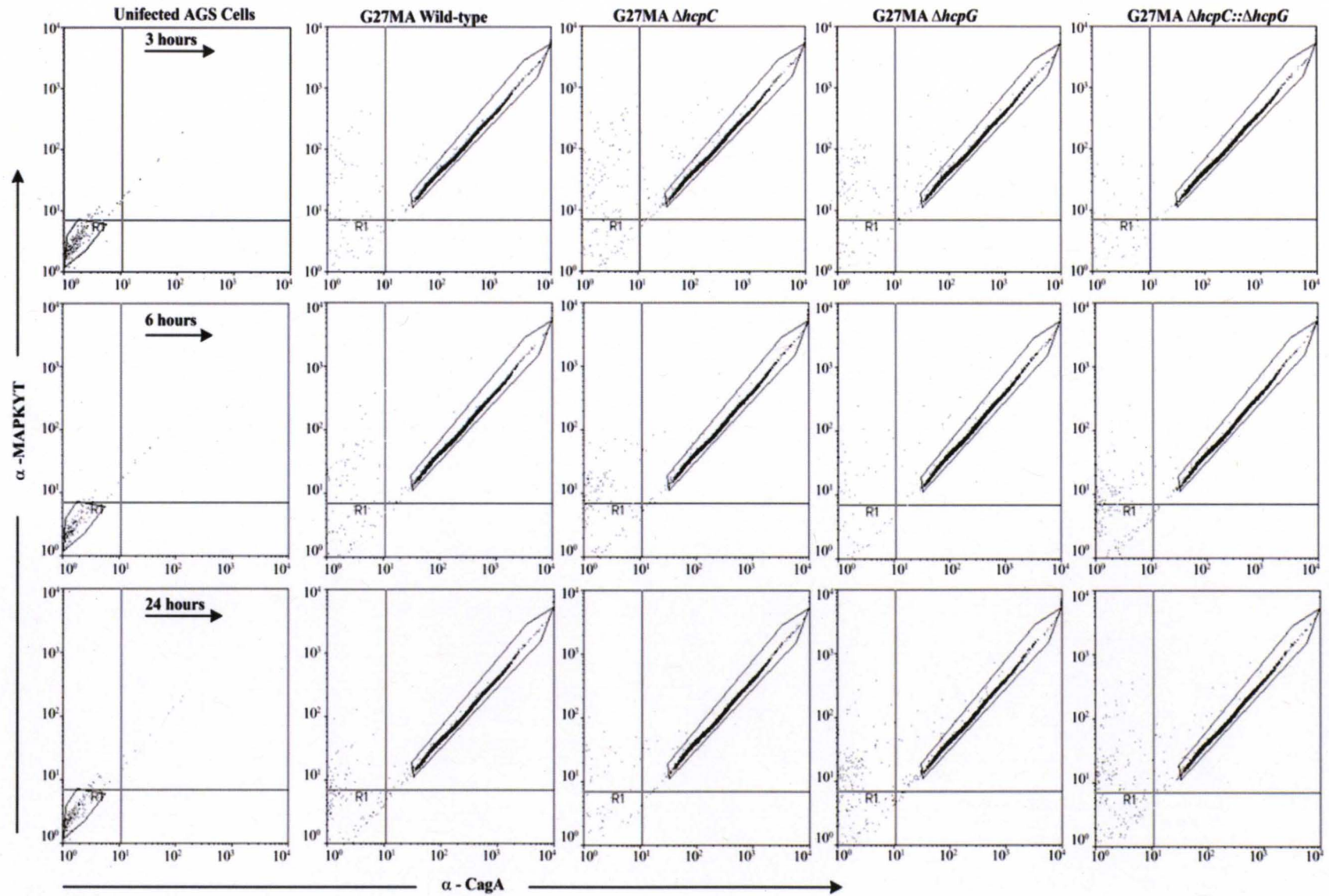


Figure 14: Dynamics of CagA and MAPKYT in G27MA – AGS infection.

Depicted here is a box plot representation of CagA fluorescence (x – axis) and MAPKYT fluorescence (y- axis) of the AGS cells infected with indicated G27MA mutants and G27MA WT strain. As expected, in uninfected AGS control, the CagA (control for equal MOI) and MAPKYT (activation is a hall mark of host epithelial cells following *H. pylori* infection) fluorescence is very low, which were then up regulated post AGS infection with indicated *H. pylori* strains. Following AGS infection for indicated time points, no difference was observed in CagA, and MAPKYT dynamics among G27MA WT and G27MA - *hcp* mutants. This suggests that dynamics seen in HspB surface localization was specific to G27MA - *hcp* mutants.

Next, I asked if the variation in HspB surface localization in G27MA - *hcp* mutants was indeed due to a defect in surface translocation of HspB or, as a result of decreased synthesis of HspB in the G27MA - *hcp* mutant bacterial strains. To test this, I assessed the expression dynamics of *hspB* in the G27MA^{WT} and G27MA - *hcp* mutant bacterial strains in the AGS cell infection model using quantitative real time PCR. Since, *H. pylori* HspB is closely associated with Urease [136], I also tested the expression dynamics of urease functional sub unit, *ureB* in G27MA - *hcp* mutants, post AGS cell infection. Results from these experiments indicated that there was no apparent reduction i.e., < 2 fold in the amount of *hspB* (Fig. 15B), and *ureB* (Fig 15C) transcript compared to the WT strain, 3 hr and 6 hr post AGS infection. Taken together, I conclude that only surface translocation of HspB is affected in the tested G27MA - *hcp* mutants and that both HcpC and HcpG play a role in transporting HspB from cell cytosol to the bacterial surface, 3 hr post infection. However, by 6 hr post infection, HcpC compensates for the loss of HcpG in G27MA Δ *hcpG* strain as no defect in HspB surface localization was observed in G27MA Δ *hcpG* strain from this time interval. A consistent defect was shown by the *hcpC* mutant however, indicating a non-reciprocal effect of HcpG on loss of *hcpC* in *H. pylori* strain G27MA Δ *hcpC*.

Next, I sought to characterize the compensatory mechanism of G27MA HcpC in G27MA Δ *hcpG* mutant strain. I hypothesized that G27MA HcpC might have played a role in rescuing the defect of HspB surface translocation in G27MA Δ *hcpG* via its transcript up regulation by 6 hrs post infection. To test this hypothesis, I quantified the level of *hcpC* transcript in G27MA WT *H. pylori* strain and G27MA Δ *hcpG* strain, 3 hr, 6 hr, and 24 hr post AGS infection. However, results from this analysis indicated no apparent up regulation i.e., > 2 fold of *hcpC* transcript in G27MA Δ *hcpG* mutant strain (Fig. 15A). Yet, it is still possible that, other regulatory mechanisms or stability of HcpC might have played a role in rescuing the observed defect.

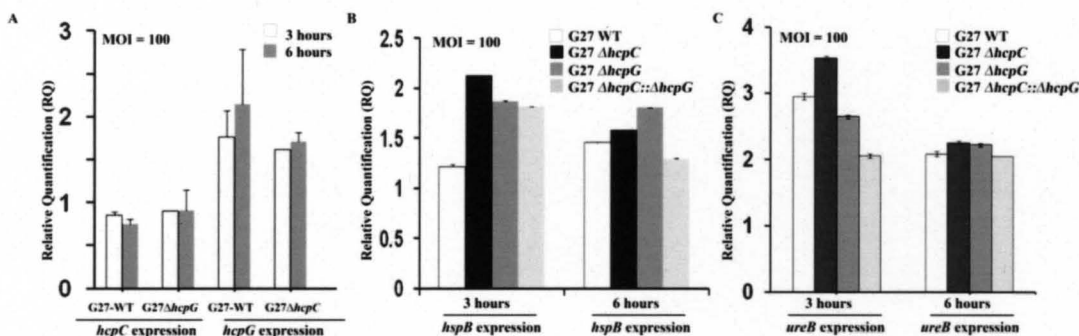


Figure 15: Synthesis of HspB is not affected in G27MA – *hcp* mutants.

(A) Regulation of G27MA - *hcpC* transcript expression in G27MA - *hcpG* mutant.

qRT-PCR was used to determine the regulation of expression of *hcpC* and *hcpG* in G27MA WT and in indicated G27MA - *hcp* mutants. There is no apparent up regulation of either *hcpC* or *hcpG* transcripts in the absence of their paralogs. The results are represented as relative quantification normalized to *ureA*. Error bars represent standard deviations of triplicate experiments.

(B) Expression dynamics of *hspB* in G27MA - *hcp* mutants.

Real Time qRT-PCR was used to determine the regulation of expression of *hspB* in G27MA WT and indicated G27MA - *hcp* mutants, 3hr and 6hr post AGS infection. There is no apparent down regulation of HspB transcript in G27MA – *hcp* mutants' following AGS infection, suggesting that only surface translocation of HspB was affected. The results are represented as relative

quantification normalized to *recA*. Error bars represent standard deviations of triplicate experiments.

(C) Expression dynamics of *ureB* in G27MA - *hcp* mutants.

Real Time qRT-PCR was used to determine the regulation of expression of *ureB* in G27MA WT and indicated G27MA - *hcp* mutants, 3hr and 6hr post AGS infection. There is no apparent up / down regulation of *UreB* transcript in G27MA - *hcp* mutants' following AGS infection. The results are represented as relative quantification normalized to *recA*. Error bars represent standard deviations of triplicate experiments.

HcpC and HcpG dependent modulation of HspB surface expression requires cellular infection.

Next, I asked if the defect seen in HspB surface localization in G27MA - *hcp* mutants is specific to the presence of host cellular components (modeled by AGS infection) or a constitutive property of G27MA - *hcp* mutants *in vitro* (BHI liquid culture). To determine this, I measured and compared the fluorescence of bacterial surface associated HspB in G27MA WT and G27MA - *hcp* mutants that are grown for 3 hr, 24 hr, and 56 hr *in vitro*. Comparison of mean geometric fluorescent intensities of the WT and *hcp* mutants revealed no difference in the mean fluorescence among the strains tested grown *in vitro* for 3 hr, 24 hr, and 56 hr (Fig. 16A, 16B, and 16C, respectively). The 56 hr time point was included as *G27MAΔhcpCΔhcpG* double mutant grown for 56 hr *in vitro*, exhibited a significant defect ($0.01 > p < 0.05$) in its growth rate (Fig. 12B). However, as shown above, decreased growth rate had no effect in HspB surface translocation. Hence, although the number of *G27MAΔhcpCΔhcpG* were significantly less than the G27MA WT by 56 hrs when grown *in vitro*, the amount of HspB translocated to the bacterial surface was not altered in the double mutant. Taken together,

my data indicate that the defect in HspB surface localization exhibited by G27MA – *hcp* mutants is specific to the cellular infection model.

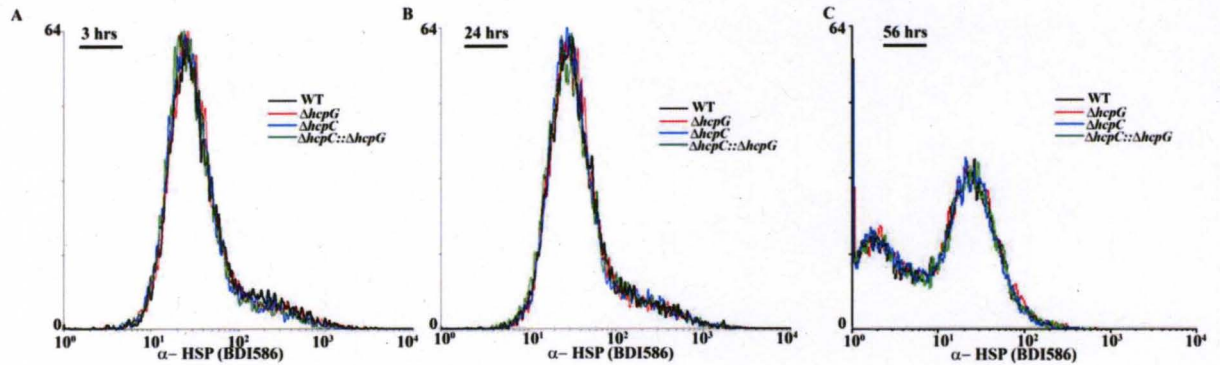


Figure 16: Cellular infection dependent surface translocation defect of HspB in G27MA – *hcp* mutants.

(A) HspB surface localization in G27MA - *hcp* mutants *in vitro*.

No defect is observed in the surface localization of HspB in G27MA - *hcp* mutants grown *in vitro* for 3hr; (B) 24hr; and (C) 56hr. HspB surface expression was detected by measuring fluorescence of HspB expressed on the surface of bacteria, using a BD FACS Calibur machine.

Summary I:

1. *hcpG* is only present in a subset of *H. pylori* strains, and in a significant number of *H. pylori* strains *hcpG* is pseudogenized or deleted.
2. Striking sequence variability was found in *hcpG* homologs in diverse *H. pylori* isolates.
3. *hcpC* and *hcpG* evolve rapidly under positive selection and HcpC contributes significantly to *H. pylori* growth and fitness during infection.
4. HcpC and HcpG are genetically redundant and contribute additively to *H. pylori* fitness.
5. HcpC is necessary and sufficient for optimal surface expression of HspB.

6. Absence of HcpG downregulates HspB surface expression during very early infection and its loss is compensated for by HcpC thereby implying that HcpC and HcpG are functionally redundant.
7. HcpG alone does not rescue the lack of HcpC, suggesting that the observed functional redundancy is non-reciprocal.

II. Role of HepC during *H. pylori* Growth and Infection.

Differential regulation of *hepC* expression in diverse *H. pylori* isolates in AGS cell infection model

Earlier studies indicated a likely role of HepC in the adaptation of *H. pylori* to diverse human hosts [42]. Thus, to elucidate the biological relevance of HepC in *H. pylori* infection, I first monitored the expression dynamics of *H. pylori* strain G27MA *hepC* grown *in vitro* and in AGS cell infection model by Reverse transcription PCR. These experiments revealed that HepC transcript was detectable *in vitro* (BHI medium) and in the AGS cell infection model (Fig. 17A). Since *hepC* evolves rapidly in diverse *H. pylori* populations with different amino acids often being selected in different geographic regions likely fine tuning host responses [42], I hypothesized that genetically diverse *H. pylori* isolates differentially regulate *hepC* transcript expression. Therefore, I used quantitative real time PCR to obtain a quantitative measure of the *hepC* transcript in AGS cell infection of diverse *H. pylori* strains. As expected, this analysis revealed that genetically diverse *H. pylori* strains differentially regulate expression of *hepC* transcript (Fig. 17B). More importantly, I determined that the Japanese *H. pylori* strain, JS7 exhibited a maximal up regulation in *hepC* expression (~ 10 fold), 3 hr and 6 hr post infection compared to *H. pylori* strains isolated from European (26695), Amerindian

(Shi470) and South African (R10) populations (Fig. 17B). Such differential regulation of *hepC* in different genomic contexts in the AGS infection suggested that *hepC*'s role in *H. pylori* strains was likely tightly regulated, which in turn indicated that HepC is likely biologically relevant during *H. pylori* infection.

HepC contributes significantly to the fitness of *H. pylori*

To further characterize the biological significance of HepC in *H. pylori* infection, I then sought to determine the contribution of HepC to *H. pylori*'s fitness in an AGS cell infection model. Since relative fitness measures provide more accurate measures at understanding evolutionary success [121], I determined the relative fitness of *G27MAΔhepC* strain compared to *G27MA* WT *H. pylori* strain for survival and replication, in AGS cell infection model. Results from competition assay revealed that *G27MAΔhepC* mutant showed a significant reduction (30% reduction) in the relative fitness, 24hr post infection (Fig. 17D) (student's *t*- test; $p < 0.01$) and a non-significant reduction 6 hr post infection (Fig. 17D). Since, *G27MAΔhepC* exhibited a significant reduction in relative fitness compared to *G27MA* WT, 24 hr post infection but not at 6 hr, I hypothesized that HepC plays a crucial role late in the infection. To test this hypothesis, I monitored the expression pattern of *hepC* in *G27MA* – AGS infection to detect if *hepC* expression was significantly up regulated 24 hrs post infection. As expected, there was an apparent up regulation (~ 13-fold) of *hepC* transcript 24 hr post infection when compared to its expression pattern at 3 hr and 6 hr post AGS infection (Fig. 17B).

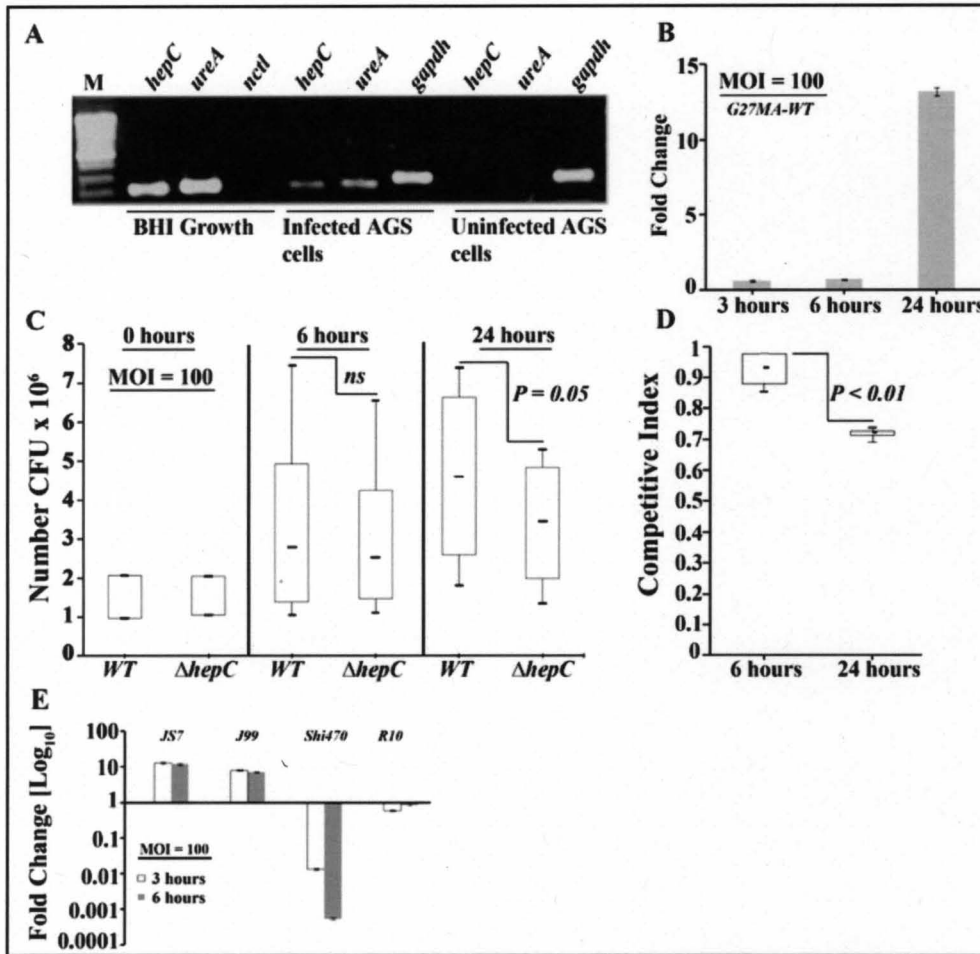


Figure 17: *hepC* appears to be relevant biologically, and required during *H. pylori* infection.

(A) *hepC* mRNA is detected during *in vitro* growth and 6 hours following infection of AGS cells. Depicted here is a 1% agarose gel with G27MA *hepC* homolog amplified by reverse transcription PCR of cDNA synthesized from RNA extracted from G27MA grown *in vitro* on Brain heart infusion agar plates and 6 hr post AGS cell infection. *recA* and *gapdh* were amplified as positive controls for *H. pylori* and AGS cell infection, respectively. *nctI* lanes indicate no amplifications from PCR using RNA as a template, suggesting no genomic DNA contamination in cDNA samples.

(B) Real-time PCR analysis shows that the *hepC* transcript is dramatically upregulated (~13-fold) following 24 hrs of infection with *H. pylori* strain G27MA.

(C) (D) Deletion of *hepC* reduces the ability of G27MA Δ *hepC* to grow and compete with the WT resulting in a 30% reduction in the relative fitness of G27MA Δ *hepC*, 24 hours post infection.

(E) Real-time PCR analysis shows that *hepC* is differentially regulated in different genomic contexts.

HepC likely targets the host cytoskeletal machinery during late infection in AGS cell culture model of infection

Ongoing work from other investigators in the lab demonstrated that HepC interacts with the multifunctional human cytoskeletal proteins Ezrin, and Vinculin. Therefore, I sought to determine the aspect of host cytoskeleton that HepC targets. Cell scattering or humming bird phenotype of AGS cells is a result of *H. pylori* infection mediated cytoskeletal dysregulation, and is a hallmark of *H. pylori* infected AGS cells [83, 84]. With this rationale and given that HepC interacts with human cytoskeletal proteins, Ezrin and Vinculin, I first asked if HepC has a role in cell scattering of AGS cells. To test this, I infected AGS cells with *H. pylori* strains G27MA WT, *G27MA Δ hepC*, *G27MA Δ virD4* (a type IV secretion system deletion mutant which is unable to induce cell scattering phenotype), 26695 WT, *26695 Δ hepC*, and measured the cell scattering phenotype 6 hr post infection. I identified that both the *hepC* mutants tested mediated cell scattering phenotype similar to their respective WT strains (Fig. 18A). This finding suggested that HepC is not likely to mediate early cytoskeletal changes. However, as our earlier finding (Fig. 17C, 17D & 17E) indicated a likely role of HepC, late in the infection, I then sought to determine the molecular basis of cytoskeletal deregulation (if any) by HepC, 24 hr post AGS infection. Quantifying the cell scattering phenotype 24 hrs post infection posed a challenge due to extreme mobility of cells. Using cytoskeletal pathway specific PCR arrays I quantified the relative expression dynamics of cytoskeletal regulators in the *G27MA Δ hepC* mutant compared to the G27MA WT, 24 hrs post infection, to identify cytoskeletal regulators that were specifically targeted by HepC. As expected, scatter plot analyses demonstrating > or < 2 fold differences in gene

expression, identified 10 genes were identified that were apparently up regulated and 1 gene that was apparently down regulated in *G27MAΔhepC* infection when compared to *G27MA* WT infection (Fig. 18B). Inferences from this experiment are as follows: 1) cytoskeletal regulators that were identified as up regulated in *G27MAΔhepC* infection were likely down regulated by HepC in WT infection; and that 2) cytoskeletal regulators that were identified as down regulated in *G27MAΔhepC* infection were likely up regulated by HepC in WT infection. Such dysregulation of cytoskeletal regulators by HepC, 24 hrs post infection suggests that HepC most likely targets host cytoskeletal machinery during late infection in a cell culture model of infection.

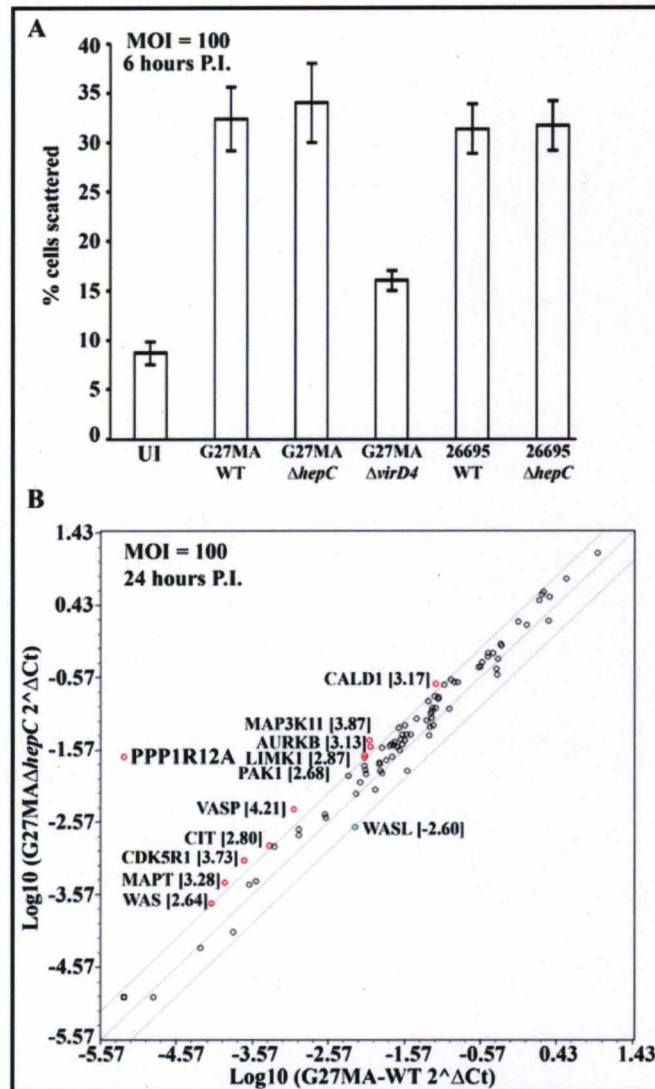


Figure 18: HepC likely targets the host cytoskeletal machinery during late infection in AGS cell culture model of infection.

(A) HepC does not mediate early cytoskeletal changes following infection.

AGS cells were infected at an MOI of 100 with indicated bacterial strains. *virD4* mutant is a Type IV secretion system mutant and is used as a control for cell scattering. 6hr post infection, cell extensions measuring over 40 μ meters were considered "scattered". No difference was observed in the scattering phenotype of AGS cells infected with WT or *hepC* mutant *H. pylori* strains. Experiment was done in triplicate and error bars represent standard deviations.

(B) HepC likely targets host cell cytoskeleton during later (24 hrs) period of AGS cell infection.

Scatter plot analysis comparing expression of cytoskeletal regulators following 24hrs post infection with Log₁₀ 2^{ΔCt} values of *G27MA hepC* mutant on y-axis and Log₁₀ 2^{ΔCt} values of *G27MA WT* on x-axis. Number of genes significantly up (≥ 2 fold) and down regulated (< 2 fold)

are shown in red and green circles, respectively. 10 genes were likely down regulated by HepC, 24 hr post AGS cell infection.

Genetically diverse *H. pylori* strains differentially dysregulate cytoskeletal regulators during an early infection of AGS cell line infection

Other investigators in our lab demonstrated that different geographic variants of HepC differ in their binding affinity to Ezrin. Given that Ezrin is a major cytoskeletal regulator that influences diverse cytoskeleton-dependent cell functions, that geographically distinct *H. pylori* strains interact differently with Ezrin, and that cytoskeletal dysregulation is a hallmark of *H. pylori* infection [83, 84], I then hypothesized that geographically distinct *H. pylori* strains likely differ in their ability to deregulate host cytoskeletal dynamics. To test this hypothesis, I first assessed for the humming bird phenotype (a consequence of cytoskeletal deregulation by *H. pylori*) of AGS cells infected with genetically diverse *H. pylori* strains. For this, I infected AGS cells with diverse *H. pylori* strains isolated from patients belonging to specific human populations: Amerindian (*Shi470*), Japan (*JS7*), Europe (*J99*), and cell culture adapted strain (*G27MA*), and monitored cell scattering phenotype, 6 hr post infection. Results from this experiment showed that *H. pylori* isolates from different geographic regions can differentially affect the cell scattering phenotype in AGS cells (p value < 0.05) (Fig. 19A & 19B). More importantly, Japanese *H. pylori* strain JS7 induced cell scattering at a higher rate as compared to infection with other tested strains. This finding is interesting in that cell scattering is a result of loss of cell-cell junctions, cytoskeletal modifications and eventual acquisition of motility, which are important in cancer progression and metastasis [84 - 86] and *H. pylori* mediated gastric cancer is most common in Japanese

populations [33, 34]. Next, I sought to further characterize and quantify the molecular basis of cytoskeletal deregulation by diverse *H. pylori* strains. To determine this, the expression of cytoskeletal regulators in AGS cells infected with diverse *H. pylori* strains for 3 hr and 6 hr, was compared with the expression of cytoskeletal regulators in uninfected AGS cells using cytoskeletal pathway specific PCR arrays, to identify genes that were differentially up or down regulated following infection. Analysis of the results obtained from this experiment clearly demonstrated the following key features (Fig. 19C):

1. Different genes were differentially up- or down-regulated at any given time point.
2. Even when the expression of the same gene was modulated by two strains, the level of modulation differs among distinct *H. pylori* strains (eg., VASP, which depending on the strain can be up regulated by as low as 2-fold to as high as 5.2-fold)
3. In addition, these experiments have identified several novel host candidate genes that are modulated during early infection with *H. pylori*. Monitoring novel signal transduction pathways should lead to better understanding of *H. pylori*'s pathogenesis.

Next, I verified PCR-array results, for cytoskeletal regulator VASP which was significantly up regulated (5.26X) in AGS cells following a 6 hr infection with *H. pylori* strain G27MA, using fluorescent activated cell sorting (FACS) analysis that compared VASP expression of uninfected AGS cells and AGS cells infected with *H. pylori* strain G27MA for 6 hr. Confirming the PCR array result, significant increase in VASP fluorescence i.e., VASP expression was observed in AGS cells infected with G27MA as

compared to uninfected AGS cells (Fig. 19D). This assay lends further credibility to the PCR array data and analysis that I performed above.

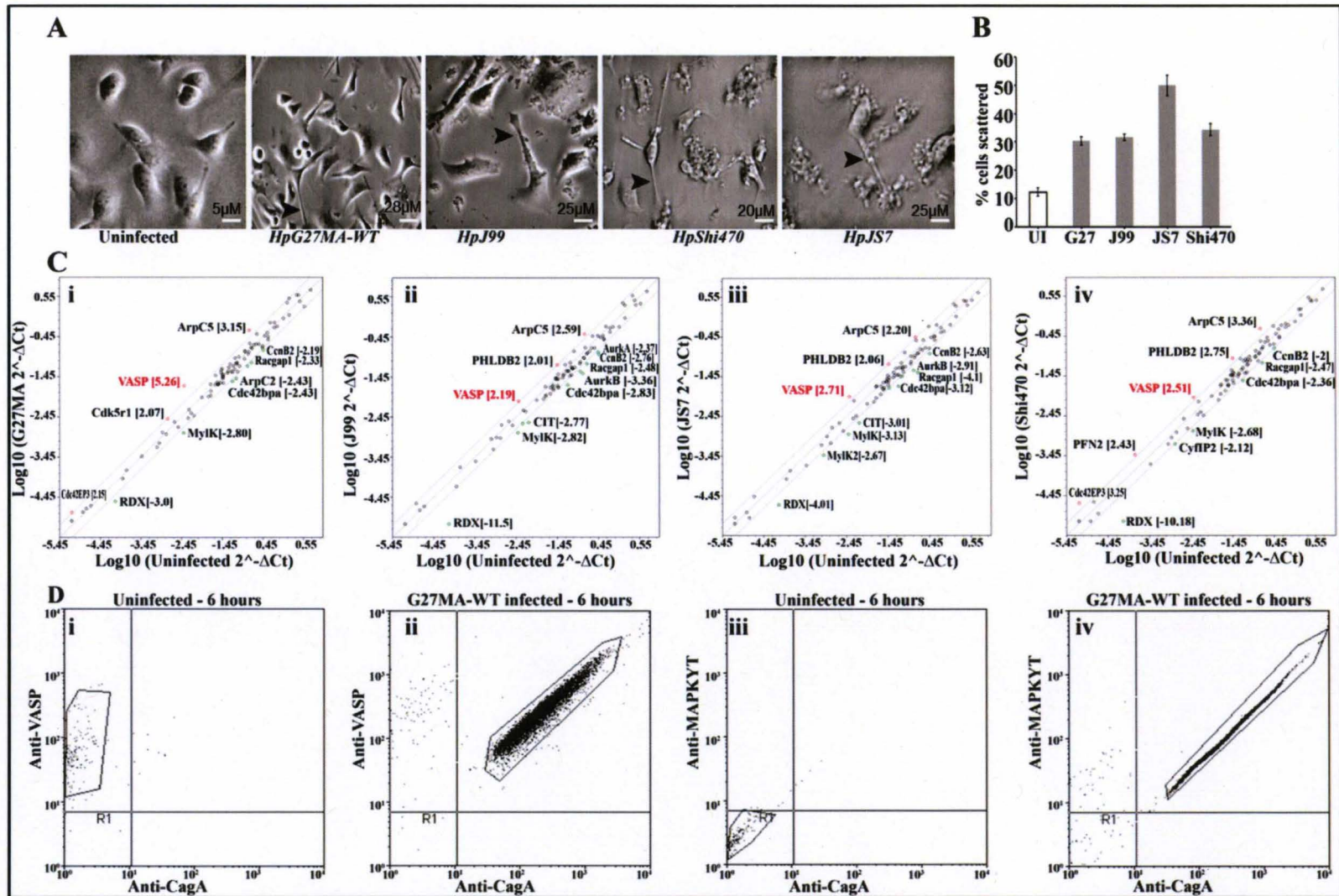


Figure 19: Differential dysregulation of cytoskeletal regulators following *H. pylori* infection with diverse *H. pylori* strains.

(A) Differential AGS cell scattering by diverse *H. pylori* strains.

Six hours post-infection, images were taken of the various wells containing AGS cells infected with indicated *H. pylori* strains. Cell extensions (arrow heads) measuring over 40 μ meters were considered “scattered” or “hummingbird” phenotype. 100 total cells in each field and three random fields were counted for each experiment. Experiment was done in triplicate and error bars represent standard deviations.

(B) Significant increase in the percentage cell scattering of AGS cells were observed following infection with diverse *H. pylori* strains with Japanese *H. pylori* strain causing the maximum affect.

(C) Molecular basis of cytoskeletal deregulation by *H. pylori*: Cytoskeletal targets are variably affected by diverse *H. pylori* strains.

Scatter plot analysis comparing expression of cytoskeletal regulators following 6hrs of infection with indicated *H. pylori* strains (on y-axis) with their expression in uninfected AGS cells (on x-axis). Number of genes significantly up (≥ 2 fold) and down regulated (< 2 fold) are shown in red and green circles, respectively. Designation of each gene was given right next to its position on the scatter plot.

(D) Significant increase of VASP expression in AGS cells following *H. pylori* G27MA infection.

(i) Depicts a dot plot representation of VASP fluorescence (boxed regions) in uninfected AGS cells and in AGS cells infected with G27MA-WT (ii). Shift of fluorescence to the right of the plot indicates increase in protein expression level; (iii) & (iv) depicts dot plots showing dynamics of CagA and phosphorylated MAPK in uninfected and G27MA infected AGS cells, respectively, which were monitored as controls to ensure the specificity of infection progression. MAPK activation is a hallmark of *H. pylori* infection and is observed only in infected AGS cells.

Summary II

1. HepC is biologically relevant during *H. pylori* infection and genetically diverse *H. pylori* strains tightly modulate *hepC* expression in AGS cell culture infection model.
2. HepC contributed significantly to *H. pylori*'s fitness.
3. HepC most likely targets host cytoskeletal machinery during late infection in a cell culture model of infection.
4. Genetically diverse *H. pylori* strains differentially dysregulate host cytoskeletal regulators during an early infection of AGS cell culture infection.

DISCUSSION AND FUTURE DIRECTIONS

The ability of *H. pylori* to chronically persist in human gastric mucosa along with striking geographic variation in the clinical outcome of infection suggests *H. pylori* strain specific exploitation and modulation of host responses. Emerging data suggests a possible role for proteins encoded by the *H. pylori slr* genes, in mediating and / or managing *H. pylori* - host interaction [42]. *H. pylori slr* genes encode secreted proteins with homology to the Sel-1 group of eukaryotic regulatory proteins that, through their interaction with other eukaryotic proteins, affect cell proliferation, apoptosis, immune response, and intracellular trafficking [43, 45]. Positive selection plays a dominant role in *H. pylori slr* gene family evolution, such that in any given *slr* protein different amino acids are favored in different geographical areas, and that the selection intensity is stronger on some *slr* genes than others in natural *H. pylori* populations [42]. Here, I extended this paradigm to *H. pylori slr* genes, *hcpC*, and *hcpG*. *hcpC* is present in all the natural *H. pylori* populations tested whereas *hcpG* is either absent, pseudogenized or, exhibited extreme polymorphisms. Different *hcpC* and *hcpG* codons evolved at different rates in different populations, although the intensity of selection to diversify is higher in *hcpG* evolution. Localization of adaptive residues to the molecular HcpC surface suggests that these may affect the affinity or specificity of its interaction with cognate host protein/ (s), fine tuning the host responses. Alternatively, some of the adaptive residues might also be

involved in immune escape, given that HcpC is a secreted and immunogenic molecule [54].

One unique feature of *H. pylori* is its outer membrane association of intrinsic cytoplasmic protein, HspB [132-135]. This feature was attributed to autolysis of bacteria and subsequent adsorption of extracellularly released HspB onto intact live bacteria [137]. However, it cannot be totally discounted that other accessory protein transporters, here HcpC, are involved in this process, given that a consistent defect in HspB surface localization is observed in G27MA Δ *hcpC* mutant strain, in a cell culture infection model. Furthermore, HcpC is able to rescue the HspB surface localization defect exhibited by G27MA Δ *hcpG* mutant strain early (3 hr) in cell culture infection model. Surface association of HspB was shown to induce humoral immune response and inflammation, leading to gastritis [138-143], and can also probably mask the integral intrinsic outer membrane proteins of *H. pylori* from host immune recognition [52, 137, 140, 144]. Thus, my observation has important implications in *H. pylori* pathogenesis by promoting long term survival of *H. pylori* in the host by evading immune surveillance. Exactly how HcpC is involved in surface translocation of HspB now merits detailed analysis. Furthermore, the G27MA Δ *hcpC* mutant, but not the G27MA Δ *hcpG* mutant, exhibited significant relative fitness reduction in a cell culture infection model. However, the combined effect of paralogs, *hcpC* and *hcpG* to *H. pylori* fitness is higher than the fitness contributions provided by each paralog individually. Taken together, these findings indicate that HcpC and HcpG are genetically redundant, but functionally non-reciprocal, and that HcpC and HcpG perform a crucial role in gastric epithelial cell infection. Given

these findings, I propose the following model for emergence, fixation and preservation of *hcpG* alleles in *H. pylori* genomes via gene duplication from ancestor *hcpC* (Figure 20).

Model for evolution of *hcpG* in *H. pylori* genomes:

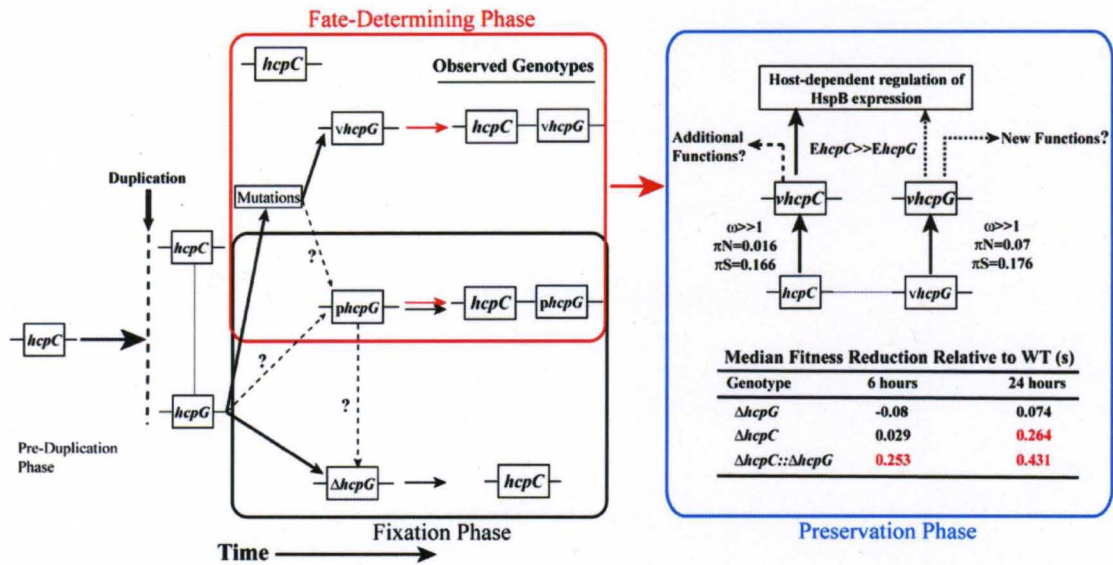


Figure 20: Evolution of stable, non-reciprocal genetic redundancy by diversifying selection following duplication and divergence in *H. pylori* Sel 1-like gene family.

Explanation of terms used:

$\Delta hcpG$: Deletion of *hcpG*; *phcpG*: Pseudogenized *hcpG* alleles; *vhcpG*: variant functional *hcpG* alleles; *EhcpC*: Efficiency of *hcpC* function; *EhcpG*: Efficiency of *hcpG* function; $\omega=d_N/d_S$; π_N = nucleotide diversity per non-synonymous site; π_S = nucleotide diversity per synonymous site.

Following gene duplication, the duplicated copy usually undergoes an initial fixation phase where the duplicate achieves fixation in the population if not lost by genetic drift [145]. Fixation phase is followed by the fate – determination phase, where the duplicated copy accumulates fate - determining mutations some of which can then be fixed by natural selection. The final phase is the preservation phase, where the fixed

change/s are maintained in the population [145]. In addition, selection pressures operating on the gene duplicates in each of the phases can significantly contribute to their preservation and functional evolution in genomes [145]. Based on the evolutionary and functional findings on paralogs, *hcpC*, and *hcpG* in this study, I propose the following model (Figure 20): Following the duplication event whereby *hcpC* duplicates to give rise to *hcpG*, fixation phase begins. In the fixation phase, *hcpG* is usually lost from the population as a consequence of genetic drift ($\Delta hcpG$). However, if *hcpG* escapes genetic drift mediated loss, it acquires random mutations in the fate determining phase. Such mutations, if they confer a functional advantage to *H. pylori*, are fixed in the genomes (*vhcpG*). If the mutations are not conferring an advantage, they will be pseudogenized (*phcpG*) and be lost from the populations ($\Delta hcpG$). However, some null mutations that can contribute to the improved fitness of *H. pylori*, are fixed in the genomes (evidence: retention of *phcpG* in some *H. pylori* genomes). Proceeding in to preservation phase, positive selection then operates on the duplicated paralog (*vhcpG*) that accumulated loss of function / degenerative mutations in fate - determining phase (evidence for accumulation of loss of function mutations by *hcpG*: HcpG is only required for optimal HspB expression during early infection, and is unable to compensate for the lack of HcpC during later phases, whereas HcpC is necessary and sufficient for optimal surface expression of HspB. Furthermore, the contribution of *hcpC* to *H. pylori* fitness, in the AGS cell culture infection model, is significantly greater than *hcpG*. However, both genes together demonstrated an additive effect on *H. pylori* fitness during 24 hrs post infection; $S_{\Delta hcpC} = 0.264$ vs. $S_{\Delta hcpG} = 0.074$, $P < 0.01$; $S_{\Delta hcpC}$ or $S_{\Delta hcpG}$ vs. $S_{\Delta hcpC::\Delta hcpG} = 0.431$, $P < 0.01$, where S =coefficient of median fitness reduction). Thus, HcpG

subfunctionalized in HcpC function, and is preserved in *H. pylori* genomes with a stable, genetically redundant, epistatic and overlapping yet non-reciprocal functional relationship with *hcpC*. There is also evidence of positive selection in the evolution of *hcpC* in the preservation phase (*hcpC*; $\omega > 1$). Yet, the selection intensity on *vhcpG* in the preservation phase is much higher than *hcpC* ($\omega_{hcpG} > \omega_{hcpC}$), indicating that *vhcpG* diversified or is diversifying significantly from the ancestor gene *hcpC* most probably for functional divergence, and that natural selection favored retention of the ancestral *hcpC* likely with an enhanced functional efficiency. Taken together, my data suggests a novel mechanism by which natural selection selects stable redundancy in duplicated genes. So, how does my data not fit in completely in any of the well-established models of gene duplication evolution? Since my data show that both *hcpC* and *hcpG* have evolved by strong diversifying selection in the preservation phase, I will briefly describe the models that take into account that positive selection is involved in the process of stable maintenance or preservation of both the duplicated genes, and explain how my data deviate from those models:

1. Duplication – degeneration – complementation (DDC model) [146]: Here, both the paralogs accumulate degenerative mutations that reduce the functional efficacy of both the duplicated genes. As a result, neither copy is sufficient to perform the original function and hence both the copies subfunctionalize, so that they both must be maintained by selection [146-148]. Moreover, symmetrical levels of polymorphisms and divergence among duplicated paralogs are expected in this model given that, both the paralogs undergo degenerative mutations to reduce the efficiency of the same function [145, 149, 150] . In contrast, here I

show that the fitness contribution of both the paralogs (*hcpC* and *hcpG*) together is significantly greater than the fitness contributions of each of the paralogs individually, that even though there is evidence of subfunctionalization in *HcpG*, natural selection favored retention of the ancestral *HcpC* function but not subfunctionalization, and that there is asymmetrical levels of non – synonymous levels of polymorphism (π_N) in the evolution of *hcpC* and *hcpG* ($\pi_{N \text{ hcpG}} > \pi_{N \text{ hcpC}}$).

2. Escape from adaptive conflict model (EAC): This model assumes that, single copy gene ancestral gene carries out distinct functions, and that it cannot improve one aspect of its performance without negatively affecting other aspects (adaptive conflict) [151-154]. This adaptive conflict is resolved when a duplication event gives one of the paralogs a chance to escape one of its roles. Therefore, duplication is accompanied by adaptive mutations in both duplicated genes to fine-tune their newly attained subfunctions, with signatures of positive selection on sequences released from adaptive conflict [151-154]. Moreover, the fitness contributions of both paralogues will be greater than the unduplicated paralog [143] and the polymorphism patterns will be similar to that of DDC model [143]. Even though my data is in harmony with this model with respect to the fitness dynamics and polymorphism patterns of the paralogs, it is the functional aspect where this model differs significantly. I show that instead of each of the paralog (*hcpC* and *hcpG*) diversifying and specializing in each of the different functions of ancestral gene, they both contribute to the same function i.e., HspB surface localization, albeit with different efficiencies.

3. Permanent heterozygote model: This model assumes that genetic variation already exists for a gene before the duplication event, that following gene duplication formation and fixation of a permanent heterozygote results in the achievement of higher fitness than either of the homozygotes in the pre - duplication phase, and that high levels of polymorphisms exist before the gene duplication event [145, 155, 156]. To consider this model for *hcpG* evolution in *H. pylori* genomes, the main assumption would be that following duplication the permanent heterozygote (here, genomes with both *hcpC* and *hcpG*) will be fixed in the populations. My data clearly contradict this assumption in that nearly 50% of *H. pylori* populations tested lack *hcpG*.
4. Multi – allelic diversifying selection: The main assumption of this model is that the functional attributes of genes under multi allelic diversifying selection, requires the genes to evolve constantly and rapidly (eg., major histocompatibility genes) for functional divergence, and hence positive selection favors fixation of new copy [145, 157, 158]. However, my data show that HcpG is subfunctionalized to HcpC function rather than acquiring a new function.

In an attempt to delineate the function / (s) of HcpG, I cloned and expressed the five unique *hcpG* allelic variants as 6 X-Histidine fusion proteins. Attempts at purifying HcpG::His, to identify the interacting host partner / (s) are still in process.

Studies in our lab have determined that *H. pylori* SLR protein HepC interacts directly with multifunctional human cytoskeletal protein, Ezrin and that the Japanese HepC variant exhibited a greater affinity for Ezrin interaction when compared to its European counterpart, suggesting that such interaction differences likely manifest in

altered Ezrin-dependent signal transduction which may directly influence the progression and thus the severity of gastric disease outcome in Japanese populations (Figure 21). It is in this context that I characterized the biological significance of HepC in *H. pylori* pathogenesis. I showed that the different geographic variants of *hepC* differ in their expression patterns in the AGS cell culture infection model, with Japanese *hepC* exhibiting maximal up regulation following infection. This finding is interesting given that HepC – Ezrin interaction affinity is also the strongest in Japanese *H. pylori* populations. Whether such increased binding affinity can be directly correlated with increased *hepC* expression in Japanese *H. pylori* populations needs to be fully characterized. HepC contributes significantly to *H. pylori*'s fitness and most likely targets host cytoskeletal machinery by inhibiting key components of the human cytoskeletal machinery, during late infection in the cell culture model. It will be noteworthy to elucidate whether such cytoskeletal deregulation is mediated by HepC – Ezrin interaction or, by other yet unidentified mechanisms of HepC – host mediated interactions. Furthermore, genetically diverse *H. pylori* strains differentially impact key cytoskeletal regulators during relatively early cell culture infection model. Yet more, amplitude of expression of the identical cytoskeletal regulators varies following infection with diverse *H. pylori* strains. It is tempting to consider that such dysregulation of cytoskeletal regulators following infection with diverse *H. pylori* strains can affect the downstream events in two different ways: 1) Dramatic: modulation and alteration of different signaling pathways, 2) Subtle: fine tuning identical signaling pathways. These findings should help develop a population - based framework in understanding molecular events behind geographically variable clinical outcome of gastric disease, and in developing

population specific biomarkers to predict the outcome of *H. pylori* infection. In addition, these experiments have identified several novel host candidate genes that are modulated during early infection with *H. pylori* and thus the need to monitor novel signal transduction pathways to better understand *H. pylori*'s pathogenesis.

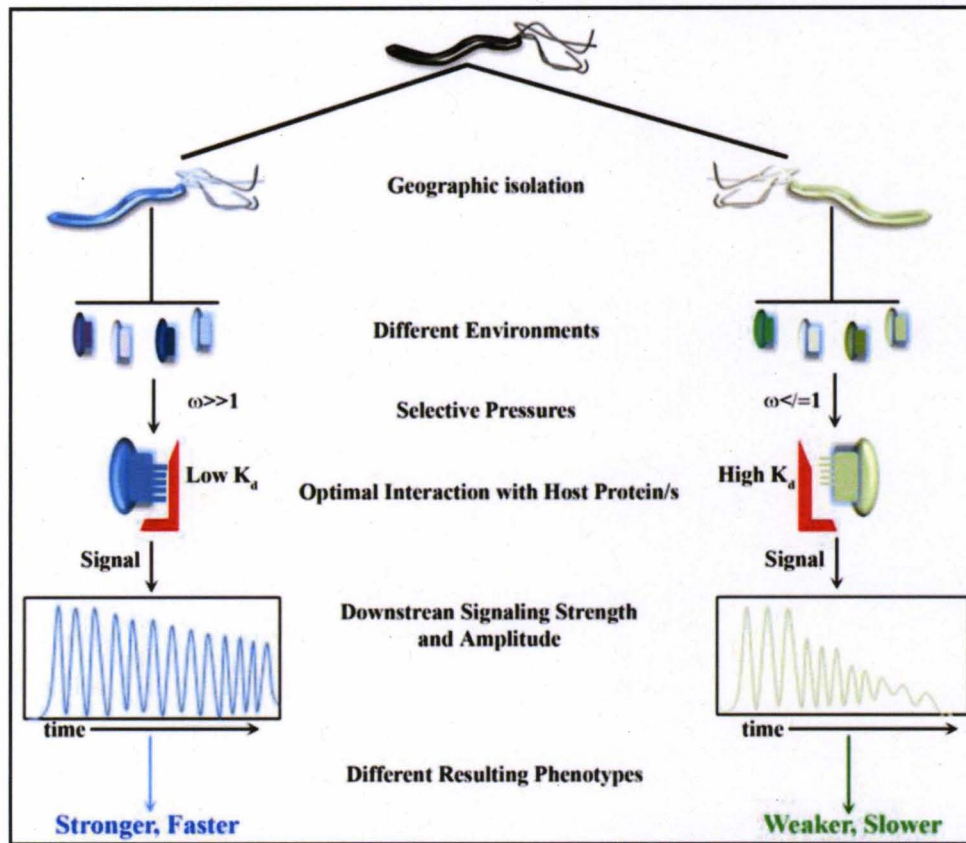


Figure 21: Causes and consequences of *H. pylori* molecular evolution.

Taken together, of the myriad of positive selection outcomes, I specifically showed that positive selection mediated genetic redundancy provides a fitness advantage to *H. pylori* (Figure 20), and that positive selection mediated modulation of *H. pylori* – host interactions likely alters the strength, duration or amplitude of downstream signaling events (Figure 21). I speculate that some of these differences may affect the progression

of *H. pylori* infection, and thereby the clinical outcome in different geographical regions. I conclude that, further search for, and analysis of, *H. pylori*'s determinants that have been subject to positive selection in particular lineages should elucidate mechanisms important in establishment and maintenance of chronic infection and disease, and perhaps provide new insights into effective management or eradication of these infections in diverse human populations.

REFERENCES

1. Falush, D., et al., *Recombination and mutation during long-term gastric colonization by Helicobacter pylori: estimates of clock rates, recombination size, and minimal age*. Proc Natl Acad Sci U S A, 2001. **98**(26): p. 15056-61.
2. Linz, B., et al., *An African origin for the intimate association between humans and Helicobacter pylori*. Nature, 2007. **445**(7130): p. 915-8.
3. Banatvala, N., et al., *The cohort effect and Helicobacter pylori*. J Infect Dis, 1993. **168**(1): p. 219-21.
4. Kersulyte, D., H. Chalkauskas, and D.E. Berg, *Emergence of recombinant strains of Helicobacter pylori during human infection*. Mol Microbiol, 1999. **31**(1): p. 31-43.
5. Schwarz, S., et al., *Horizontal versus familial transmission of Helicobacter pylori*. PLoS Pathog, 2008. **4**(10): p. e1000180.
6. Rothenbacher, D. and H. Brenner, *Burden of Helicobacter pylori and H. pylori-related diseases in developed countries: recent developments and future implications*. Microbes Infect, 2003. **5**(8): p. 693-703.
7. Suerbaum, S. and C. Josenhans, *Helicobacter pylori evolution and phenotypic diversification in a changing host*. Nat Rev Microbiol, 2007. **5**(6): p. 441-52.
8. *Schistosomes, liver flukes and Helicobacter pylori*. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7-14 June 1994. IARC Monogr Eval Carcinog Risks Hum, 1994. **61**: p. 1-241.
9. Mbulaiteye, S.M., M. Hisada, and E.M. El-Omar, *Helicobacter Pylori associated global gastric cancer burden*. Front Biosci, 2009. **14**: p. 1490-504.
10. Kate, V., et al., *Prevalence of Helicobacter pylori infection in disorders of the upper gastrointestinal tract in south India*. Natl Med J India, 1998. **11**(1): p. 5-8.
11. Liu, Y., et al., *Geographic pathology of Helicobacter pylori gastritis*. Helicobacter, 2005. **10**(2): p. 107-13.

12. Roder, D.M., *The epidemiology of gastric cancer*. Gastric Cancer, 2002. 5(Supplement 1): p. 5 - 11.
13. Fenger, H.J., Gudmand-Hoyer, E., *Peptic ulcer in Greenland Inuit: evidence for a low prevalence of duodenal ulcer*. Int J Circumpolar Health. , 1997. 56(3): p. 64-69.
14. Dwyer, B., Sun, N.X., Kaldor, J., Tee, W., Lambert, J., Luppino, M., Flannery, G., *Antibody response to Campylobacter pylori in an ethnic group lacking peptic ulceration*. Scand J Infect Dis. , 1988. 20(1): p. 63-68.
15. Cover, T.L., et al., *H. pylori Pathogenesis*. Principles of Bacterial Pathogenesis ed. E.A. Groisman. 2001, New York: Academic Press.
16. Blaser, M.J. and J.C. Atherton, *Helicobacter pylori persistence: biology and disease*. J. Clin. Invest., 2004. 113(3): p. 321-333.
17. Covacci, A., et al., *Helicobacter pylori Virulence and Genetic Geography*. Science, 1999. 284(5418): p. 1328-1333.
18. Go, M.F., et al., *Population genetic analysis of Helicobacter pylori by multilocus enzyme electrophoresis: extensive allelic diversity and recombinational population structure*. J. Bacteriol., 1996. 178(13): p. 3934-3938.
19. Akopyanz, N., Bukanov, N.O., Westblom, T.U., Kresovich S., and Berg, D.E., *DNA diversity among clinical isolates of Helicobacter pylori detected by PCR-based RAPD fingerprinting*. Nucleic Acids Research, 1992. 20(19): p. 5137-5142.
20. Feil, E.J. and B.G. Spratt, *RECOMBINATION AND THE POPULATION STRUCTURES OF BACTERIAL PATHOGENS*. Annual Review of Microbiology, 2001. 55(1): p. 561-590.
21. Kersulyte, D., et al., *Differences in Genotypes of Helicobacter pylori from Different Human Populations*. J. Bacteriol., 2000. 182(11): p. 3210-3218.
22. Aspholm-Hurtig, M., et al., *Functional Adaptation of BabA, the H. pylori ABO Blood Group Antigen Binding Adhesin*. Science, 2004. 305(5683): p. 519-522.
23. Albert, T.J., et al., *Mutation discovery in bacterial genomes: metronidazole resistance in Helicobacter pylori*. Nat Meth, 2005. 2(12): p. 951.

24. Achtman, M., et al., *Recombination and clonal groupings within Helicobacter pylori from different geographical regions*. *Molecular Microbiology*, 1999. **32**(3): p. 459-470.
25. Falush, D., et al., *Traces of Human Migrations in Helicobacter pylori Populations*. *Science*, 2003. **299**(5612): p. 1582-1585.
26. Kalia, A., et al., *Evolutionary Dynamics of Insertion Sequences in Helicobacter pylori*. *J. Bacteriol.*, 2004. **186**(22): p. 7508-7520.
27. Cooke, C.L., J.L. Huff, and J.V. Solnick, *The role of genome diversity and immune evasion in persistent infection with Helicobacter pylori*. *FEMS Immunology and Medical Microbiology*, 2005. **45**(1): p. 11.
28. Y. Yamaoka, M.S.O., A. R. Sepulveda, O. Gutierrez, N. Figura, J.G. Kim, T. Kodama, K. Kashima, D.Y. Graham *Molecular epidemiology of Helicobacter pylori: separation of H. pylori from East Asian and non-Asian countries*. *Epidemiol Infect.* , 200. **124**(1): p. 91-96.
29. Yamazaki, S., et al., *Distinct Diversity of vacA, cagA, and cagE Genes of Helicobacter pylori Associated with Peptic Ulcer in Japan*. *J. Clin. Microbiol.*, 2005. **43**(8): p. 3906-3916.
30. van Doorn, L.-J., et al., *Expanding Allelic Diversity of Helicobacter pylori vacA*. *J. Clin. Microbiol.*, 1998. **36**(9): p. 2597-2603.
31. Atherton, J.C., et al., *Mosaicism in Vacuolating Cytotoxin Alleles of Helicobacter pylori*. *J. Biol. Chem.*, 1995. **270**(30): p. 17771-17777.
32. Atherton, J.C., *THE PATHOGENESIS OF HELICOBACTER PYLORI-INDUCED GASTRO-DUODENAL DISEASES*. *Annual Review of Pathology: Mechanisms of Disease*, 2006. **1**(1): p. 63-96.
33. Higashi, H., et al., *Biological activity of the Helicobacter pylori virulence factor CagA is determined by variation in the tyrosine phosphorylation sites*. *PNAS*, 2002. **99**(22): p. 14428-14433.
34. Higashi, H., et al., *SHP-2 Tyrosine Phosphatase as an Intracellular Target of Helicobacter pylori CagA Protein*. *Science*, 2002. **295**(5555): p. 683-686.
35. Wang, J., et al., *Regional Variation among vacA Alleles of Helicobacter pylori in China*. *J. Clin. Microbiol.*, 2003. **41**(5): p. 1942-1945.

36. Wang, G., M.Z. Humayun, and D.E. Taylor, *Mutation as an origin of genetic variability in Helicobacter pylori*. Trends Microbiol, 1999. 7(12): p. 488-93.
37. Aras, R.A., et al., *Extensive repetitive DNA facilitates prokaryotic genome plasticity*. Proc Natl Acad Sci U S A, 2003. 100(23): p. 13579-84.
38. Wright, S., *Evolution and the genetics of populations*. Experimental results and evolutionary deductions. Vol. 3. 1977, Chicago: The University of Chicago Press.
39. Graur, D. and W.-H. Li, *Fundamentals of Evolution*. Second ed. -. 2000, Sunderland, MA: Sinauer Associates, Inc.
40. Yang, Z., *Inference of selection from multiple species alignments*. Current Opinion in Genetics & Development, 2002. 12(6): p. 688.
41. Ilver, D., et al., *Helicobacter pylori Adhesin Binding Fucosylated Histo-Blood Group Antigens Revealed by Retagging*. Science, 1998. 279(5349): p. 373-377.
42. Ogura, M., et al., *Helicobacter pylori evolution: lineage- specific adaptations in homologs of eukaryotic Sell-like genes*. PLoS Comput Biol, 2007. 3(8): p. e151.
43. Grant, B. and I. Greenwald, *The Caenorhabditis elegans sel-1 Gene, a Negative Regulator of lin-12 and glp-1, Encodes a Predicted Extracellular Protein*. Genetics, 1996. 143(1): p. 237-247.
44. Amsen, D., et al., *Instruction of Distinct CD4 T Helper Cell Fates by Different Notch Ligands on Antigen-Presenting Cells*. Cell, 2004. 117(4): p. 515.
45. Cattaneo, M., et al., *Identification of a region within SEL1L protein required for tumour growth inhibition*. Gene, 2004. 326: p. 149.
46. Neuber, O., et al., *Ubx2 links the Cdc48 complex to ER-associated protein degradation*. Nat Cell Biol, 2005. 7(10): p. 993.
47. Lee, C.S., et al., *Neurogenin 3 is essential for the proper specification of gastric enteroendocrine cells and the maintenance of gastric epithelial cell identity*. Genes Dev., 2002. 16(12): p. 1488-1497.
48. Sjolund, J., et al., *The Notch pathway in cancer: Differentiation gone awry*. European Journal of Cancer, 2005. 41(17): p. 2620.
49. Zhang, H., et al., *Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry*. Nat Biotechnol, 2003. 21(6): p. 660-6.

50. Gardner, R.G., et al., *Endoplasmic reticulum degradation requires lumen to cytosol signaling. Transmembrane control of Hrd1p by Hrd3p.* J Cell Biol, 2000. **151**(1): p. 69-82.
51. Okabe, M., T. Yakushi, and M. Homma, *Interactions of MotX with MotY and with the PomA/PomB sodium ion channel complex of the Vibrio alginolyticus polar flagellum.* J Biol Chem, 2005. **280**(27): p. 25659-64.
52. Sabarth, N., et al., *Identification of surface proteins of Helicobacter pylori by selective biotinylation, affinity purification, and two-dimensional gel electrophoresis.* J Biol Chem, 2002. **277**(31): p. 27896-902.
53. Bumann, D., et al., *Proteome analysis of secreted proteins of the gastric pathogen Helicobacter pylori.* Infect Immun, 2002. **70**(7): p. 3396-403.
54. Haas, G., et al., *Immunoproteomics of Helicobacter pylori infection and relation to gastric disease.* Proteomics, 2002. **2**(3): p. 313-24.
55. Cirillo, S.L., J. Lum, and J.D. Cirillo, *Identification of novel loci involved in entry by Legionella pneumophila.* Microbiology, 2000. **146** (Pt 6): p. 1345-59.
56. Newton, H.J., et al., *Identification of Legionella pneumophila-specific genes by genomic subtractive hybridization with Legionella micdadei and identification of lpnE, a gene required for efficient host cell entry.* Infect Immun, 2006. **74**(3): p. 1683-91.
57. Reed, J.W., J. Glazebrook, and G.C. Walker, *The exoR gene of Rhizobium meliloti affects RNA levels of other exo genes but lacks homology to known transcriptional regulators.* J Bacteriol, 1991. **173**(12): p. 3789-94.
58. Ponting, C.P., et al., *Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer.* J Mol Biol, 1999. **289**(4): p. 729-45.
59. Ponting, C.P., *Proteins of the endoplasmic-reticulum-associated degradation pathway: domain detection and function prediction.* Biochem J, 2000. **351** Pt 2: p. 527-35.
60. Mittl, P.R. and W. Schneider-Brachert, *Sell-like repeat proteins in signal transduction.* Cell Signal, 2007. **19**(1): p. 20-31.

61. Alm, R.A., et al., *Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori*. Nature, 1999. **397**(6715): p. 176.
62. Tomb, J.-F., et al., *The complete genome sequence of the gastric pathogen Helicobacter pylori*. Nature, 1997. **388**(6642): p. 539.
63. Ponting, C.P., et al., *Eukaryotic Signalling Domain Homologues in Archaea and Bacteria. Ancient Ancestry and Horizontal Gene Transfer*. Journal of Molecular Biology, 1999. **289**(4): p. 729.
64. Mittl, P.R.E., et al., *The Cysteine-rich Protein A from Helicobacter pylori Is a beta -Lactamase*. J. Biol. Chem., 2000. **275**(23): p. 17693-17699.
65. Cao, P., et al., *Extracellular Release of Antigenic Proteins by Helicobacter pylori*. Infect. Immun., 1998. **66**(6): p. 2984-2986.
66. Krishnamurthy, P., et al., *Identification of a novel penicillin-binding protein from Helicobacter pylori*. J Bacteriol, 1999. **181**(16): p. 5107-10.
67. Deml, L., et al., *Characterization of the Helicobacter pylori cysteine-rich protein A as a T-helper cell type 1 polarizing agent*. Infect Immun, 2005. **73**(8): p. 4732-42.
68. Dumrese, C., et al., *The secreted Helicobacter cysteine-rich protein A causes adherence of human monocytes and differentiation into a macrophage-like phenotype*. FEBS Letters, 2009. **583**(10): p. 1637-1643.
69. Hocking, D., et al., *Isolation of recombinant protective Helicobacter pylori antigens*. Infect Immun, 1999. **67**(9): p. 4713-9.
70. Parkhill, J., et al., *The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences*. Nature, 2000. **403**(6770): p. 665.
71. Suerbaum, S., et al., *The complete genome sequence of the carcinogenic bacterium Helicobacter hepaticus*. PNAS, 2003. **100**(13): p. 7901-7906.
72. Baar, C., et al., *Complete genome sequence and analysis of Wolinella succinogenes*. PNAS, 2003. **100**(20): p. 11690-11695.
73. Perry, G.H., et al., *Diet and the evolution of human amylase gene copy number variation*. Nat Genet, 2007. **39**(10): p. 1256-60.

74. Kimura, M., *The neutral theory of molecular evolution*. Sci Am, 1979. **241**(5): p. 98-100, 102, 108 passim.
75. Ohta, T., *Mechanisms of molecular evolution*. Philos Trans R Soc Lond B Biol Sci, 2000. **355**(1403): p. 1623-6.
76. Hughes, A.L., *Near neutrality: leading edge of the neutral theory of molecular evolution*. Ann N Y Acad Sci, 2008. **1133**: p. 162-79.
77. Jukes, T.H., *The neutral theory of molecular evolution*. Genetics, 2000. **154**(3): p. 956-8.
78. Takahata, N., *Neutral theory of molecular evolution*. Curr Opin Genet Dev, 1996. **6**(6): p. 767-72.
79. Kimura, M., *The neutral theory of molecular evolution and the world view of the neutralists*. Genome, 1989. **31**(1): p. 24-31.
80. Nei, M., Y. Suzuki, and M. Nozawa, *The neutral theory of molecular evolution in the genomic era*. Annu Rev Genomics Hum Genet, 2010. **11**: p. 265-89.
81. Kimura, M., *The neutral theory of molecular evolution: a review of recent evidence*. Jpn J Genet, 1991. **66**(4): p. 367-86.
82. Li, W.H. and T. Gojobori, *Rapid evolution of goat and sheep globin genes following gene duplication*. Mol Biol Evol, 1983. **1**(1): p. 94-108.
83. Asahi, M., et al., *Helicobacter pylori CagA protein can be tyrosine phosphorylated in gastric epithelial cells*. J Exp Med, 2000. **191**(4): p. 593-602.
84. Ridley, A.J., et al., *Cell migration: integrating signals from front to back*. Science, 2003. **302**(5651): p. 1704-9.
85. Broussard, J.A., D.J. Webb, and I. Kaverina, *Asymmetric focal adhesion disassembly in motile cells*. Curr Opin Cell Biol, 2008. **20**(1): p. 85-90.
86. Boyer, B., A.M. Valles, and N. Edme, *Induction and regulation of epithelial-mesenchymal transitions*. Biochem Pharmacol, 2000. **60**(8): p. 1091-9.
87. Lo, S.H., *Focal adhesions: what's new inside*. Dev Biol, 2006. **294**(2): p. 280-91.
88. Kwok, T., et al., *Helicobacter exploits integrin for type IV secretion and kinase activation*. Nature, 2007. **449**(7164): p. 862-6.
89. Mitra, S.K. and D.D. Schlaepfer, *Integrin-regulated FAK-Src signaling in normal and cancer cells*. Curr Opin Cell Biol, 2006. **18**(5): p. 516-23.

90. Parsons, J.T., et al., *Focal adhesion kinase: a regulator of focal adhesion dynamics and cell movement*. *Oncogene*, 2000. **19**(49): p. 5606-13.
91. Caron-Lormier, G. and H. Berry, *Amplification and oscillations in the FAK/Src kinase system during integrin signaling*. *J Theor Biol*, 2005. **232**(2): p. 235-48.
92. Higashi, H., et al., *SHP-2 tyrosine phosphatase as an intracellular target of Helicobacter pylori CagA protein*. *Science*, 2002. **295**(5555): p. 683-6.
93. Yamazaki, S., et al., *The CagA protein of Helicobacter pylori is translocated into epithelial cells and binds to SHP-2 in human gastric mucosa*. *J Infect Dis*, 2003. **187**(2): p. 334-7.
94. Backert, S., et al., *Phosphorylation of tyrosine 972 of the Helicobacter pylori CagA protein is essential for induction of a scattering phenotype in gastric epithelial cells*. *Mol Microbiol*, 2001. **42**(3): p. 631-44.
95. Selbach, M., et al., *Src is the kinase of the Helicobacter pylori CagA protein in vitro and in vivo*. *J Biol Chem*, 2002. **277**(9): p. 6775-8.
96. Stein, M., et al., *c-Src/Lyn kinases activate Helicobacter pylori CagA through tyrosine phosphorylation of the EPIYA motifs*. *Mol Microbiol*, 2002. **43**(4): p. 971-80.
97. Higashi, H., et al., *Helicobacter pylori CagA induces Ras-independent morphogenetic response through SHP-2 recruitment and activation*. *J Biol Chem*, 2004. **279**(17): p. 17205-16.
98. Wessler, S., et al., *B-Raf/Rap1 signaling, but not c-Raf-1/Ras, induces the histidine decarboxylase promoter in Helicobacter pylori infection*. *FASEB J*, 2002. **16**(3): p. 417-9.
99. Suzuki, M., et al., *Interaction of CagA with Crk plays an important role in Helicobacter pylori-induced loss of gastric epithelial cell adhesion*. *J Exp Med*, 2005. **202**(9): p. 1235-47.
100. Brandt, S., et al., *Use of a novel coinfection system reveals a role for Rac1, H-Ras, and CrkII phosphorylation in Helicobacter pylori-induced host cell actin cytoskeletal rearrangements*. *FEMS Immunol Med Microbiol*, 2007. **50**(2): p. 190-205.

101. Cory, G.O. and A.J. Ridley, *Cell motility: braking WAVES*. Nature, 2002. **418**(6899): p. 732-3.
102. Al-Ghoul, L., et al., *Analysis of the type IV secretion system-dependent cell motility of Helicobacter pylori-infected epithelial cells*. Biochem Biophys Res Commun, 2004. **322**(3): p. 860-6.
103. Moese, S., et al., *Helicobacter pylori induces AGS cell motility and elongation via independent signaling pathways*. Infect Immun, 2004. **72**(6): p. 3646-9.
104. Weydig, C., et al., *CagA-independent disruption of adherence junction complexes involves E-cadherin shedding and implies multiple steps in Helicobacter pylori pathogenicity*. Exp Cell Res, 2007. **313**(16): p. 3459-71.
105. Churin, Y., et al., *Pathogenicity island-dependent activation of Rho GTPases Rac1 and Cdc42 in Helicobacter pylori infection*. Mol Microbiol, 2001. **40**(4): p. 815-23.
106. Ferrero, R.L., et al., *Construction of isogenic urease-negative mutants of Helicobacter pylori by allelic exchange*. J Bacteriol, 1992. **174**(13): p. 4212-7.
107. Sainsus, N., et al., *Liquid culture medium for the rapid cultivation of Helicobacter pylori from biopsy specimens*. Eur J Clin Microbiol Infect Dis, 2008. **27**(12): p. 1209-17.
108. Kitsos, C.M. and C.T. Stadlander, *Helicobacter pylori in liquid culture: evaluation of growth rates and ultrastructure*. Curr Microbiol, 1998. **37**(2): p. 88-93.
109. Kiss, J., M. Szabo, and F. Olsz, *Site-specific recombination by the DDE family member mobile element IS30 transposase*. Proc Natl Acad Sci U S A, 2003. **100**(25): p. 15000-5.
110. Kalia, A., et al., *Evolutionary dynamics of insertion sequences in Helicobacter pylori*. J Bacteriol, 2004. **186**(22): p. 7508-20.
111. Posada, D. and K.A. Crandall, *MODELTEST: testing the model of DNA substitution*. Bioinformatics, 1998. **14**(9): p. 817-8.
112. Huelsenbeck, J.P. and B. Rannala, *Phylogenetic methods come of age: testing hypotheses in an evolutionary context*. Science, 1997. **276**(5310): p. 227-32.

113. Yang, Z.H., et al., *Codon-substitution models for heterogeneous selection pressure at amino acid sites*. Genetics, 2000. **155**(1): p. 431-449.
114. Yang, Z., *Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution*. Mol Biol Evol, 1998. **15**(5): p. 568-73.
115. Dailidiene, D., et al., *Contraselectable streptomycin susceptibility determinant for genetic manipulation and analysis of Helicobacter pylori*. Appl Environ Microbiol, 2006. **72**(9): p. 5908-14.
116. Torii, N., et al., *Spontaneous mutations in the Helicobacter pylori rpsL gene*. Mutat Res, 2003. **535**(2): p. 141-5.
117. Chalker, A.F., et al., *Systematic identification of selective essential genes in Helicobacter pylori by genome prioritization and allelic replacement mutagenesis*. J Bacteriol, 2001. **183**(4): p. 1259-68.
118. Tan, S. and D.E. Berg, *Motility of urease-deficient derivatives of Helicobacter pylori*. J Bacteriol, 2004. **186**(3): p. 885-8.
119. Aspholm, M., et al., *Helicobacter pylori adhesion to carbohydrates*. Methods Enzymol, 2006. **417**: p. 293-339.
120. Mimuro, H., et al., *Grb2 is a key mediator of helicobacter pylori CagA protein activities*. Mol Cell, 2002. **10**(4): p. 745-55.
121. Davis, K.M., et al., *Resistance to mucosal lysozyme compensates for the fitness deficit of peptidoglycan modifications by Streptococcus pneumoniae*. PLoS Pathog, 2008. **4**(12): p. e1000241.
122. de Vries, N., et al., *The stress-induced hsp12 gene shows genetic variation among Helicobacter pylori strains*. FEMS Immunol Med Microbiol, 2003. **38**(1): p. 45-51.
123. BELL, G., *Selection: The mechanism of evolution*. 1949.
124. Fischer, W., et al., *Systematic mutagenesis of the Helicobacter pylori cag pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8*. Mol Microbiol, 2001. **42**(5): p. 1337-48.
125. Keates, S., et al., *Differential activation of mitogen-activated protein kinases in AGS gastric epithelial cells by cag+ and cag- Helicobacter pylori*. J Immunol, 1999. **163**(10): p. 5552-9.

126. Ding, S.Z., M.F. Smith, Jr., and J.B. Goldberg, *Helicobacter pylori* and mitogen-activated protein kinases regulate the cell cycle, proliferation and apoptosis in gastric epithelial cells. *J Gastroenterol Hepatol*, 2008. **23**(7 Pt 2): p. e67-78.
127. Biunno, I., et al., *SEL1L*, the human homolog of *C. elegans sel-1*: refined physical mapping, gene structure and identification of polymorphic markers. *Hum Genet*, 2000. **106**(2): p. 227-35.
128. Cattaneo, M., et al., *Cloning and functional analysis of SEL1L promoter region, a pancreas-specific gene*. *DNA Cell Biol*, 2001. **20**(1): p. 1-9.
129. Cattaneo, M., et al., *The expression of SEL1L and TAN-1 in normal and neoplastic cells*. *Int J Biol Markers*, 2000. **15**(1): p. 26-32.
130. Orlandi, R., et al., *SEL1L expression decreases breast tumor cell aggressiveness in vivo and in vitro*. *Cancer Res*, 2002. **62**(2): p. 567-74.
131. Luthy, L., M.G. Grutter, and P.R.E. Mittl, *The crystal structure of helicobacter cysteine-rich protein C at 2.0 angstrom resolution: Similar peptide-binding sites in TPR and SEL1-like repeat proteins*. *Journal of Molecular Biology*, 2004. **340**(4): p. 829-841.
132. Bode, G., et al., *Ultrastructural localization of urease of Helicobacter pylori*. *Med Microbiol Immunol*, 1993. **182**(5): p. 233-42.
133. Eschweiler, B., et al., *In situ localization of the 60 k protein of Helicobacter pylori, which belongs to the family of heat shock proteins, by immuno-electron microscopy*. *Zentralbl Bakteriologie*, 1993. **280**(1-2): p. 73-85.
134. Hawtin, P.R., A.R. Stacey, and D.G. Newell, *Investigation of the structure and localization of the urease of Helicobacter pylori using monoclonal antibodies*. *J Gen Microbiol*, 1990. **136**(10): p. 1995-2000.
135. Spiegelhalder, C., et al., *Purification of Helicobacter pylori superoxide dismutase and cloning and sequencing of the gene*. *Infect Immun*, 1993. **61**(12): p. 5315-25.
136. Evans, D.J., Jr., et al., *Urease-associated heat shock protein of Helicobacter pylori*. *Infect Immun*, 1992. **60**(5): p. 2125-7.
137. Phadnis, S.H., et al., *Surface localization of Helicobacter pylori urease and a heat shock protein homolog requires bacterial autolysis*. *Infect Immun*, 1996. **64**(3): p. 905-12.

138. Craig, P.M., et al., *Helicobacter pylori* secretes a chemotactic factor for monocytes and neutrophils. *Gut*, 1992. **33**(8): p. 1020-3.
139. Mooney, C., et al., *Neutrophil activation by Helicobacter pylori*. *Gut*, 1991. **32**(8): p. 853-7.
140. Dunn, B.E., G.I. Perez-Perez, and M.J. Blaser, *Two-dimensional gel electrophoresis and immunoblotting of Campylobacter pylori proteins*. *Infect Immun*, 1989. **57**(6): p. 1825-33.
141. Macchia, G., et al., *The Hsp60 protein of Helicobacter pylori: structure and immune response in patients with gastroduodenal diseases*. *Mol Microbiol*, 1993. **9**(3): p. 645-52.
142. Mai, U.E., et al., *Surface proteins from Helicobacter pylori exhibit chemotactic activity for human leukocytes and are present in gastric mucosa*. *J Exp Med*, 1992. **175**(2): p. 517-25.
143. Rautelin, H., et al., *Incidence of Helicobacter pylori strains activating neutrophils in patients with peptic ulcer disease*. *Gut*, 1993. **34**(5): p. 599-603.
144. Doig, P. and T.J. Trust, *Identification of surface-exposed outer membrane antigens of Helicobacter pylori*. *Infect Immun*, 1994. **62**(10): p. 4526-33.
145. Innan, H. and F. Kondrashov, *The evolution of gene duplications: classifying and distinguishing between models*. *Nat Rev Genet*, 2010. **11**(2): p. 97-108.
146. Force, A., et al., *Preservation of duplicate genes by complementary, degenerative mutations*. *Genetics*, 1999. **151**(4): p. 1531-45.
147. Lynch, M. and A. Force, *The probability of duplicate gene preservation by subfunctionalization*. *Genetics*, 2000. **154**(1): p. 459-73.
148. Lynch, M., et al., *The probability of preservation of a newly arisen gene duplicate*. *Genetics*, 2001. **159**(4): p. 1789-804.
149. Nielsen, R., *Molecular signatures of natural selection*. *Annu Rev Genet*, 2005. **39**: p. 197-218.
150. Kreitman, M., *Methods to detect selection in populations with applications to the human*. *Annu Rev Genomics Hum Genet*, 2000. **1**: p. 539-59.

151. Wen, Z., et al., *CYP6B1 and CYP6B3 of the black swallowtail (Papilio polyxenes): adaptive evolution through subfunctionalization*. Mol Biol Evol, 2006. **23**(12): p. 2434-43.
152. Des Marais, D.L. and M.D. Rausher, *Escape from adaptive conflict after duplication in an anthocyanin pathway gene*. Nature, 2008. **454**(7205): p. 762-5.
153. Piatigorsky, J. and G. Wistow, *The recruitment of crystallins: new functions precede gene duplication*. Science, 1991. **252**(5009): p. 1078-9.
154. Hittinger, C.T. and S.B. Carroll, *Gene duplication and the adaptive evolution of a classic genetic switch*. Nature, 2007. **449**(7163): p. 677-81.
155. Proulx, S.R. and P.C. Phillips, *Allelic divergence precedes and promotes gene duplication*. Evolution, 2006. **60**(5): p. 881-92.
156. Otto, S.P. and P. Yong, *The evolution of gene duplicates*. Adv Genet, 2002. **46**: p. 451-83.
157. Penn, D.J., K. Damjanovich, and W.K. Potts, *MHC heterozygosity confers a selective advantage against multiple-strain infections*. Proc Natl Acad Sci U S A, 2002. **99**(17): p. 11260-4.
158. Doherty, P.C. and R.M. Zinkernagel, *A biological role for the major histocompatibility antigens*. Lancet, 1975. **1**(7922): p. 1406-9.
159. B.J. Marshall, *Unidentified curved bacillus on gastric epithelium in active chronic gastritis*, Lancet **1** (1983), pp. 1273–1275.
160. B.J. Marshall and J.R. Warren, *Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration*, Lancet **1** (1984), pp. 1311–1315.
161. Liddell HG and Scott R (1966). *A Lexicon: Abridged from Liddell and Scott's Greek-English Lexicon*. Oxford [Oxfordshire]: Oxford University Press. ISBN 0-19-910207-4.

APPENDIX

I. Confidently predicted domains, repeats and motifs of HcpC and HcpG in the available *H. pylori* genomes using SMART analysis and Signal IP HMM prediction.

Protein	Name	Begin	End	E-value	Probability
G27_HcpG	Signal peptide	1	25		0.991
	SEL1	104	135	0.0289	NA
	SEL1	136	171	5.75E+01	NA
	SEL1	172	207	1.94E-12	NA
	SEL1	208	243	4.12E-07	NA
	SEL1	244	279	5.83E-07	NA
	SEL1	280	315	6.15E-06	NA
J99_HcpG	Signal Peptide	1	25		0.991
	SEL1	105	136	0.0356	NA
	SEL1	137	172	4.78E+01	NA
	SEL1	173	208	1.46E-07	NA
	SEL1	209	244	6.15E-06	NA
B8_HcpG	Signal Peptide	1	27		1
	SEL1	104	135	0.0289	NA
	SEL1	136	171	5.75E+01	NA
	SEL1	172	207	1.94E-12	NA
	SEL1	208	243	4.12E-07	NA
	SEL1	244	279	5.83E-07	NA
	SEL1	280	315	6.15E-06	NA
HpV225_HcpG	Signal Peptide	1	24		0.991
	SEL1	28	59	2.60E+01	NA
	SEL1	101	132	0.285	NA
	SEL1	133	166	0.327	NA
G27_HcpC	Signal Peptide	1	25		0.991
	SEL1	29	60	2.34E+01	NA
	SEL1	61	96	1.98E-09	NA
	SEL1	97	132	6.1E-10	NA
	SEL1	133	168	5.31E-10	NA
	SEL1	169	204	1.18E-07	NA
	SEL1	205	240	1.36E-07	NA
	SEL1	241	276	1.82E-08	NA

II. CFU counts obtained in the AGS cell infection model competition assay between G27MA WT and *slr* mutants used in the study

CFU counts from competition experiments between G27MA WT and G27Δ<i>hepC</i>						
	G27WT			G27D<i>hepC</i>		
	0hr	6hr	24hr	0hr	6hr	24hr
Exp't 1		4080000	6360000		3480000	4680000
Exp't 2	2060000	7440000	7380000	2040000	6540000	5280000
Exp't 3	970000	1044000	1800000	1050000	1104000	1344000
Exp't 4	970000	1488000	2832000	1050000	1572000	2208000
CFU counts from competition experiments between G27MA WT and G27Δ<i>hepC</i>						
	G27WT			G27D<i>hepC</i>		
	0hr	6hr	24hr	0hr	6hr	24hr
Exp't 1	464000	468000	1340000	468000	498000	1020000
Exp't 2	464000	756000	1090000	468000	680000	809000
Exp't 3	464000	522000	990000	468000	564000	758000
Exp't 4	514000	876000	1580000	508000	844000	1060000
Exp't 5	514000	766000	1320000	508000	696000	926000
CFU counts from competition experiments between G27MA WT and G27Δ<i>hepG</i>						
	G27WT			G27D<i>hepG</i>		
	0hr	6hr	24hr	0hr	6hr	24hr
Exp't 1	420000	636000	1632000	426000	744000	1452000
Exp't 2	420000	438000	1164000	426000	492000	1080000
Exp't 3	420000	636000	1044000	426000	582000	984000
Exp't 4	534000	1152000	9000000	540000	1080000	9840000
Exp't 5	534000	656000	10560000	540000	720000	10560000
CFU counts from competition experiments between G27MA WT and G27Δ<i>hepC</i>:Δ<i>hepG</i>						
	G27WT			G27D<i>hepC</i>:Δ<i>hepG</i>		
	0hr	6hr	24hr	0hr	6hr	24hr
Exp't 1	586000	862000	1780000	592000	686000	1020000
Exp't 2	586000	792000	1860000	592000	564000	1160000
Exp't 3	586000	910000	1460000	592000	686000	786000
Exp't 4	586000	946000	1230000	592000	662000	697000
Exp't 5	536000	648000	1080000	532000	480000	670000

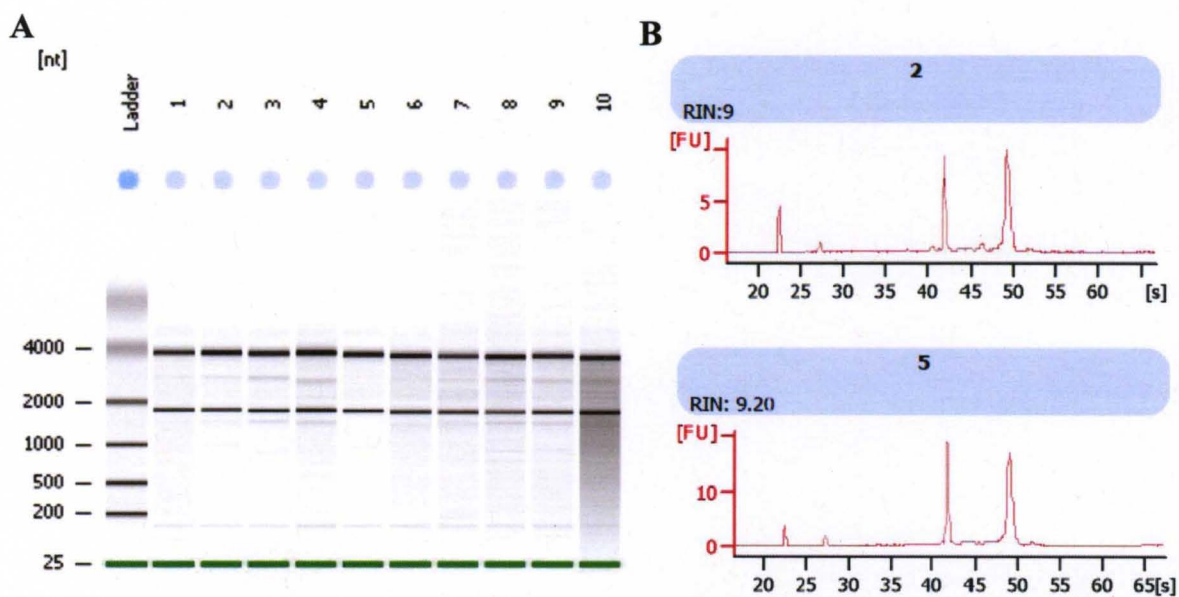
III. CFU counts in *in vitro* competition assays.

CFU counts from <i>in vitro</i> competition experiments between G27MA WT and G27ΔhcpC						
	G27WT			G27ΔhcpC		
	0hr	12hr	56hr	0hr	12hr	56hr
Exp't 1	5640000	16920000	67800000	5980000	17000000	68600000
Exp't 2	5640000	16200000	70200000	5980000	15800000	72800000
Exp't 3	5640000	16520000	69800000	5980000	16600000	75800000
CFU counts from competition experiments between G27MA WT and G27ΔhcpG						
	G27WT			G27ΔhcpG		
	0hr	12hr	56hr	0hr	12hr	56hr
Exp't 1	5640000	19200000	83200000	5480000	18000000	79600000
Exp't 2	5640000	18700000	78600000	5480000	19800000	76800000
Exp't 3	5640000	19000000	80800000	5480000	18800000	78800000
CFU counts from competition experiments between G27MA WT and G27ΔhcpC:ΔhcpG						
	G27WT			G27ΔhcpC-G		
	0hr	12hr	56hr	0hr	12hr	56hr
Exp't 1	5640000	17000000	72000000	5860000	17000000	65600000
Exp't 2	5640000	17600000	68600000	5860000	16800000	69800000
Exp't 3	5640000	18500000	65800000	5860000	17600000	70800000

IV. CFU counts obtained for growth assays of G27MA WT and *hcp* mutants in AGS cell culture infection model.

Growth dynamics of G27WT and G27 - <i>hcp</i> mutants in an AGS cell infection												
	G27WT			G27 Δ <i>hcpG</i>			G27 Δ <i>hcpC</i>			G27 Δ <i>hcpC</i> : <i>AhcpG</i>		
	0hr	6hr	24hr	0hr	6hr	24hr	0hr	6hr	24hr	0hr	6hr	24hr
Exp't 1	2300000	4460000	11500000	2280000	5020000	11800000	2500000	4280000	10700000	2420000	4120000	10200000
Exp't 2	2300000	4780000	12600000	2280000	4860000	11600000	2500000	4420000	12000000	2420000	4480000	11200000
Exp't 3	2820000	4920000	15800000	2780000	4560000	18600000	2620000	4460000	11200000	2860000	4140000	12000000
Exp't 4	2820000	4860000	16200000	2780000	4460000	15600000	2620000	4120000	12000000	2860000	4020000	11000000
Exp't 5	2980000	4320000	18400000	3040000	4340000	19000000	2900000	4200000	13000000	2880000	3980000	10800000

V. Assessment of RNA quality for PCR arrays.



(A) The assessment of RNA integrity with Agilent 2100 bioanalyzer shows the electropherogram of extracted RNA samples (1 – 10) used in PCR arrays. (B) RIN visualization using the Agilent expert software for representative samples from lanes 2 & 5.

Curriculum Vitae

Kalyani Putty

Address and Contact Information 139 Life Sciences Building
Department of Biology
University of Louisville
Louisville, KY 40292
Phone: 502-287-2426
EMail: kalyaniputty@gmail.com, k0putt01@louisville.edu

A. Education

Dec 1999-Jan 2005 Bachelors of Veterinary Science and Animal Husbandry (B.V.Sc.&A.H); equivalent to Doctor of Veterinary Medicine (D.V.M) Acharya N.G.Ranga Agricultural University, Hyderabad.

Jan 2006-Present Working towards Ph.D (to be graduated in Dec 2010) in Dept of Biology, Division of Molecular Cellular and Developmental Biology, University of Louisville, Louisville, KY.

B. Positions Held

Jul. 2004 – July. 2005 Junior Veterinary Resident, Veterinary Poly clinic, Guntur.

Jan. 2006 – Present Graduate Teaching Assistant, Dept of Biology, University of Louisville, Louisville, KY.

C. Academic and Professional Honors

May 2009 Awarded Centre for Genetics and Molecular Medicine (CGeMM) Travel award for attending/presenting at 109th American Society of Microbiology (ASM) Conference, Philadelphia, PA.

D. Research (Jan 2006 – Dec 2010)

My dissertation research focuses on understanding the pathogenesis of gastric pathogen *Helicobacter pylori* and the role of its genetic diversity in geographical differences seen in the clinical outcome of the *H. pylori* infection.

i. Works Published

1. Ogura M, Perez JC, Mittl PRE, Lee H-K, Dailide G, Tan S, Ito Y, Secka O, Dailidienne D, **Putty K**, Berg DE and Kalia A. 2007. *Helicobacter pylori* Evolution: Lineage- Specific Adaptations in Homologs of Eukaryotic Sell-Like Genes. **PLoS Computational Biology**.3(8): e151.
2. Palaniappan Sethu, **Kalyani Putty**, Yongsheng Lian, and Awdhesh Kalia. 2010. Connecting Microbial Population Genetics with Microbial Pathogenesis: Engineering Microfluidic Cell Arrays for High-throughput Interrogation of Host-Pathogen Interaction. **Handbook of Research on Computational and Systems Biology: Interdisciplinary Applications (IGI-Global)**. *In Press*.

ii. Work completed, Manuscripts to be submitted.

1. **Putty K**, Marcus SM, Bogadi L, Hunter A, Mittle PRE, Berg DE, Sethu P, and Kalia A. 2010. Evolution of stable genetic, and functionally non-reciprocal, redundancy in *Helicobacter pylori* cysteine-rich paralogs of Sell-like gene family. **Nature** (*In Prep., To be Submitted*). NOTE: completed work presented in part at 109th ASM, May 2009.
2. **Putty K**, Hunter A, Bogadi L, Darka O, Berg DE Arur S, and Kalia A. 2010. High-affinity interaction of *Helicobacter pylori* HepC with human Ezrin, and its modulation by lineage-specific positive selection. **PLoS One** (*In prep., To be submitted*).
3. Revels EB, Roy EG, **Putty K**, Al-Khodor S, Habyarimana F, Jones S, Miller RD, Abu Kwaik Y and Kalia A. 2011. Molecular evolution of polyubiquitin recruitment in *Legionella pneumophila*. **PLoS One** (*In Prep., To be Submitted*).
NOTE: completed work presented in part at "Posters at the Capitol", Frankfort, KY.

iii. Presentations at International and National Scholarly Meetings.

1. **Putty K**, Marcus SM, Berg DE and Kalia A. 2006. Geographic subdivision facilitates adaptations in *Helicobacter pylori* populations; Pathogen evolution in adaptive landscapes *In*, Research Louisville, Louisville, KY.
2. **Putty K**, Marcus S, Chhabra A, Kersulyte D, Berg DE and Kalia A. 2009. Positive selection and gene duplication drive structural and functional evolution of *Helicobacter* cysteine rich proteins, HcpB, HcpC, and HcpK. *In*, Proceedings of the 109th, General meeting of the American Society for Microbiology, May 12-May 21, Philadelphia, PA.
3. Roy EG, Revels EB, **Putty K**, Al-Khodor S, Habyrimana F, Abu-Kwaik Y, and Kalia A. 2009. Molecular evolution of ankyrin repeat proteins involved in intracellular survival of *Legionella pneumophila*. *In*, Proceedings of Kentucky Posters at the Capitol meeting.
4. Bogadi L, Holley K, **Putty K**, Shriprada S, Sethu P, Kalia A. 2010. Engineering a prototype microfluidic cell array for highthroughput interrogations of host-pathogen interaction. *In*, SROP poster presentations 2010, Louisville, KY.

iv. Research Interests.

My research interests in a broad sense are: 1) To understand the mechanisms of bacterial pathogenesis, 2) Identifying and developing molecular targets for therapeutic purpose and vaccine preparations.

E. Teaching (Spring 2006 – Fall 2010)

i. BIOL104 (Laboratory for Introduction to Biological Systems – 1 credit hour)

This course provides hands – on laboratory experience for non – science students and is designed to develop skills in scientific methodology, observation, and critical thinking.

Role: Teaching Assistant.

ii. BIOL258 (Introduction to Microbiology Laboratory - 1 credit hour)

This course teaches basic microbiological techniques that are necessary to study and safely handle microorganisms and viruses.

Role: Lead Teaching Assistant.

iii. BIOL400 (Histology Lab – 4 credit hours)

This course targets pre – medical students and teaches microscopic anatomy and function of vertebrate tissues and organs.

Role: Teaching Assistant.